Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.062729

ARTICLE

A Novel CAPTCHA Recognition System Based on Refined Visual Attention

Zaid Derea[1,2,*], Beiji Zou[1], Xiaoyan Kui[1,*], Monir Abdullah[3], Alaa Thobhani[1] and Amr Abdussalam[4]

[1]School of Computer Science and Engineering, Central South University, Changsha, 410083, China
[2]College of Computer Science and Information Technology, Wasit University, Wasit, 52001, Iraq
[3]Department of Computer Science and Artificial Intelligence, College of Computing and Information Technology, University of Bisha, Bisha, 67714, Saudi Arabia
[4]Electronic Engineering and Information Science Department, University of Science and Technology of China, Hefei, 230026, China

*Corresponding Authors: Zaid Derea. Email: zabdulameer@uowasit.edu.iq; Xiaoyan Kui. Email: xykui@csu.edu.cn

Received: 26 December 2024; Accepted: 02 January 2025; Published: 26 March 2025

ABSTRACT: Improving website security to prevent malicious online activities is crucial, and CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) has emerged as a key strategy for distinguishing human users from automated bots. Text-based CAPTCHAs, designed to be easily decipherable by humans yet challenging for machines, are a common form of this verification. However, advancements in deep learning have facilitated the creation of models adept at recognizing these text-based CAPTCHAs with surprising efficiency. In our comprehensive investigation into CAPTCHA recognition, we have tailored the renowned UpDown image captioning model specifically for this purpose. Our approach innovatively combines an encoder to extract both global and local features, significantly boosting the model's capability to identify complex details within CAPTCHA images. For the decoding phase, we have adopted a refined attention mechanism, integrating enhanced visual attention with dual layers of Long Short-Term Memory (LSTM) networks to elevate CAPTCHA recognition accuracy. Our rigorous testing across four varied datasets, including those from Weibo, BoC, Gregwar, and Captcha 0.3, demonstrates the versatility and effectiveness of our method. The results not only highlight the efficiency of our approach but also offer profound insights into its applicability across different CAPTCHA types, contributing to a deeper understanding of CAPTCHA recognition technology.

KEYWORDS: Text-based CAPTCHA recognition; refined visual attention; web security; computer vision

1 Introduction

CAPTCHAs, or Completely Automated Public Turing tests to tell Computers and Humans Apart, represent a critical component of internet security, strategically crafted to discern between genuine human users and automated bots [1–3]. With their diverse formats spanning text, image, audio, and video, these challenges serve a unified purpose of fortifying online defenses against malicious entities [4–7]. Among the various iterations, text-based CAPTCHAs have emerged as a widely adopted and effective means of safeguarding online platforms [8,9]. Their simplicity and ease of implementation make them accessible across a broad spectrum of applications. However, the relentless progress of technology, particularly in deep learning algorithms [10,11], has posed significant challenges to the security of traditional CAPTCHAs.

As machine learning techniques become increasingly sophisticated, CAPTCHA systems must evolve to stay ahead of emerging threats. Adversarial attacks and advanced algorithms have led to the development of more resilient CAPTCHA designs, incorporating elements like distortion, rotation, and context-based

challenges to thwart automated recognition. Despite these advancements, the arms race between CAPTCHA developers and attackers persists, underscoring the ongoing need for innovation in online security measures. As the digital landscape continues to evolve, CAPTCHA technology must adapt to maintain its efficacy in safeguarding against automated threats.

Deep learning's remarkable ability in feature extraction has led to significant advancements across various domains, including image restoration and object detection [12–16], rendering it an attractive option for robust CAPTCHA recognition systems. However, this capability poses challenges for text-based CAPTCHAs, as traditional methods struggle with feature extraction and are susceptible to image noise. Consequently, there is a growing trend towards employing deep learning paradigms for CAPTCHA recognition, with approaches falling into segmentation-based and segmentation-free categories [17,18]. Segmentation-based methods involve character dissection followed by recognition but often encounter efficiency and efficacy issues. In contrast, segmentation-free algorithms directly recognize and classify CAPTCHA characters without segmentation, showcasing promising accuracy and efficiency.

This work pioneers a novel approach to CAPTCHA recognition, departing from conventional segmentation-based models to introduce an innovative deep-learning-based system. Rather than relying on segmentation, the proposed system leverages image captioning, a burgeoning subfield that seamlessly integrates computer vision and natural language processing (NLP), to redefine the CAPTCHA recognition paradigm. Image captioning empowers artificial intelligence with the remarkable ability to comprehend and describe complex visual scenes in natural language. This task requires not only robust computer vision capabilities but also a deep semantic understanding. This research is centered on the integration of image captioning techniques, incorporating both bottom-up and top-down attention mechanisms, into the field of CAPTCHA recognition. This groundbreaking approach harnesses the power of human-like image understanding and natural language generation to recognize CAPTCHAs that are not only robust but also more secure. The paper explores the concept of employing bottom-up and top-down attention mechanisms in image captioning for CAPTCHA recognition, investigating its potential implications for enhancing online security.

In this paper, we propose a novel CAPTCHA recognition system named Up Down based CAPTCHA Recognition with Refined Visual Attention (CRRVA) model, as it is illustrated in Fig. 1. Our model introduces a specialized recurrent neural network (RNN) architecture, which builds upon the established UpDown model [19] commonly used in image captioning tasks. This adaptation is specifically tailored for the complex task of CAPTCHA recognition. By incorporating convolutional neural networks (CNNs) into our framework, we aim to extract both global and local features from CAPTCHA images, thereby enhancing the model's ability to discern subtle details. Moreover, instead of using the conventional visual attention mechanism for attending to the visual features of the CAPTCHA image, we harness a more enhanced and refined visual attention mechanism [20], which shows a better ability to capture local regions of the CAPTCHA image. We also integrate dual layers of Long Short-Term Memory (LSTM) networks in the decoder part of the model to enhance the prediction of the CAPTCHA characters. This strategic enhancement enables further refinement and optimization of CAPTCHA recognition performance, ensuring improved accuracy and robustness in deciphering CAPTCHA images.

The comprehensive evaluation covers datasets sourced from prominent platforms, including Bank of China (BoC), Weibo, Gregwar, and Captcha 0.3, representing a diverse range of CAPTCHA schemes. These datasets comprise manually collected BoC CAPTCHAs from the official Bank of China website, Weibo CAPTCHAs from the popular Chinese social media platform Weibo, and CAPTCHAs generated using the Gregwar CAPTCHA library and Captcha 0.3. Our meticulous analysis demonstrates remarkable success rates in defeating targeted CAPTCHAs, achieved without the need for segmentation. Specifically, we achieve a

success rate of 96.89% on BoC CAPTCHAs and 95.05% on Weibo CAPTCHAs. Importantly, the applicability of our proposed algorithm extends beyond CAPTCHA recognition, offering potential utility in domains characterized by similar dataset structures. In essence, this paper presents a pioneering deep-learning-driven CAPTCHA recognition system characterized by its efficiency and minimal complexity, eliminating the need for segmentation. By harnessing the power of image captioning methodologies, it achieves remarkable success in decoding text-based CAPTCHAs. The beauty of our approach lies in its simplicity and versatility, positioning it as a promising solution for bolstering online security. Our contributions span theoretical advancements in CAPTCHA development and tangible applications aimed at fortifying internet security.

