https://doi.org/10.7840/kics.2025.50.9.1331

A CAPTCHA System with Face Landmark Multiplication Operation

Minkyu Jo◆, Eunjin Hwang*, Jaeun Kim*, Seung Hyun Jeon°

ABSTRACT

Modern web and mobile applications face significant security threats from malicious automated bots, large-scale unauthorized data extraction, and server overload. Traditional CAPTCHA systems have been widely implemented to mitigate these threats; however, they often lead to an impaired user experience due to repetitive problem-solving and lengthy processing times. To overcome these limitations, this paper proposes a CAPTCHA solution that integrates Generative Artificial Intelligence (Gen AI) with object detection models and simple arithmetic operations, seamlessly combining them with web development techniques. This approach allows humans to quickly and effortlessly solve challenges using image recognition and basic multiplication, while making it difficult for automated programs to succeed—thus ensuring both security and usability. This dynamic CAPTCHA system enhances security by generating diverse and complex challenges that are difficult for bots to bypass, using Generative AI and object detection models. User evaluations show that the system provides an intuitive and efficient experience, making it a promising alternative to traditional CAPTCHA systems. It strikes a balance between security, accessibility, and user experience, offering strong defense against automated threats.

Key Words : CAPTCHA, Gen AI, DDPM-IP, Faster R-CNN, Web Development Integration

Ⅰ. Introduction

CAPTCHAs are security mechanisms designed to differentiate humans from automated bots and to mitigate data extraction and system overload caused by such a bot[1]. However, conventional CAPTCHA systems often have inherent limitations that lead to user frustration[2]. For instance, users may be required to solve the same challenge multiple times, or system errors may force a retry even after a correct response has been entered, negatively affecting user experience and increasing abandonment user dropout[3].

In efforts to bolster security, traditional CAPTCHA systems sometimes compromise user-friendliness. Complex character distortions and challenging image-recognition tasks can frustrate users, especially those with visual or auditory impairments, creating significant accessibility barriers. To address these issues, a new CAPTCHA system is needed—one that balances both user experience and security. Ideally, it should be intuitive and efficient for humans to solve but impervious to automated bots.

Although CAPTCHA systems are fundamentally designed to distinguish humans from bots, users in practice face various challenges[3,4]. For example, text-based CAPTCHAs often prompt incorrect answers due to distorted characters or cluttered backgrounds, and system errors may require repeated attempts even after a correct response has been given. This increases user fatigue and is a primary driver

◆ First Author : Daejeon University, Department of Computer Engineering, jomk5023@gmail.com, 학생회원
° Corresponding Author : Daejeon University, Department of Computer Engineering, creemur@dju.kr, 정회원
* Daejeon University, Department of Computer Engineering, eunijhwang@gmail.com, byu0408@gmail.com
논문번호 : 202502-043-A-RE, Received February 24, 2025; Revised April 3, 2025; Accepted April 18, 2025

of system abandonment[4]. Image-based CAPTCHAs also suffer from large database constraints, causing users to see the same tasks repeatedly—particularly when they need to select or rotate specific objects. Moreover, cultural and cognitive differences can further complicate task comprehension for some users[5,6].

Audio-based CAPTCHAs were introduced as an alternative to visual ones. However, audio clips with excessive background noise or unclear pronunciation can hinder comprehension, particularly for non-native speakers. Consequently, in attempting to enhance security, these traditional CAPTCHA systems often increase user frustration by making the challenges more difficult[7,8].

The primary issues this thesis aims to address are twofold. First, it seeks to minimize unnecessary repetition and maximize efficiency by offering intuitive, straightforward challenges—using human face images and basic multiplication tasks that humans can easily recognize. Second, it leverages Generative Artificial Intelligence (Gen AI) to generate dynamic challenges, thus boosting both data diversity and robustness. Traditional, statically generated CAPTCHAs rely on fixed databases, which pose security risks once the limited data pool is repeatedly exploited. In contrast, by integrating Denoising Diffusion Probabilistic Models with Input Perturbation (DDPM-IP) and Faster Region-based Convolutional Neural Network (Faster R-CNN), we propose a dynamic CAPTCHA system capable of generating an effectively infinite dataset[9,10].

In this paper, we present a CAPTCHA system that combines generated human face images with landmark-based multiplication challenges to tackle these problems. The proposed system offers several advantages. First, it maximizes user convenience by using elements (facial landmarks and simple math) that humans can intuitively understand, eliminating the need for complex text recognition or image selection. Second, it enhances robustness. Dynamically generated images from DDPM-IP and Faster R-CNN exhibit high diversity and complexity, making them harder for bots to learn or circumvent[11]. Third, it improves efficiency. Simple arithmetic-based authentication conserves system resources and reduces user-server interactions. In this paper, we generate challenges from real human facial landmarks using generative images. By incorporating basic math operations, we overcome the limitations of traditional CAPTCHAs and simultaneously enhance both security and usability. The user evaluation results show that the proposed face landmark multiplication-based CAPTCHA system was successfully assessed by 30 participants. The participants found the system to be intuitive, easy to use, and more enjoyable and faster compared to traditional CAPTCHA systems. Although the average solving time was slightly slower, it is expected to improve significantly with increased familiarity with the system. These evaluation results suggest that the proposed system has the potential to meet both security and user experience requirements.

Ⅱ. Related Works

CAPTCHAs have evolved to incorporate various media, including text, images, audio, and video, to deter automated bot programs. While text-based CAPTCHAs have been preferred for their low latency and minimal storage requirements, they are still susceptible to optical character recognition (OCR) and machine learning techniques, even when employing methods such as font distortion, noise, and text overlap.