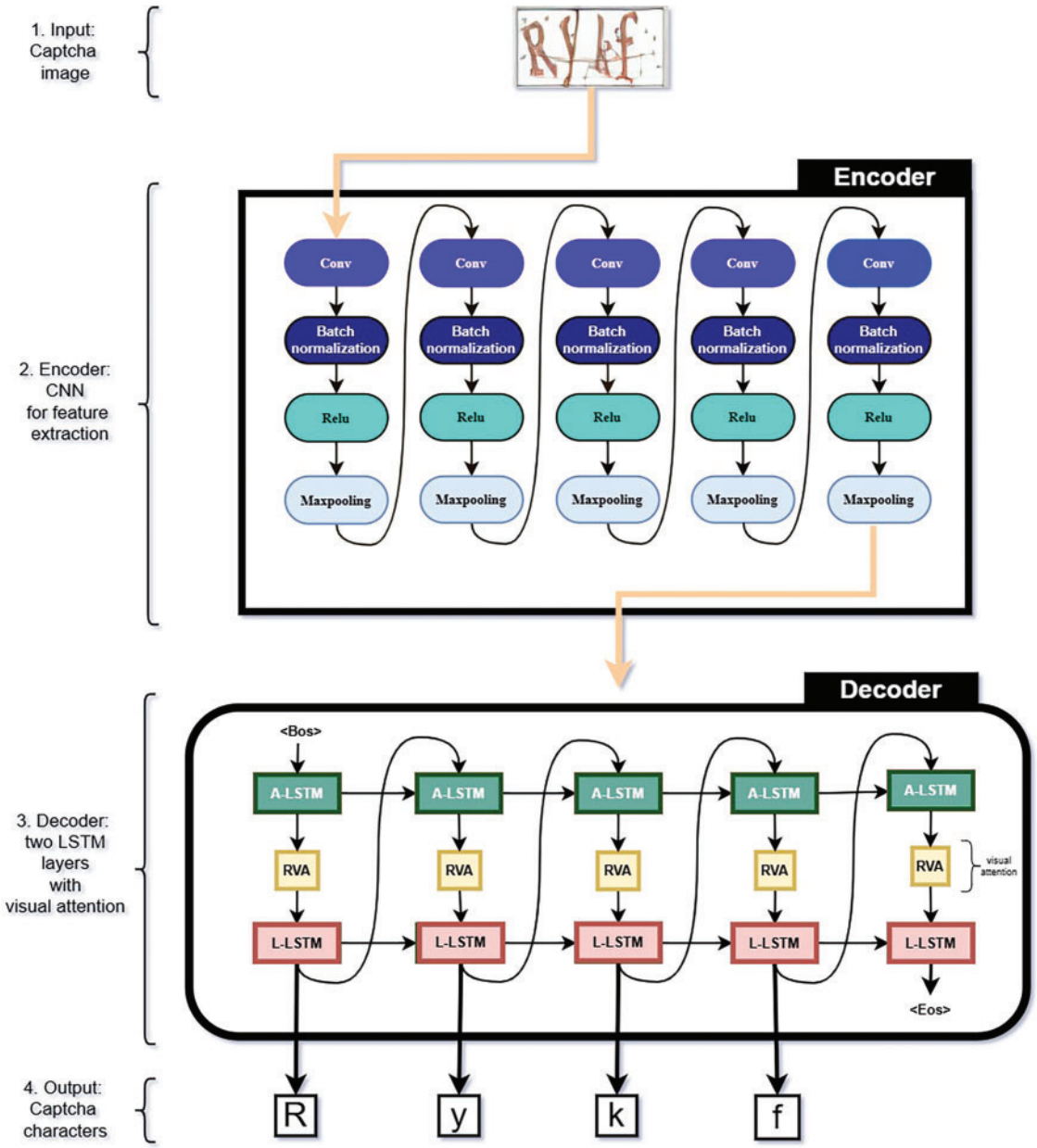

Figure 1: The diagram illustrates the entire process flow of the proposed CAPTCHA recognition framework

The proposed CAPTCHA recognition system effectively addresses key challenges in traditional CAPTCHA systems, including character overlapping and distortion, which are increasingly vulnerable to advanced deep learning attacks. By integrating a refined visual attention mechanism and leveraging a tailored UpDown model, this system enhances focus on relevant image regions, eliminates the need for segmentation, and improves recognition accuracy across various complex CAPTCHA schemes. This approach represents a significant advancement over existing methods, offering a robust and versatile solution to bolster online security against automated threats. This paper presents several key contributions:

- Introducing an innovative model for CAPTCHA recognition designed to merge the realms of image captioning and CAPTCHA recognition into a cohesive framework.
- We unveil the integration of a refined visual attention module, significantly boosting the model's proficiency in concentrating on visual representations.
- Our methodology employs CNNs to extract both global and local features, complemented by an advanced visual attention mechanism. Additionally, we integrate two layers of LSTM to further enhance CAPTCHA recognition capabilities.
- Our research entails thorough experiments conducted across four diverse dataset schemes: Weibo, BoC, Gregwar, and Captcha 0.3, yielding competitive outcomes.

2 Related Work

The concept of the Automated Turing Test has remained a topic of extensive discussion over the years. One of the notable achievements of Alta Vista was its introduction of a method for selectively limiting access to computer systems. This marked a significant milestone as the first practical implementation of a CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) designed to thwart automated bot registrations on web pages. Subsequently, the field of CAPTCHA research witnessed a proliferation of related works. While CAPTCHA exists in different forms, including image-based, text-based, and audio-based CAPTCHAs [3,21], text CAPTCHA is the most prevalent and valuable.

At their inception, text CAPTCHAs employed a simple design featuring alphanumeric characters against a plain background. However, these early text CAPTCHAs quickly fell victim to commonplace machine learning techniques like Support Vector Machines (SVM) and Optical Character Recognition (OCR) technology. An innovative approach emerged as [22] introduced the shape context matching method, achieving success rates of 33% and 92% on EZ-Gimpy and Gimpy schemes, respectively. Chellapilla et al. [23] had successfully circumvented a variety of Human Interactive Proofs (HIPs) and highlighted the significance of segmentation in confounding machine learning algorithms. Consequently, measures were taken to enhance segmentation resistance in text CAPTCHAs, such as character overlapping and character sticking.

However, Yan et al. [24] developed a character segmentation technique with broad applicability, strategically targeting multiple text-based schemes utilized by major companies such as Yahoo!, Microsoft, and Google. Their approach yielded an overall success rate surpassing 60%. This innovative method revolutionized CAPTCHA-breaking techniques, notably adopting a pixel-counting approach to overcome most Captchaservice.org schemes, boasting an exceptionally high success rate of nearly 100%. Moreover, their relentless pursuit of advancements in CAPTCHA recognition persisted. For instance, they successfully tackled a scheme reliant on a novel segmentation-resistant mechanism [25], achieving an impressive success rate of 78%.