Moreover, they pose usability challenges, as users often struggle with highly distorted text[4-6]. Image-based CAPTCHAs offer improved usability compared to text-based CAPTCHAs and typically involve selecting or rotating specific objects. However, they require a large image database and may become less efficient due to server processing time and download latency. Furthermore, using fixed image databases creates vulnerabilities that attackers can exploit through repeated attempts or learning[7].

Audio-based CAPTCHAs rely on auditory input to authenticate users, requiring them to listen to an audio clip (which may include background noise) and then enter the spoken content. However, accent variations and background noise can be especially challenging for non-native speakers, thereby limiting overall usability[8,12]. Video-based CAPTCHAs rely on audio-visual content; however, their practicality is limited by high data requirements and network latency[7]. Gamified CAPTCHAs are designed to provide a logical and engaging user experience, but their complexity can lead to user fatigue. In recent years, deep learning-based techniques such as Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs) have been used to compromise CAPTCHAs, significantly threatening the security of traditional CAPTCHA designs[5]. While CNNs and GANs offer high accuracy, they also require high-quality training data and substantial computational resources, resulting in high initial costs.

Traditional CAPTCHA systems are vulnerable to attacks due to design flaws, with pattern matching, segmentation attacks, and end-to-end machine learning models commonly reported as attack techniques[5,13]. Gen AI-based CAPTCHA systems have emerged as an innovative solution to overcome these limitations, enhancing both security and usability[14-16].

The Gen AI-based CAPTCHA system addresses many limitations of traditional CAPTCHA systems and introduces an innovative approach that enhances both security and usability[14-16]. Image-based CAPTCHAs, for example, often require 20 to 40 seconds to solve, and some solutions remain vulnerable to automated bot attacks (e.g., comparing response times between CapSolver and AntiCaptcha). In contrast, Gen AI-based CAPTCHAs resolve more quickly—typically in 5 to 15 seconds—and are designed to make the challenge-solving process engaging for users, thus significantly improving usability. Notably, while CAPTCHA systems based on behavioral data (such as Google reCAPTCHA) raise privacy concerns, the proposed system maintains a high level of security without collecting personal information.

Previous research shows that systems like FunCAPTCHA reinforce security by requiring logical reasoning and spatial imagination; however, they still exhibit vulnerabilities exploitable by certain AI techniques. By contrast, Gen AI-based CAPTCHAs can generate challenges that are not easily learned by AI models, continually evolving based on user feedback and thereby gaining an advantage in the AI "arms race." Consequently, Gen AI-based CAPTCHAs offer faster resolution times, stronger security, and enhanced user experience compared to traditional systems, establishing a new benchmark for CAPTCHA research and practical implementation[15-16]. The CAPTCHA system proposed in this paper addresses multiple limitations of existing CAPTCHA systems and provides an innovative approach that improves both security and user convenience. Table 1 presents a logical and objective comparison between the proposed system and three representative CAPTCHA solutions, highlighting their key shortcomings and demonstrating how the proposed system mitigates these issues while delivering superior performance[17].

Large Language Model (LLM)-based CAPTCHA systems are adept at handling complex tasks by utilizing advanced reasoning; however, they also face several key limitations. First, they rely on a multi-step solution process, meaning that if one step fails, the entire task fails. In contrast, the CAPTCHA system proposed in this paper avoids this structural weakness by resolving challenges through object detection and simple mathematical operations in a single step.

Second, LLMs experience significant performance degradation in tasks that involve multiple objects and relationships, with success rates for multi-criteria tasks dropping to 25%. To maintain high accuracy, this paper adopts a straightforward yet effective approach using images and landmark-based challenges. Third, LLM-based systems frequently encounter hal-

Table 1. Comparative Analysis of CAPTCHA Systems.

CAPTCHA	Inference Ability	Versatility	Dependency	Calculation cost
LLM-based CAPTCHA[14]	○	○	High	-
Diff-CAPTCHA[14]	○	△	High	High
Cycle-GAN Based Text CAPTCHA[18]	-	○	High	High

lucination (misinformation) issues when dealing with long instructions or multi-step tasks[14], which can compromise CAPTCHA reliability and security. However, the system proposed here operates on generated images and object detection, circumventing the hallucination problem[18].

Cycle-GAN-based text CAPTCHA attack systems utilize synthetic data to train on CAPTCHAs. Nevertheless, due to the lack of similarity between synthetic and real data, these systems exhibit low success rates (33.8%, 36.1%) on complex CAPTCHA schemes such as those employed by Microsoft and Tencent. In contrast, our proposed system leverages high-quality data from the CelebFaces Attributes Dataset (CelebA) and Gen AI, overcoming these limitations. Moreover, text-based approaches often see declining success rates as text length and complexity increase, where even a single character error can cause a complete failure. By employing landmark-based challenges, we eliminate text-length constraints and enable simple, intuitive problem-solving.

Traditional CAPTCHA systems also place a heavy emphasis on data collection and labeling. To address this, we utilize DDPM-IP to dynamically generate data, thereby reducing resource dependency and boosting data diversity. Diff-CAPTCHA, which relies on Denoising Diffusion Probabilistic Models (DDPM), produces high-quality images but suffers from high computational overhead and slow sampling. Here, we adopt DDPM-IP to improve sampling efficiency and deliver a fast response that is well-suited for real-time traffic.

Furthermore, Diff-CAPTCHA can degrade user readability due to complex characters and background interference, undermining the user experience. In our proposed system, we incorporate simple landmark-based questions to maximize user convenience. Lastly, while Diff-CAPTCHA depends on a specific dataset, limiting its use in diverse environments, we combine CelebA with DDPM-IP to construct a large-scale dataset and validate usability under a variety of conditions.