Adopting a novel approach, certain websites like Yahoo! implemented a unique method that involves the use of contour lines to create interconnected hollow characters, referred to as hollow CAPTCHAs. Gao et al. managed to circumvent these hollow CAPTCHAs by employing character reconstruction and CNN (Convolutional Neural Network) based recognition techniques [11]. Following this, Microsoft introduced

a dual-layer CAPTCHA system, which consisted of two rows of either English letters or digits. Gao et al. effectively breached this system as well, utilizing deep learning strategies and layer segmentation to achieve a success rate of 44.6% [26]. Despite the introduction of advanced text CAPTCHA resistance methods, including noise arcs, distortion, hollow designs, and multi-layered formats, it becomes clear that current text CAPTCHA algorithms are vulnerable to being decoded with significant accuracy through the use of neural networks, such as RNNs and CNNs.

Thobhani et al. [27] introduced a novel approach leveraging Convolutional Neural Networks (CNNs) with binary images, showcasing remarkable accuracy and a reduction in system size. However, this method encounters challenges in CAPTCHA recognition, particularly due to the substantial requirement for annotated training data. In a different vein, Derea et al. [28] presented the CRNGS algorithm, which combines deep learning with character grouping to facilitate CAPTCHA recognition without the need for segmentation. This approach employs adjustable softmax layers to fine-tune performance across various CAPTCHA schemes. Khatavkar et al. [29] developed a segmentation-free OCR model that applies the Connectionist Temporal Classification (CTC) loss technique for efficient text CAPTCHA classification. Meanwhile, Chang et al. [30] conducted an in-depth analysis of security vulnerabilities in behavior-verification-based slider CAPTCHAs, a relatively unexplored domain, proposing a universal methodology for identifying target trajectories and simulating user actions with high accuracy.

Anderson et al. [19] pioneered an innovative approach that merges bottom-up and top-down attention mechanisms, significantly improving visual question-answering and image captioning systems. This methodology aims to mimic human image interpretation and question-answering processes, drawing inspiration from human visual perception. Subsequent research has widely adopted [19] for image captioning enhancements. For instance, Huang et al. [31] showcased how integrating multimodal attribute detectors during the training phase of image captioning models can augment the descriptiveness and relevance of the generated captions. Moreover, the incorporation of topics extracted from caption corpora into captioning tasks has been shown to influence sentence generation significantly [32]. Hossen et al. [33] introduced a Guided Visual Attention (GVA) technique for enhancing the creation of image captions. This method improves the quality of captions by refining the distribution of attentional focus.

In the work by [20], the key innovation lies in the integration of a refined visual attention (RVA) framework into the image captioning process. Unlike traditional visual attention mechanisms, which may suffer from decreased effectiveness due to small numerical values in their inputs, RVA dynamically adjusts visual attention weights based on the language context of previously generated words. This adjustment is achieved by using a fully connected layer to adapt the dimensions of visual features and applying a sigmoid function to obtain a probability distribution for reweighting the input to softmax.

In our proposed approach, we integrate innovative image captioning techniques, encompassing both bottom-up and top-down attention mechanisms and refined visual attention with CAPTCHA recognition. This method harnesses human-like image comprehension and natural language generation to recognize CAPTCHAs. This article delves into the utilization of these attention mechanisms in CAPTCHA recognition to augment online security.

3 Methodology

In our thorough exploration of CAPTCHA recognition, we utilize a customized RNN network based on the well-known UpDown model, typically used in image captioning, tailored specifically for CAPTCHA recognition. Our novel method incorporates CNNs to extract comprehensive global and local features, thereby enhancing the model's capability to detect fine details within CAPTCHA images. We employ dual

layers of LSTM networks along with a refined visual attention mechanism to further enhance and refine CAPTCHA recognition accuracy. The schematic overview of our proposed model is illustrated in Fig. 1.

3.1 Visual Features of Input Image

In our CAPTCHA recognition system, the first step is to extract the visual features from the input image, allowing for their subsequent utilization in the language model's processing. The initial stage of recognizing the CAPTCHA for an image revolves around obtaining the visual representations of the input image. This is accomplished by employing a CNN in the encoder part of our model to generate a set of image features V crucial for CAPTCHA recognition. Specifically, we extract these features from the final output of the max-pooling layer in the CNN model, which comprises five convolutional layers and five max-pooling layers. The adopted activation function is ReLU and batch normalization is applied. The architecture of the adopted CNN is illustrated in Fig. 2. The CNN model produces N extracted visual features vectors, comprising the visual matrix $V \in \mathbb{R}^{N \times h}$. Subsequently, each of the extracted visual vectors is denoted as $v_i \in \mathbb{R}^h$ and $i \in \{1, 2, \ldots, N\}$. The obtained visual matrix V from the input image I through the CNN network can be described by:

$$V = CNN(I) \tag{1}$$

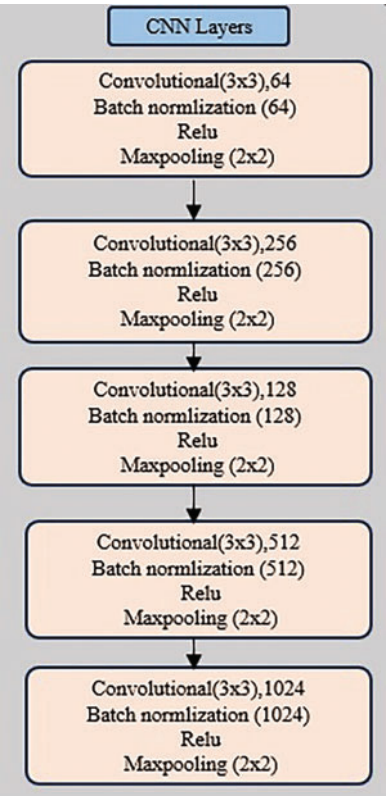

Figure 2: The architecture of the adopted CNN of our model

Also the visual vectors matrix V can be expressed as a collection of the N individual visual vectors v_i as follows:

$$V = \{v_1, v_2, \ldots, v_N\} \tag{2}$$

The mean of the extracted visual vectors representing the global visual features is denoted by $\bar{v} \in \mathbb{R}^h$ and is obtained using the following formula:

$$\bar{v} = \frac{1}{N} \sum_{i=1}^{N} v_i \tag{3}$$

3.2 Refined Visual Attention (RVA)

In our CAPTCHA recognition system, we have adopted the refined visual attention module for attending to the local visual features of the input image [20]. This new attention module provides more influence to the hidden state of the attention LSTM on the visual attention. This is achieved by adding an additional linear layer followed by a sigmoid function to the traditional visual attention module, resulting in better concentration on the relevant image's regions in the current time step. The refined visual attention module is depicted in Fig. 3 and is described with the following formulas:

$$\alpha_t^i = W_c \cdot \tanh\left(h_t^a \cdot W_a + W_b \cdot v_i\right) \tag{4}$$

$$x_t = W_d \cdot h_t^a \tag{5}$$

$$\bar{x}_t = \text{sigmoid}(x_t) \tag{6}$$

$$\hat{x}_t = \alpha_t * \bar{x}_t \tag{7}$$

$$\beta_t = \text{softmax}\left(\hat{x}_t\right) \tag{8}$$

$$\hat{v}_t = \sum_{i=1}^{N} v_i \odot \beta_t^i \tag{9}$$

where $W_a \in R^{g \times e}$, $W_b \in R^{h \times e}$, $W_c \in R^e$, and $W_d \in R^{g \times N}$ are trainable weights. $h_t^a \in R^g$, $\alpha_t \in R^N$, $\beta_t \in R^N$, and $x_t \in R^N$. The $*$ symbol refers to the element-wise multiplication operation. $\hat{v}_t \in R^h$ signifies the resulted context vector.