Ⅲ. Face Landmark Multiplication Operation CAPTCHA

3.1 Dataset Preprocessing

For dataset preprocessing, we used the CelebA from Kaggle, which contains 202,599 facial images collected from 10,177 celebrities. Each image is annotated with five key landmarks: the right eye, left eye, nose, right corner of the mouth, and left corner of the mouth. Because CelebA features a wide variety of facial angles and backgrounds, it is highly suitable for tasks like facial attribute recognition, face detection, landmark localization, and face editing or synthesis, all of which demand large-scale and diverse data. In this paper, we constructed an Auto CAPTCHA system based on this dataset, resizing the original 127×218 images to 128×128 for consistent data formatting. We then designed a process that automatically generates and verifies the visual elements needed for CAPTCHA by linking landmark positions with attribute information derived from the transformed images. Through this approach, it becomes feasible to conduct research on a large scale, while also improving model generalizability by incorporating diverse backgrounds and facial angles.

3.2 Diffusion-Driven Image Generation with DDPM-IP

In this paper, we adopted the DDPM-IP model as the core framework for image generation because it not only produces high-quality images but also generates samples that closely match realistic distributions. Conventional DDPM models create samples through an incremental computationally expensive and susceptible to cumulative errors over multiple steps. To overcome these limitations, the DDPM-IP model introduces a correction term that aligns the scores of model-generated images more closely with those of real data, thus guiding the generated samples to a more realistic distribution. Notably, the DDPM-IP model excels on high-resolution portrait datasets such as CelebA, producing sharp, lifelike images. Building on these strengths, we aim to generate high-quality, well-structured portrait images to system. We selected DDPM-IP as the optimal form a foundational dataset for the CAPTCHA solution because it demonstrates robust performance across diverse data conditions and aligns well with our research objectives[8]. As shown in Fig. 1, the DDPM-IP model achieves notably high performance on large-scale, high-resolution portrait datasets such as CelebA. When dealing with varied facial angles and backgrounds, it effectively reconstructs fine facial details in a natural manner, yielding images whose visual quality is comparable to real photographs.

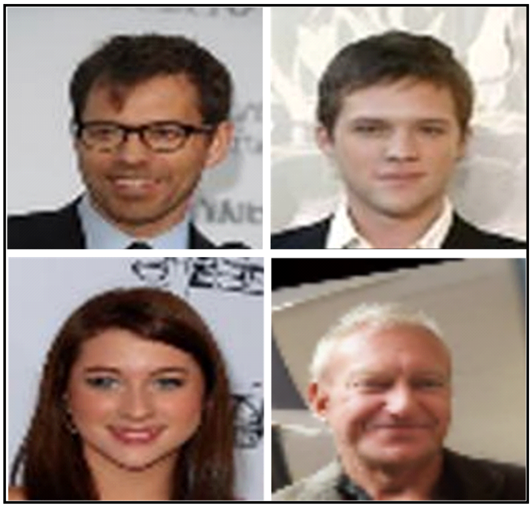


Fig. 1. Four images generated by the DDPM-IP MODEL.

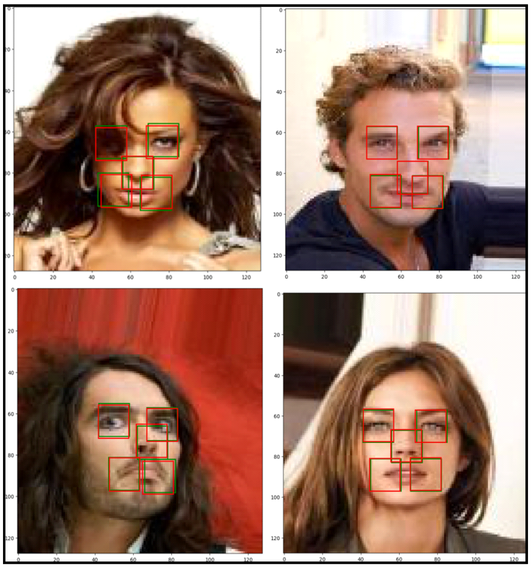


Fig. 2. Faster R-CNN Utilizing the CelebA Dataset.

3.3 Object Detection

In this paper, we employed the Faster R-CNN model for facial landmark extraction and object detection in our CAPTCHA system. This deep learning-based framework comprises two main components: a Region Proposal Network (RPN) and an object detector. The RPN transforms input images into feature maps via convolutional layers and efficiently identifies regions that are likely to contain objects, thereby reducing the initial number of candidate regions and minimizing computational costs while maintaining high detection accuracy. The object detector then classifies the objects in these proposed regions and predicts precise bounding boxes. We chose Faster R-CNN for the following reasons. First, it provides the precision and real-time capabilities needed for accurate facial landmark extraction in CAPTCHA systems. Second, it delivers robust detection performance across various facial angles and complex backgrounds, enhancing the reliability of landmark extraction.

Third, it effectively learns from large-scale datasets like CelebA, which include detailed landmark information—thus improving data processing and accuracy in CAPTCHA. Moreover, its high versatility facilitates easy adaptation to applications beyond CAPTCHA systems.

In this paper, RPNs were used to quickly generate candidate object regions, followed by classification and bounding box adjustment to detect the final object. Specifically, we trained the model to generate bounding boxes for five major landmarks present in the CelebA dataset.