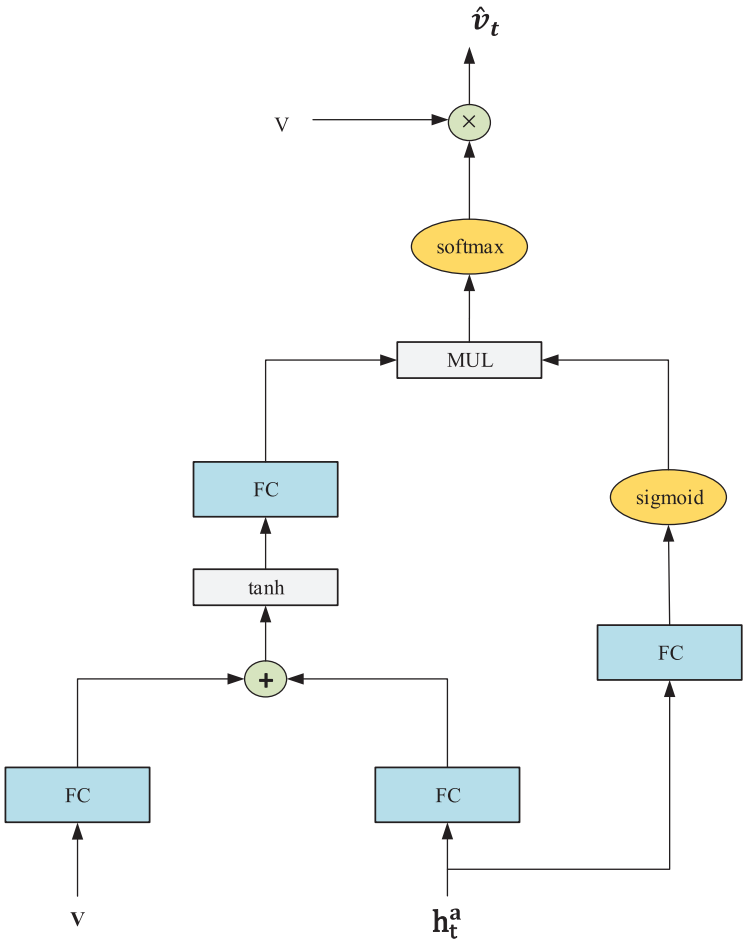

Figure 3: The diagram depicts the internal structure of the refined visual attention

3.3 Language Model

The architecture of our language model is depicted in Fig. 4. Our approach utilizes the UpDown architecture as its core framework, which incorporates two LSTM layers: one for language processing, referred to as $LSTM_l$, and another for attention, referred to as $LSTM_a$. The hidden states for both the attention LSTM, represented as $h_t^a \in \mathbb{R}^g$, and the language LSTM, represented as $h_t^l \in \mathbb{R}^g$, are detailed in the equations that follow:

$$h_t^{\mathrm{a}} = LSTM_a\left(h_{t-1}^{\mathrm{a}}; \left[h_{t-1}^{\mathrm{l}}, \bar{v}, E \cdot y_{t-1}\right]\right) \tag{10}$$

$$h_t^{\mathrm{l}} = LSTM_l\left(h_{t-1}^{\mathrm{l}}; \left[\hat{v}_t, h_t^{\mathrm{a}}\right]\right) \tag{11}$$

where $E \in \mathbb{R}^{c \times f}$ represents the matrix of character embeddings and $y_{t-1} \in \mathbb{R}^c$ indicates the character generated at the previous timestep. Subsequently, the output from the language LSTM, h_t^l, is processed through a linear layer followed by a softmax activation layer, resulting in a probability distribution over all CAPTCHA characters, given as:

$$p_t = \mathrm{softmax}\left(W_p \cdot h_t^l\right) \tag{12}$$

where $p_t \in \mathbb{R}^c$ and $W_p \in \mathbb{R}^{g \times c}$.

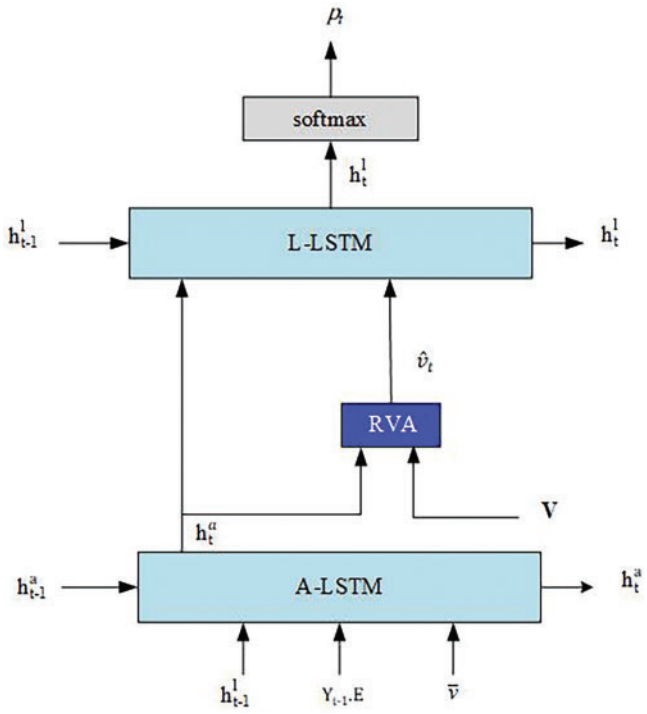

Figure 4: The diagram illustrates the internal architecture of the language model. A-LSTM refers to the attention LSTM while the L-LSTM refers to the language LSTM

3.4 Loss Functions

We utilize cross-entropy (XE) for training our model, employing the standard cross-entropy loss $Loss_{XE}$, which is defined as:

$$Loss_{XE} = \frac{1}{T}\sum_{t=1}^{T} -\log\left(p_t\left(y_t \mid y_{1:t-1}, I\right)\right) \tag{13}$$

4 Experiments and Results

In this section, we furnish detailed insights into the datasets utilized for training, validating, and evaluating the CRRVA model. After delving into the datasets, we provide an exhaustive description of the CNN architecture and attention mechanism employed within the CRRVA model, alongside a comprehensive overview of the training parameters applied. Subsequent to setting up the model, we meticulously evaluate its accuracy using various metrics. Furthermore, we conduct an extensive comparative analysis, juxtaposing the performance of the CRRVA algorithm with that of other existing CAPTCHA recognition systems. This comparative scrutiny not only underscores the strengths of the CRRVA model but also pinpoints areas for potential enhancement, thus offering valuable insights into its overall effectiveness and limitations.

4.1 Used Datasets

In order to tackle CAPTCHA recognition challenges, it's essential to source CAPTCHA images due to the limited availability of comprehensive CAPTCHA datasets. To address this, we collect CAPTCHA

images through two primary methods: first, by extracting them from real-time online sources, and second, by utilizing CAPTCHA generation software. Our research involves the utilization of four distinct schemes of CAPTCHA datasets, namely, Bank of China (BoC) (accessible at https://ebsnew.boc.cn/boc15/login.html, accessed on 29 February 2024), Weibo (available at https://www.weibo.com, accessed on 02 February 2023), Captcha 0.3, and Gregwar. Fig. 5 illustrates examples from each of these four CAPTCHA schemes, showcasing the diversity in the datasets we have compiled for our experiments.