In Fig. 2, the green bounding boxes represent the actual landmark locations recorded in the CelebA dataset, while the red bounding boxes show the predictions made by the Faster R-CNN model. In all four images, the predicted and actual values closely match, demonstrating the high accuracy of the model. By using Faster R- CNN, this research improves the security of CAPTCHAs and the processing speed. In this paper, we implemented a single code that integrates the results obtained from the two models, Figs. 1 and 2, and generates the result image by assigning the landmark values predicted by Faster R-CNN on top of the portrait image generated by the

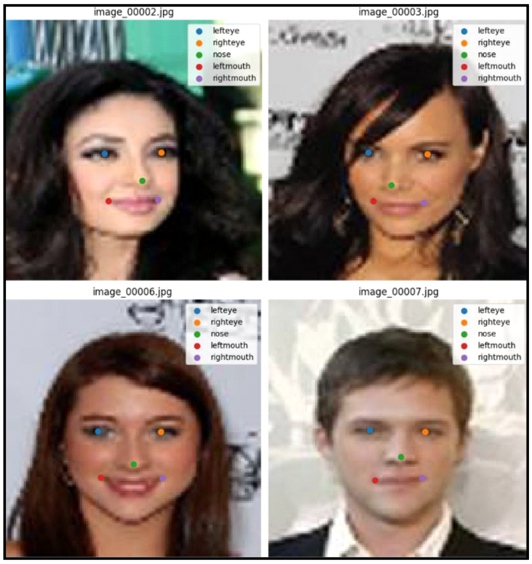


Fig. 3. Resulting images generated through the collaboration of DDPM-IP and Faster R-CNN.

DDPM-IP model. In Fig. 3, we can see that the high-resolution face image generated by the DDPM-IP model is well represented by the accurate landmark locations predicted by the Faster R-CNN. Fig. 5. shows images stored in the database. The interconnection between the models maximizes the synergy between generative and object detection models, which can be used in a variety of future applications, including more sophisticated face synthesis, data augmentation, and personalized CAPTCHA generation. Fig. 4 shows the overall modeling structure for representing five landmarks in an image generated by DDPM-IP and Faster R-CNN.

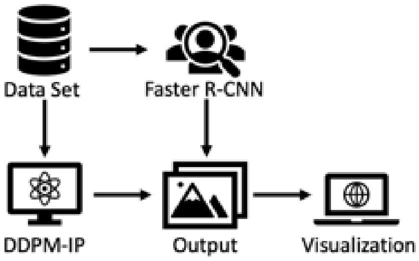
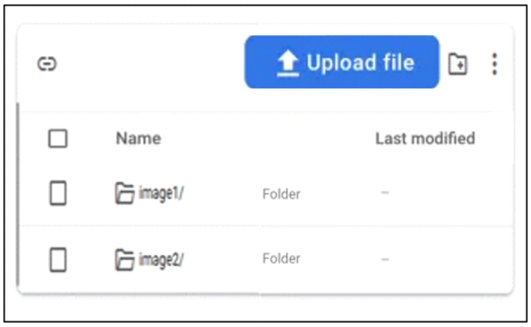


Fig. 4. Integrated System Architecture for Object Detection and Gen AI.

3.4 CAPTCHA System Implementation

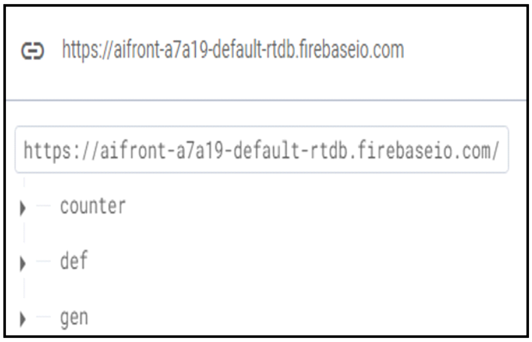
The CAPTCHA developed in this paper is implemented as a web-based platform and utilizes HTML, CSS, and JavaScript to provide an intuitive and efficient interface. When a user requests data through the web interface, the request is sent to the server where the AI model is built via FastAPI, and the server processes it and returns the generated image and coordinate data. The returned data is stored in a Firebase-based cloud database for efficient data management and accessibility. This architecture enables seamless integration between front-end and back-end and ensures real-time data processing and scalability. Firebase Storage is organized into two directories (image1, image2) that manage existing images and images created with generative models, respectively.

This enables efficient search and management of data and enhances integration with CAPTCHA systems. Fig. 5 illustrates the storage architecture, showing how existing (image1) and newly generated (image2) images are organized in separate directories while maintaining an efficient flow for data retrieval and updates. Fig. 6 describes this data structure and the functions of each group, and Fig. 7 explains how file and index names are determined based on the "counter" function described of Fig. 6. The function described of Fig. 6 serves to track and manage images within the database by managing the unique identifiers of the generated images, while the "def" function described of Fig. 6 stores the coordinates of landmarks in the images contained in the image1 directory, which are used to analyze and validate key features. The "gen" function described of Fig. 6 stores the landmark coordinates of images generated using Faster R-CNN in the image2 directory and reflects the object



□	Name		Last modified
□	image1/	Folder	–
□	image2/	Folder	–

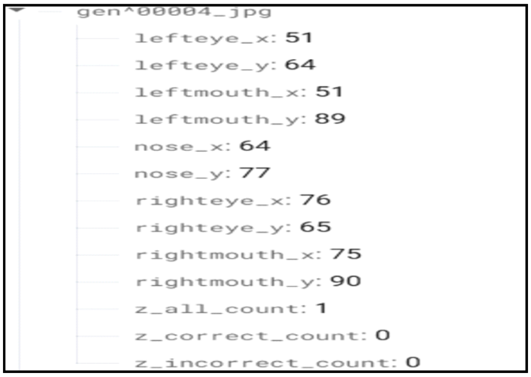
Fig. 5. Images stored in the database.



```
https://aifront-a7a19-default-rtdb.firebaseio.com

https://aifront-a7a19-default-rtdb.firebaseio.com/
▸ counter
▸ def
▸ gen
```

Fig. 6. Structure and Function of Realtime Database Groups in CAPTCHA System.



```
▾ gen^00004_jpg
    lefteye_x: 51
    lefteye_y: 64
    leftmouth_x: 51
    leftmouth_y: 89
    nose_x: 64
    nose_y: 77
    righteye_x: 76
    righteye_y: 65
    rightmouth_x: 75
    rightmouth_y: 90
    z_all_count: 1
    z_correct_count: 0
    z_incorrect_count: 0
```

Fig. 7. Figure X: Determination of File and Index Names Based on Counter Values.

detection results. Variables such as z_all_count, z_correct_count, and z_incorrect_count are updated in real time Security Architecture according to the CAPTCHA verification results. When the configured (or designated) verification count (z_all_count) is exhausted, the image is moved to the image1 directory or deleted based on the success/failure ratio. This compensates for the limited dataset of image1 data, keeps CAPTCHA questions varied and fresh, and prevents the bot from weakening security due to repeated learning.