Figure 5: Illustrative examples of CAPTCHA schemes employed within the CRRVA model

4.1.1 Bank of China CAPTCHA Dataset

With a vast network of over 10,000 branches spanning the globe, the Bank of China holds a prominent position as one of China's foremost commercial banks. Recognizing the escalating threats posed by automated attacks, the bank has proactively implemented a robust CAPTCHA system to bolster its online security measures. This advanced CAPTCHA system is meticulously designed, incorporating sophisticated elements such as character overlap, distortion, warping, and rotation. By presenting CAPTCHAs consisting of four characters, comprising a mix of uppercase English letters and numerical digits, the Bank of China aims to create formidable barriers against automated bots attempting to infiltrate its online platforms. Our meticulously compiled dataset comprises a staggering 70,000 diverse CAPTCHA images sourced directly from the Bank of China's CAPTCHA system. This comprehensive dataset not only reflects the institution's steadfast commitment to cybersecurity but also serves as a testament to its proactive approach to safeguarding sensitive financial information. It's noteworthy that specific characters, such as G, C, Q, I, O, L, S, 0, 1, and 5, have been deliberately excluded from these CAPTCHAs to heighten security protocols and mitigate the risk of automated attacks. By providing researchers and cybersecurity experts with access to this extensive dataset, the Bank of China not only facilitates the development and refinement of advanced CAPTCHA recognition algorithms but also contributes significantly to ongoing efforts aimed at fortifying online security measures on a global scale. This initiative underscores the bank's dedication to staying ahead of emerging threats and ensuring the utmost protection of confidential financial data across digital platforms worldwide.

4.1.2 Weibo CAPTCHA Dataset

Weibo, standing as one of China's foremost social media platforms, commands a staggering monthly active user base of 586 million as of 2022, solidifying its position as a dominant force in the digital landscape. With a distinguished ranking by Alexa, Weibo underscores its commitment to online security through the implementation of a rigorous CAPTCHA system. This CAPTCHA system, designed with meticulous attention to detail, incorporates a range of protective measures, including character overlapping, distortion, warping, and rotation. Such measures are strategically employed to fortify the security of user accounts, thwarting malicious attempts at unauthorized access. Typically presenting users with CAPTCHAs comprising four characters, encompassing both numeric digits and uppercase English letters, Weibo's system intentionally omits specific characters such as I, D, G, U, Q, 0, 1, and 5. This deliberate exclusion serves to enhance security protocols, making it more challenging for automated bots to bypass the authentication process. Our dataset, meticulously curated and comprising 70,000 Weibo CAPTCHA images, serves as a testament to Weibo's unwavering dedication to cybersecurity and user privacy protection. Each image within the dataset has been carefully labeled to facilitate comprehensive analysis, providing valuable insights into the efficacy of Weibo's CAPTCHA system and aiding in the development of advanced security algorithms. Obtained through a rigorous manual collection process, this dataset underscores Weibo's proactive approach to bolstering online security measures.

4.1.3 Captcha 0.3 CAPTCHA Dataset

In its open-source nature, Captcha 0.3 enables users to easily generate CAPTCHAs. In our selection of CAPTCHA images, we've opted for a four-character arrangement, encompassing a wide spectrum of possibilities—these characters can be numeric, spanning from 0 to 9, or uppercase and lowercase English letters, ranging from A to Z and a to z, respectively. Our meticulous efforts have yielded a comprehensive dataset comprising a substantial 70,000 CAPTCHA images. Each of these CAPTCHA images was meticulously generated with randomness as a core principle, assuring the absence of any repetition or duplication

within the dataset. Elevating the level of security intrinsic to these CAPTCHAs, we've introduced elements such as intersecting lines across the characters and the inclusion of noisy dots scattered throughout the background. To craft these CAPTCHA images, we've opted for the "liberbaskerville-regular" font. To provide a tangible glimpse into the CAPTCHA schemes within our dataset, Fig. 5 showcases samples of these CAPTCHA images.

4.1.4 Gregwar CAPTCHA Dataset

The Gregwar library in PHP stands out for its efficient CAPTCHA generation capabilities, offering a robust defense against automated bot attacks. By incorporating a range of security enhancements, such as intricate noise lines, elegant backgrounds, and rotational elements, Gregwar CAPTCHAs are meticulously crafted to withstand even the most determined bot attacks. Comprising four characters drawn from three distinct character classes—numeric digits, uppercase English letters, and lowercase English letters—these CAPTCHAs provide a formidable challenge to automated bots. Our dataset is the result of meticulous creation, encompassing a vast set of 70,000 unique Gregwar CAPTCHA images. During the generation process, each CAPTCHA's four characters were meticulously selected at random, ensuring a dataset free from duplications or repetitions. This commitment to randomness and diversity enhances the dataset's effectiveness and security, reinforcing its utility for CAPTCHA recognition algorithms and security assessments.

4.1.5 Preprocessing Steps

Each dataset of CAPTCHAs is meticulously organized into three distinct subsets: a robust training set comprising 50,000 diverse CAPTCHA images, a meticulously curated testing set containing 10,000 carefully selected CAPTCHA images, and a comprehensive validation set encompassing 10,000 meticulously chosen CAPTCHA images. Embedded within the filename of each CAPTCHA image across the four dataset schemes lies a label representing a unique four-character string extracted from the CAPTCHA itself.

Initially, a meticulous preprocessing step involves converting all CAPTCHA images to grayscale and resizing them to uniform dimensions of 64×256 to ensure consistency across the dataset. It is imperative to note that the selection process for images within the training, validation, and testing sets is conducted randomly, ensuring the elimination of any potential biases and maintaining the integrity of the dataset for accurate evaluation and training purposes.

4.2 Experimental Settings

In this work, we employ the CNN architecture based on Fig. 2 to extract features from images, yielding object features with dimensions of 16×1024. We utilize 1000-dimensional character embeddings and set the hidden state length for both LSTM networks to 1000.

In our CAPTCHA Recognition model, a visual features vector of size $h = 1024$ is utilized to represent the input image. The LSTM hidden state size $g = 1000$ facilitates capturing intricate linguistic patterns during CAPTCHA generation. Each character is mapped to a vector of length $f = 1000$. Information is extracted from $N = 16$ visual feature vectors associated with the image. Moreover, we used internal hidden attention with size $e = 512$ to improve the model's ability to focus on relevant image parts during CAPTCHA generation. c represents the number of characters adopted in a CAPTCHA scheme, which varies from one CAPTCHA scheme to another. Specifically, c is 28, 26, 62, and 62 in Weibo, BoC, Gregwar, and Captcha 0.3 CAPTCHA schemes, respectively.

For training our CAPTCHA recognition networks, we utilize the Adam optimizer and conduct 120 epochs of training. The initial learning rate is set to 0.0005 and decreases by a factor of 0.8 every 5 epochs during training. The batch size remains fixed at 50, with scheduled sampling increasing by 5% every 5 epochs during training until reaching 25%. Gradients are clipped to a maximum absolute value of 0.1, and a dropout ratio of 0.5 is employed in our model. Testing employs a beam size of 3 and a beam search strategy, and the networks are implemented using the PyTorch framework.

4.3 Model Accuracy

Table 1 provides a comprehensive overview of the CAPTCHA accuracies achieved by the CRRVA model across various CAPTCHA schemes, including Weibo, BoC, Gregwar, and Captcha 0.3. Notably, the CRRVA model demonstrated remarkable performance in the BoC CAPTCHA scheme, achieving an accuracy of 96.89%, correctly identifying 9689 out of 10,000 images. This underscores its efficacy in handling the complexities of BoC CAPTCHAs.

Table 1: The four CAPTCHA schemes, i.e., Gregwar, Captcha 0.3, Weibo, and BoC, used to evaluate individual character accuracy and overall CAPTCHA accuracy for CRRVA

	Gregwar	**Captcha 0.3**	**Weibo**	**BoC**
1st character	84.72%	99.25%	98.68%	98.95%
Accuracy	(8472/10,000)	(9925/10,000)	(9868/10,000)	(9895/10,000)
2nd character	78.04%	97.89%	98.18%	99.17%
Accuracy	(7804/10,000)	(9789/10,000)	(9818/10,000)	(9917/10,000)
3rd character	76.62%	97.63%	98.02%	99.24%
Accuracy	(7662/10,000)	(9763/10,000)	(9802/10,000)	(9924/10,000)
4th character	84.72%	98.90%	98.64%	99.29%
Accuracy	(8472/10,000)	(9890/10,000)	(9864/10,000)	(9,929/10,000)
Total character	81.02%	98.41%	98.38%	99.16%
Accuracy	(32,410/40,000)	(39,367/40,000)	(39,352/10,000)	(39,665/40,000)
Overall CAPTCHA	47.08%	94.78%	95.05%	96.89%
Accuracy	(4708/10,000)	(9478/10,000)	(9505/10,000)	(9689/10,000)

In the Weibo CAPTCHA scheme, the CRRVA model achieved an accuracy of 95.05%, accurately recognizing 9505 out of 10,000 images. This further solidifies its robustness across diverse CAPTCHA formats. Similarly, in the Captcha 0.3 scheme, the CRRVA model showcased a high recognition accuracy of 94.78%, correctly identifying 9478 images out of 10,000. This consistent performance highlights the model's reliability across different CAPTCHA variations. However, the Gregwar CAPTCHA scheme presented a notable challenge, with the CRRVA model achieving an accuracy of only 47.08%, correctly recognizing 4708 images out of 10,000. Despite this lower accuracy rate, the model's ability to decipher a significant portion of Gregwar CAPTCHAs indicates its potential for improvement and adaptation to more intricate CAPTCHA designs.

Transitioning to overall character accuracy, the CRRVA model showcased robust performance across different CAPTCHA schemes. Notably, in the BoC CAPTCHA scheme, an outstanding accuracy of 99.16% was attained, correctly identifying 39,665 out of 40,000 images. This highlights the model's exceptional ability to accurately decipher BoC CAPTCHAs.

Similarly, in the Weibo CAPTCHA scheme, the accuracy reached 98.38%, with 39,352 images recognized correctly out of 40,000. This consistent high performance underscores the model's reliability across diverse CAPTCHA formats. The Captcha 0.3 scheme also demonstrated strong accuracy, achieving a rate of 98.41% with 39,367 images correctly identified out of 40,000. This further validates the model's effectiveness in handling various CAPTCHA variations. Even in the challenging Gregwar CAPTCHA scheme, the model exhibited commendable accuracy, reaching 81.02% and accurately identifying 32,410 images out of 40,000. Despite the complexity inherent in Gregwar CAPTCHAs, the model's ability to decipher a significant portion of them showcases its adaptability and potential for improvement.

For individual character accuracies, the CRRVA model exhibited outstanding performance across all schemes, showcasing its robustness in character recognition tasks. In the BoC CAPTCHA scheme, the model achieved remarkable accuracies for each character: 98.95% for the first character, 99.17% for the second character, 99.24% for the third character, and 99.16% for the fourth character. Similarly, in the Weibo CAPTCHA scheme, the CRRVA model maintained high accuracy rates, with accuracies of 98.68%, 98.18%, 98.02%, and 98.64% for the respective characters. In the Captcha 0.3 scheme, the model demonstrated consistent accuracy, with accuracies of 99.25%, 97.89%, 97.63%, and 98.90% for the first, second, third, and fourth characters, respectively. Despite the challenges posed by the Gregwar CAPTCHA scheme, the CRRVA model showcased notable accuracy rates, achieving 84.72%, 78.04%, 76.62%, and 84.72% for the individual characters.

Figs. 6–9 illustrate the overall character recognition accuracies for the BoC, Weibo, Captcha 0.3, and Gregwar datasets throughout 120 training epochs. These figures provide a comprehensive view of how the model's performance evolves during the training process for different datasets, offering insights into its learning behavior and generalization capability across varied data types.

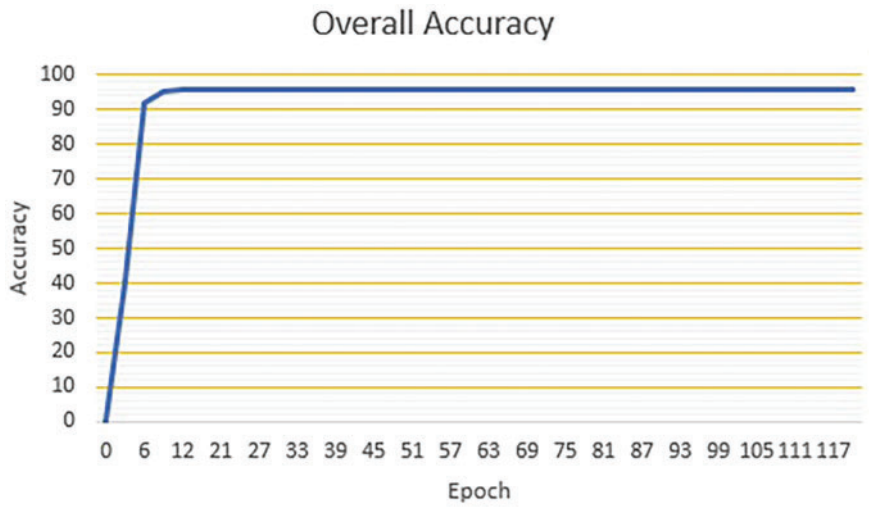

Figure 6: The overall character recognition accuracy of CRRVA on the BoC dataset during the training phase

In Fig. 6, the accuracy curve for the BoC dataset shows a consistent upward trend, indicating a steady improvement in character recognition as training progresses. The accuracy starts at a relatively low level in the initial epochs, reflecting the model's initial phase of learning. As the epochs increase, there is a noticeable increase in accuracy, with the curve becoming smoother and gradually plateauing towards the later epochs. By the end of the training period, the model achieves an impressive accuracy of 96.89%, demonstrating its effectiveness in recognizing characters from the BoC dataset.