The CAPTCHA authentication flow starts when the user connects to the web and the authentication logic kicks in. During this process, the user must solve a challenge generated by the CAPTCHA server, and if successful, they are directed to the Real Data path to access normal pages or services. On the other hand, if the user is unable to solve the problem or submits an incorrect answer, they are directed to a restricted or error page and are provided with Fake Data. CAPTCHA success and failure results are sent back to the server and recorded in the database to continuously expand the data pool between Real Data and Gen AI Image and enhance the security of the CAPTCHA system. This structure prevents bots from defeating CAPTCHAs through iterative learning and contributes to the security and reliability of CAPTCHAs by reflecting real-time user authentication results. Fig. 8 illustrates how these authentication flows and data management structures enhance the security of the CAPTCHA system.

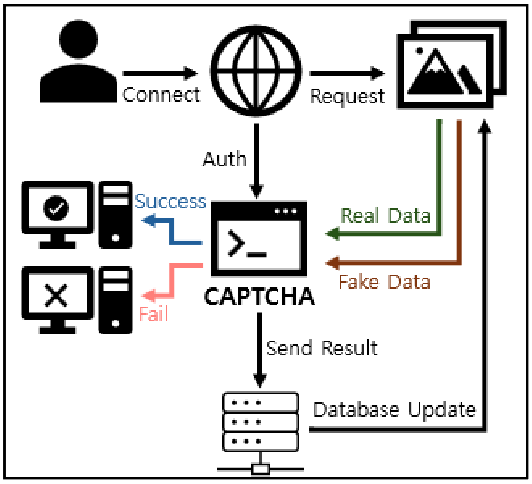
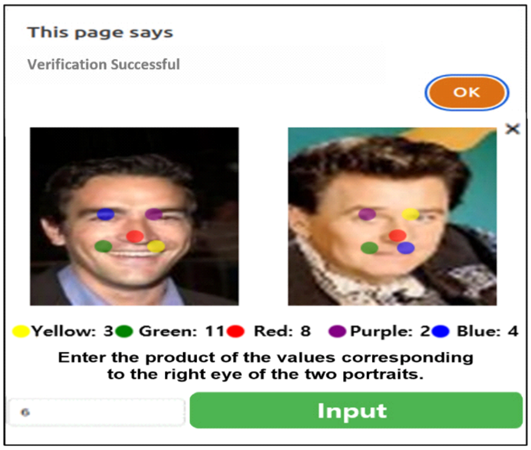


Fig. 8. CAPTCHA-Based User Authentication and Data.

Ⅳ. Verification of the proposed CAPTCHA

We developed a system to compare the left image taken from the CelebA dataset with the right image generated by the DDPM-IP model. Both images were processed using the object detection model Faster R-CNN, which assigns five landmarks to each face (left eye, right eye, nose, left corner of the mouth, and right corner of the mouth). Based on these landmarks, we built a fast CAPTCHA system that allows users to easily distinguish between a human and a bot by performing simple multiplication operations. This method is fast and accurate and can effectively block bots by verifying that the user is entering the correct answer.

Fig. 9 shows the finalized Auto CAPTCHA. As shown in Fig. 10, the left image is selected from the CelebA dataset and the right image is generated by the DDPM-IP model. The user is given a CAPTCHA that requires them to solve a simple multiplication problem while viewing the two images simultaneously. For example, if the left image is given a right eye value of 2 and the right image is given a right eye value of 3, the result of multiplying the respective right eye values is 6. If the user enters these values correctly, they will see an "Authentication Successful" message, and this simple problem solving can effectively block bot access in a short amount of time. This structure expands the possibilities of CAPTCHA systems in terms of providing engaging experience for users without requiring complex string input, while maintaining high security.



This page says

Verification Successful

● Yellow: 3 ● Green: 11 ● Red: 8 ● Purple: 2 ● Blue: 4

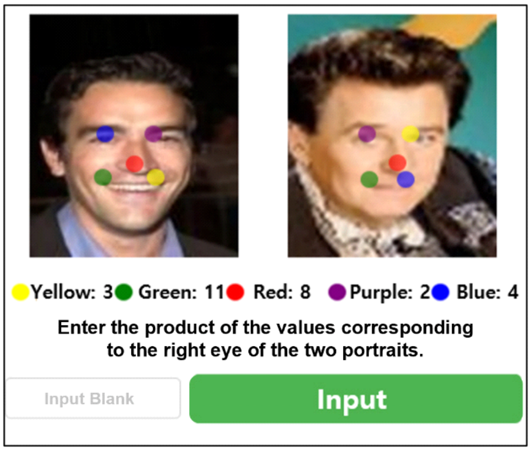
Enter the product of the values corresponding to the right eye of the two portraits.

6

Fig. 9. Final Automatic CAPTCHA System.

Referring to Fig. 11, the value for the left corner of the mouth in the left image is set to 4, while the value for the left corner of the mouth in the right (generated) image is set to 2, resulting in the correct answer of 8. However, if the user enters 7, the web-page displays an "Authentication Failed" message and automatically refreshes to present a new image and multiplication question. Through this process, the CAPTCHA system can effectively block not only simple incorrect entries, but also bots that make repeated attempts, further enhancing security.

Table 2 shows that the proposed CAPTCHA system in this paper has clear advantages over existing systems in terms of security, user experience, and efficiency through the combination of Gen AI and Faster R-CNN. The proposed system maximizes data diversity and generation speed, reduces user fatigue through a simple problem-solving approach, and enhances security. It shows promise as a practical and scalable CAPTCHA solution for a variety of web environments. LLM-based CAPTCHAs are highly data-dependent, relying on a large amount of pre-trained data and a complex model structure. They also require a multi-step solving process, so an error in a particular step can cause the entire operation to fail. Performance degrades sharply, especially for problems that require multiple objects and relationships to be considered, and hallucination (misinformation) can

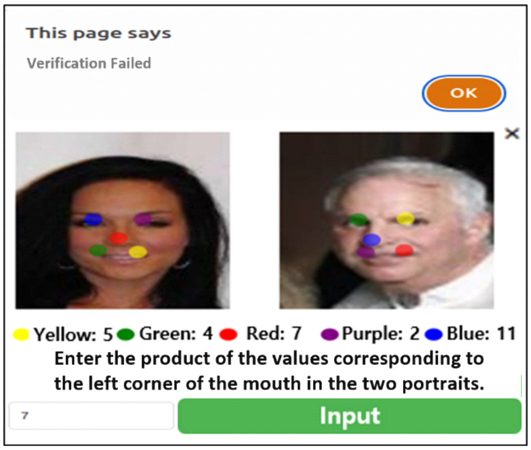


● Yellow: 3 ● Green: 11 ● Red: 8 ● Purple: 2 ● Blue: 4

Enter the product of the values corresponding to the right eye of the two portraits.

Input Blank

Fig. 10. Final Automatic CAPTCHA System - Correct Image.



This page says

Verification Failed

● Yellow: 5 ● Green: 4 ● Red: 7 ● Purple: 2 ● Blue: 11

Enter the product of the values corresponding to the left corner of the mouth in the two portraits.

7

Fig. 11. Final Automatic CAPTCHA System - Incorrect Image.

Table 2. Comparative Analysis of CAPTCHA Systems in Terms of Performance and Characteristics.

CAPTCHA	Dependency	Calculation cost
LLM-based CAPTCHA[14]	High	-
Diff-CAPTCHA[14]	High	High
Cycle-GAN Based Text CAPTCHA[18]	High	High
Proposed CAPTCHA system	Low	Low

be a frequent problem for long commands or multi-step tasks. The proposed CAPTCHA system, on the other hand, overcomes these structural weaknesses by combining object detection and simple math operations to solve the problem in a single step.

Diff-CAPTCHA is based on DDPM to generate high-quality images, but its high computational cost and sampling speed make it unsuitable for real-time applications. Moreover, complex characters and background interference sometimes make it difficult for users to read CAPTCHAs, which can degrade the user experience. In contrast, the proposed CAPTCHA system utilizes DDPM-IP to improve sampling efficiency and provide a fast response rate suitable for real-time traffic. Additionally, we maximize user convenience by using simple landmark-based questions.

Cycle-GAN-based text CAPTCHAs utilize synthetic data to generate CAPTCHAs, but their performance is limited by the lack of similarity between synthetic and real data. In particular, the have a low success rate (33.8% and 36.1%) on complex CAPTCHA schemes such as Microsoft and Tencent, and the success rate decreases further as text length and complexity increase. Furthermore, a single-character prediction error is more likely to lead to overall failure. The proposed CAPTCHA system, on the other hand, is designed to utilize landmark-based questions to enable simple and intuitive problem solving, free from the constraints of text length and complexity.

The proposed CAPTCHA system overcomes the limitations of existing CAPTCHA systems by leveraging high-quality data generated via CelebA and Gen AI. While existing systems suffer from high data dependency and computational overhead, the proposed CAPTCHA uses DDPM-IP to dynamically generate data, reduce resource dependency, and improve sampling efficiency. In addition, it is designed to maintain high security without collecting personal information, providing a safer and more practical alternative to existing CAPTCHA systems.

Ⅴ. User Evaluation

To evaluate the usability of the proposed CAPTCHA system with face landmark multiplication, participants were asked to respond to the system[19]. The evaluation was conducted with a total of 30 participants, randomly selected from the Department of Computer Engineering, where the three authors of this paper are affiliated, considering diversity in age and gender. The experiment began with a brief explanation of general CAPTCHA systems, followed by an interactive session in which participants used the proposed face landmark multiplication-based CAPTCHA system. During the experiment, response time and error rate were measured. After completing the task, a survey was conducted to collect participants' subjective feedback on the system.

The survey consisted of the following four questions:

- I want to use this CAPTCHA system frequently.
- I think this CAPTCHA system is more fun than other CAPTCHA systems.
- I feel that this CAPTCHA system is much faster than other CAPTCHA systems.
- I feel that this CAPTCHA system is easy to use.

Responses were based on a 5-point Likert scale: *Strongly disagree*, *Disagree*, *Neutral*, *Agree*, and *Strongly agree*. To quantify subjective feedback, each response was converted into a numerical score ranging from 1 to 5 (Strongly disagree = 1, ..., Strongly agree = 5).

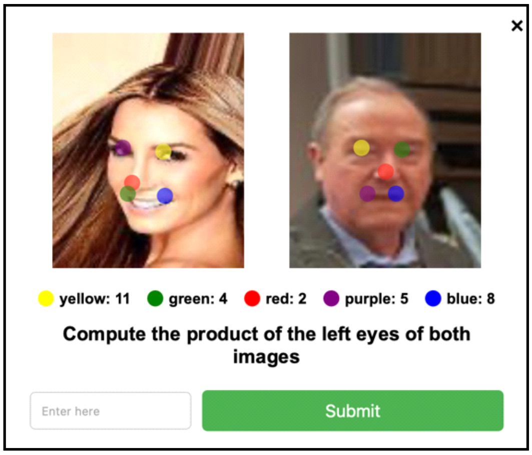
Fig. 12 and 13 show images related to the usability evaluation of the proposed CAPTCHA system. The evaluation was carried out in this way by 30 participants. Fig. 14 shows the survey form used to assess subjective satisfaction after completing the CAPTCHA system. Table 3 shows that the proposed

CAPTCHA system with face landmark multiplication operation achieved a success rate of 100% (error rate 0%) with an average solving time of 13.5 seconds, based on evaluations from 30 randomly selected participants. The minimum solving time was 4.90 seconds, and the maximum solving time was 37.12 seconds.