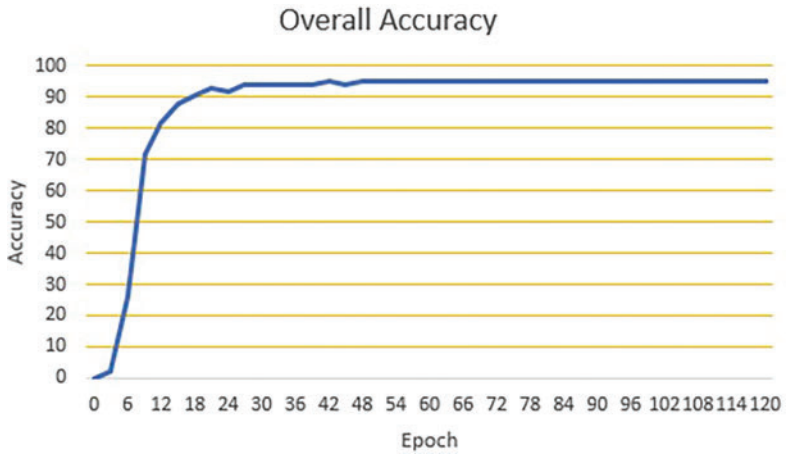

Figure 7: The overall character recognition accuracy of CRRVA on the Weibo dataset during the training phase

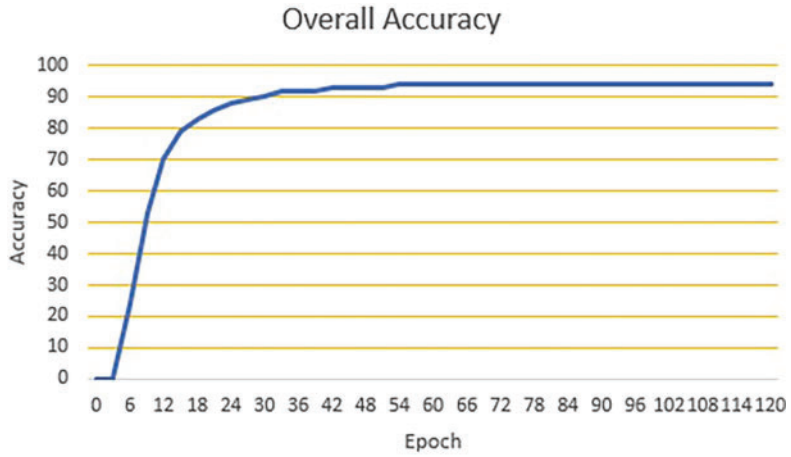

Figure 8: The overall character recognition accuracy of CRRVA on the Captcha 0.3 dataset during the training phase

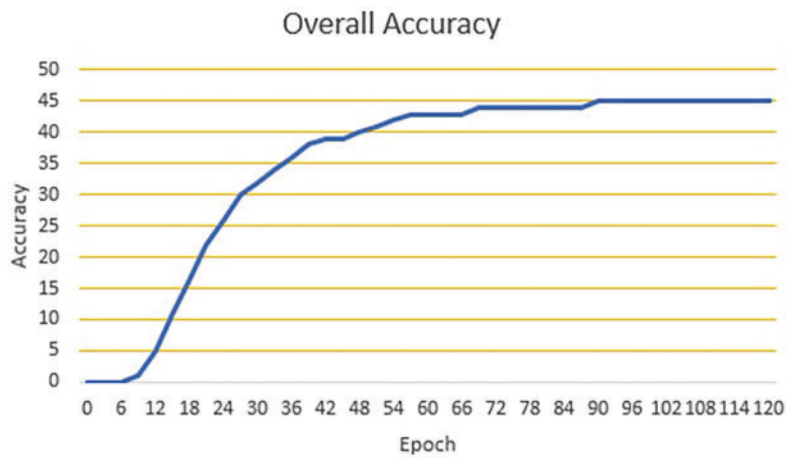

Figure 9: The overall character recognition accuracy of CRRVA on the Gregwar dataset during the training phase

Fig. 7 presents the accuracy progression for the Weibo dataset. Similar to the BoC dataset, the model shows an initial phase of rapid improvement, followed by a more gradual increase in accuracy. The accuracy curve for the Weibo dataset exhibits more fluctuations compared to the BoC dataset, indicating potential variability in the data or additional complexity in recognizing characters. Despite these fluctuations, the overall trend remains positive, with the model achieving a high accuracy of 95.05% by the end of the 120 epochs, reflecting its robust performance on the Weibo dataset.

In Fig. 8, the accuracy curve for the Captcha 0.3 dataset reveals a different pattern. The initial epochs show a slower rate of improvement, suggesting that the model takes longer to learn the distinguishing features of this dataset. However, as training continues, there is a marked increase in accuracy, indicating that the model is eventually able to capture the necessary patterns for effective character recognition. The plateauing of the curve towards the end of the epochs signifies that the model is reaching its peak performance for the Captcha 0.3 dataset. The final accuracy achieved is 94.78%, showcasing the model's strong capability in handling CAPTCHA challenges.

Fig. 9 displays the accuracy for the Gregwar dataset. The accuracy curve here demonstrates a steady and consistent rise throughout the training epochs, similar to the BoC dataset. The relatively smooth and continuous increase in accuracy suggests that the model is efficiently learning from the Gregwar dataset with fewer interruptions or overfitting issues. However, the final accuracy achieved is 47.08%, indicating that this dataset presents significant challenges for the model. The lower accuracy suggests the need for further optimization and potential refinement in handling the complexities of the Gregwar dataset.

These impressive results across diverse CAPTCHA schemes underscore the CRRVA model's ability to accurately identify individual characters, highlighting its effectiveness and versatility in character recognition tasks and emphasizing the efficient role of the RVA module for capturing the CAPTCHA characters' details.

4.4 Comparison Results

To comprehensively assess the strengths and weaknesses of the CRRVA model, we conducted a rigorous comparison with several widely adopted CAPTCHA recognition algorithms, utilizing the same datasets for a fair evaluation. Among the notable contenders in the field of image-text CAPTCHA recognition, we included the Multilabel CNN, CRABI, and CRNN models. The Multilabel CNN architecture employs a single convolutional neural network with multiple softmax output layers, where each layer is responsible for recognizing a different character within the CAPTCHA. In contrast, CRABI simplifies the process by bypassing segmentation. It uses binary images attached to CAPTCHA copies to efficiently locate and recognize characters using a basic CNN, which has a single softmax output layer. Meanwhile, the CRNN combines convolutional layers with recurrent layers with complex and resource-intensive architecture.

The comprehensive comparison results, meticulously documented in Tables 2–5, corresponding to the BoC, Weibo, Captcha 0.3, and Gregwar CAPTCHA schemes, respectively, shed light on the performance of each model across various CAPTCHA scenarios. Our evaluation focused on two critical metrics: total character accuracy and overall CAPTCHA accuracy.

Conversely, the CRABI model emerges as the frontrunner in the Gregwar CAPTCHA scheme, outperforming other contenders, including the CRRVA model. This highlights the diversity and complexity of CAPTCHA structures and underscores the importance of tailoring recognition algorithms to specific CAPTCHA types.