The average age of the participants was 22.4 years. The evaluation was conducted on a system equipped with an Intel i9-10980Xe CPU, two NVIDIA GeForce RTX 3090 GPUs, and 256GB of RAM, ensuring high computational performance. Additionally, benchmark results from Geekbench 6, which measures the performance, showed a single-core score of 1,325 and a multi-core score of 9,899, indicating that the system can run efficiently even in lightweight environments.

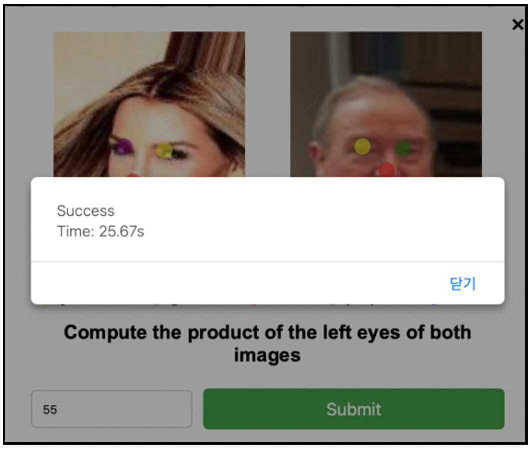
Table 3. Summary of Usability Evaluation Metrics.

Metric	Mean
Success Rate	100%
Average Solving Times	13.5 second
Question 1	4.1
Question 2	4.3
Question 3	3.9
Question 4	4.3



Compute the product of the left eyes of both images

Fig. 12. Evaluation of the Usability of the Proposed CAPTCHA System.



Success
Time: 25.67s

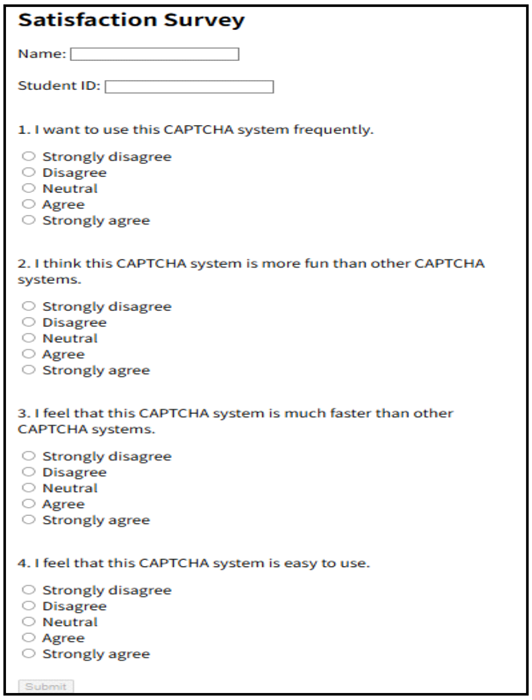
Compute the product of the left eyes of both images

55

Fig. 13. Evaluation of the Usability of the Proposed CAPTCHA System – Success Image.

The user feedback, quantified on a 5-point Likert scale, showed the following average scores: 4.1 for Question 1 (frequent use), 4.3 for Question 2 (more fun), 3.9 for Question 3 (faster than others), and 4.3 for Question 4 (easy to use).

Although the average solving time was slower than expected, it is likely due to participants encountering the proposed CAPTCHA system for the first time and being unfamiliar with it. With increased familiarity, the solving time is expected to improve significantly.



Satisfaction Survey

Name: ________

Student ID: ________

1. I want to use this CAPTCHA system frequently.

○ Strongly disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly agree

2. I think this CAPTCHA system is more fun than other CAPTCHA systems.

○ Strongly disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly agree

3. I feel that this CAPTCHA system is much faster than other CAPTCHA systems.

○ Strongly disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly agree

4. I feel that this CAPTCHA system is easy to use.

○ Strongly disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly agree

Fig. 14. A Survey aimed at assessing subjective satisfaction.

Ⅵ. Conclusions

To simultaneously enhance security and user experience, this study proposes a novel CAPTCHA system that generates images using the CelebA dataset and the DDPM-IP model, detects facial landmarks via Faster R-CNN, and designs challenges in the form of simple math operations. Compared to traditional systems, this approach significantly improves both usability and robustness. The system is supported by a Firebase-based real-time data management framework and a web-based architecture, ensuring high scalability, easy maintenance, and continuous performance improvement through automatic updates of the image pool. User testing with 30 participants demonstrated a 100% success rate and an average solving time of 13.5 seconds, with high ratings for usability, enjoyment, and ease of use. The proposed system combines multiplication with landmark detection and natural language processing, making it difficult for bots, due to its integrated visual and linguistic complexity, yet intuitive and accessible for human users. These results highlight the system's effectiveness in achieving a strong balance between security and user convenience, offering a promising direction for future CAPTCHA development.

References

[1] L. von Ahn, et al., "CAPTCHA: Using hard AI problems for security," in *Int. Conf. Theory and Applications of Cryptographic Techniques*, pp. 294-311, Berlin, Germany, May 2003. (https://doi.org/10.1007/3-540-39200-9_18)

[2] A. Searles, et al., "An empirical study & evaluation of modern CAPTCHAs," in *Proc. 32nd USENIX Security 23*, pp. 3081-3097, Anaheim, USA, Aug. 2023. (https://doi.org/10.48550/arXiv.2307.12108)

[3] N. Tanthavech and A. Nimkoompai, "CAPTCHA: Impact of website security on user experience," in *Proc. 2019 4th Int. Conf. Intell. Inf. Technol.*, pp. 37-41, Da Nang, Vietnam, Feb. 2019. (https://doi.org/10.1145/3321454.3321459)

[4] Y.-W. Chow, W. Susilo, and P. Thorncharoensri, "CAPTCHA design and security issues," in *Advances in Cyber Security: Principles, Techniques, and Appl.*, Springer, pp. 69-92, 2019. (https://doi.org/10.1007/978-981-13-1483-4_4)

[5] P. Wang, et al., "An experimental investigation of text-based CAPTCHA attacks and their robustness," *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1-38, Nov. 2023. (https://doi.org/10.1145/3559754)

[6] Z. Ouyang, et al., "A cloud endpoint coordinating CAPTCHA based on multi-view stacking ensemble," *Comput. & Secur.*, vol. 103, p. 102178, 2021. (https://doi.org/10.1016/j.cose.2021.102178)

[7] H. Weng, et al., "Towards understanding the security of modern image CAPTCHAs and underground CAPTCHA-solving services," *Big Data Mining and Analytics*, vol. 2, no. 2, pp. 118-144, Jun. 2019. (https://doi.org/10.26599/BDMA.2019.9020001)

[8] N. Tariq and F. A. Khan, "Match-the-sound CAPTCHA," in *Proc. 14th Int. Conf. Inf. Technol.-New Generations*, pp. 803-808, Las Vegas, USA, Apr. 2017. (https://doi.org/10.1007/978-3-319-54978-1_99)

[9] M. Ning, et al., Input perturbation reduces exposure bias in diffusion models(2023), Retrieved Dec., 23, 2024, from https://arxiv.org/abs/2301.11706 (https://doi.org/10.48550/arXiv.2301.11706)

[10] S. Ren, et al., "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137-1149, Jun. 2016. (https://doi.org/10.48550/arXiv.1506.01497)

[11] G. Goswami, et al., "FaceDCAPTCHA: Face detection based color image CAPTCHA," *Future Generation Computer Syst.*, vol. 31, pp. 59-68, Feb. 2014. (https://doi.org/10.1016/j.future.2012.08.013)

[12] J. Lazar, et al., "The SoundsRight CAPTCHA: An improved approach to audio human

interaction proofs for blind users," in *Proc. SIGCHI Conf. Human Factors in Computing Syst.*, pp. 2267-2276, Austin, USA, May 2012. (https://doi.org/10.1145/2207676.2208385)

[13] N. Tariq, et al., CAPTCHA types and breaking techniques: Design issues, challenges, and future research directions(2023), Retrieved Dec., 23, 2024. (https://doi.org/10.48550/arXiv.2307.10239)

[14] G. Deng, et al., Oedipus: LLM-enhanced reasoning CAPTCHA solver(2024), Retrieved Dec., 23, 2024. (https://doi.org/10.48550/arXiv.2405.07496)

[15] J. Zhang, et al., "A secure annuli CAPTCHA system," *Comput. & Secur.*, vol. 125, p. 103025, Feb. 2023. (https://doi.org/10.1016/j.cose.2022.103025)

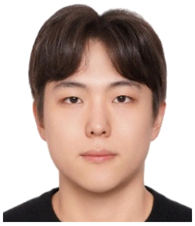
[16] P. Wang, et al., "Extended research on the security of visual reasoning CAPTCHA," *IEEE Trans. Dependable Secure Comput.*, vol. 20, no. 6, pp. 4976-4992, Jan. 2023. (https://doi.org/10.1109/TDSC.2023.3238408)

[17] M. Guerar, et al., "Gotta CAPTCHA'Em all: A survey of 20 years of the human-or-computer dilemma," *ACM Comput. Surv.*, vol. 54, no. 9, pp. 1-33, Oct. 2021. (https://doi.org/10.1145/3477142)

[18] C. Li, et al., "End-to-end attack on text-based CAPTCHAs based on cycle-consistent generative adversarial network," *Neurocomputing*, vol. 433, pp. 223-236, Dec. 2021. (https://doi.org/10.1016/j.neucom.2020.11.057)

[19] W. Yang and T. Kwon, "Emerging image cue CAPTCHA resisting automated and human-solver-based attacks," *J. Korea Inst. Inf. Secur. & Cryptology*, vol. 27, no. 3, pp. 531-540, Jun. 2017. (https://doi.org/10.13089/JKIISC.2017.27.3.531)

Minkyu Jo



Mar. 2020~Current : B.S. in Computer Engineering, Dae-jeon Univ.

<Research Interests> Computer vision, bioinformatics, protein structure prediction.

[ORCID:0009-0002-7985-6065]

Eunjin Hwang

Mar. 2022~Current : B.S. in Computer Engineering, Dae-jeon Univ.

<Research Interests> Autonomous driving, computer vision.

[ORCID:0009-0009-9004-9177]

Jaeun Kim

Feb. 2025 : B.S. in Computer Engineering, Dae-jeon Univ.

Mar. 2025~Current : M.S. in Computer Engineering, Dae-jeon Univ.

<Research Interests> Computer vision, machine learning, data analysis.

[ORCID:0009-0001-0261-0270]

Seung Hyun Jeon

Feb. 2017 : Ph.D. School of Electrical Engineering, KAIST, Korea

Jul. 2018~Mar. 2023 : KT R&D Center, KT, Korea

Mar. 2023~Current : Dept. of Computer Engineering, Dae-jeon Univ.

<Research Interests> Machine learning, blockchain networks, energy consumption models for networks.

[ORCID:0000-0001-7303-4672]