Table 2: Comparison results using BoC CAPTCHA scheme

	CRABI	Multilabel	CRNN	CRRVA
Testing total Character accuracy	98.44% (39,379/40,000)	99.03% (39,614/40,000)	–	99.16% (39,665/40,000)
Testing overall CAPTCHA accuracy	94.33% (9433/10,000)	96.39% (9639/10,000)	96.47% (9647/10,000)	96.89% (9689/10,000)

Table 3: Comparison results using Weibo CAPTCHA scheme

	CRABI	Multilabel	CRNN	CRRVA
Testing total Character accuracy	97.89% (39,156/40,000)	96.03% (38,411/40,000)	–	98.38% (39,352/10,000)
Testing overall CAPTCHA accuracy	92.68% (9268/10,000)	86.24% (8624/10,000)	91.05% (9105/10,000)	95.05% (9505/10,000)

Table 4: Comparison results using Captcha 0.3 CAPTCHA scheme

	CRABI	Multilabel	CRNN	CRRVA
Testing total	96.11%	98.71%	–	98.41%
Character accuracy	(38,444/40,000)	(39,485/40,000)	–	(39,367/40,000)
Testing overall	85.93%	95.33%	83.57%	94.78%
CAPTCHA accuracy	(8593/10,000)	(9533/10,000)	(8357/10,000)	(9478/10,000)

Table 5: Comparison results using Gregwar CAPTCHA scheme

	CRABI	Multilabel	CRNN	CRRVA
Testing total	85.28%	83.31%	–	81.02%
Character accuracy	(34,111/40,000)	(33,322/40,000)		(32,410/40,000)
Testing overall	54.20%	51.23%	49.98%	47.08%
CAPTCHA Accuracy	(5420/10,000)	(5123/10,000)	(4998/10,000)	(4708/10,000)

Analyzing the outcomes, Tables 2 and 3 reveal that the CRRVA model surpasses all other methods in both total character accuracy and overall CAPTCHA accuracy for the BoC and Weibo CAPTCHA schemes. Its exceptional performance underscores its effectiveness in accurately deciphering characters and capturing the overall context of the CAPTCHA images.

While the CRRVA model secures the top position in the BoC and Weibo schemes, it still achieves commendable results in the Captcha 0.3 scheme, ranking second to the Multilabel model. This slight deviation in performance indicates the nuanced nature of CAPTCHA challenges and the need for adaptive models to maintain consistent efficacy across different CAPTCHA designs.

Overall, the comparison results underscore the versatility and robustness of our proposed CRRVA model, which exhibits superior performance across multiple CAPTCHA schemes. This signifies its potential

for widespread adoption in real-world applications where accurate and efficient CAPTCHA recognition is paramount.

5 Qualitative Evaluation

In addition to conducting quantitative score analysis, it is imperative to evaluate the accuracy of the CAPTCHAs generated by CRRVA qualitatively. Fig. 10 showcases a curated selection of sample images from the test dataset, each paired with its corresponding CAPTCHA. Within Fig. 10, every image is matched with a CAPTCHA text generated by our model across the four datasets: BoC, Weibo, Captcha 0.3, and Gregwar. For instance, the image situated at the top-left corner of the first row and column serves as an illustration for the BoC CAPTCHA "PTX7" and the top-right "92FW" serves as an illustration for the Gregwar CAPTCHA. Notably, our CRRVA model adeptly captures the textual content within the image. The performance of our CRRVA model, alongside the quality of its generated CAPTCHAs, remains consistently high. This is evident from the scores detailed in Table 1, as well as exemplified by the sample CAPTCHAs showcased in Fig. 10.

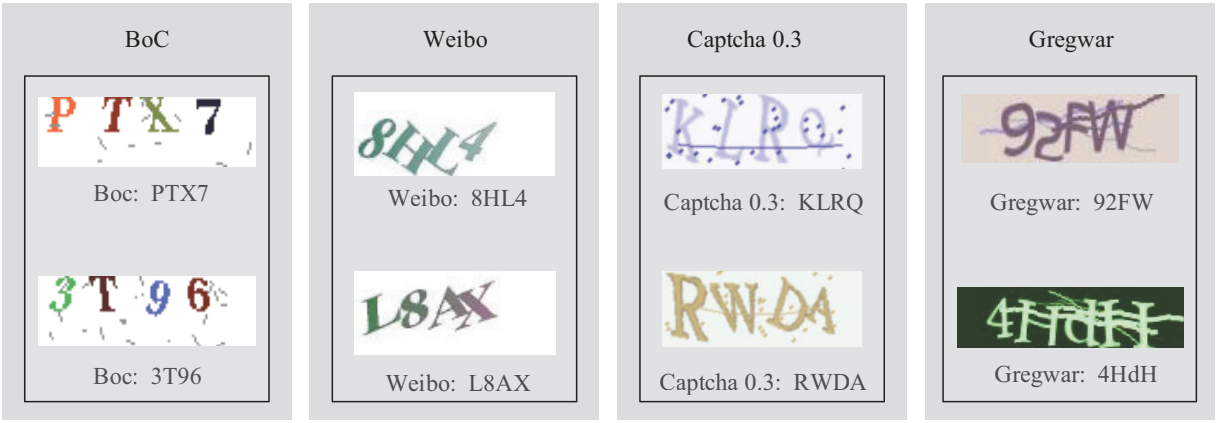

Figure 10: Examples of some CAPTCHAs correctly recognized by CRRVA for the four datasets: BoC, Weibo, Captcha 0.3, and Gregwar

As shown in Fig. 11, some CAPTCHAs were incorrectly recognized. For example, in the CAPTCHA image on the left, the characters "d" and "q" were incorrectly recognized due to the presence of overlapping colored lines intersecting the text and the challenging positioning of the characters. The common failure cases arise from various resistance mechanisms designed to obstruct automated recognition. These mechanisms include distorted and stretched characters, overlaying colored lines intersecting the text, and added background noise, all of which increase the complexity of the CAPTCHA. Furthermore, the characters are positioned closely together with some overlap, complicating the segmentation process. Additionally, the low contrast between the text and the background poses significant challenges for machine-based CAPTCHA solvers.

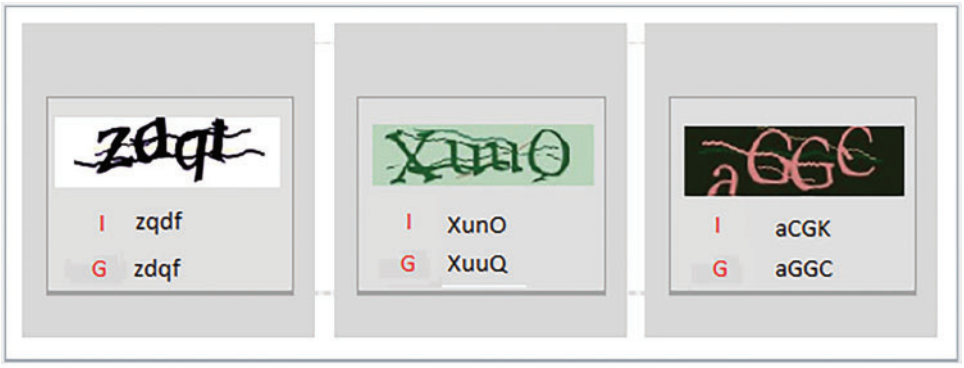

Figure 11: Examples of CAPTCHAs incorrectly recognized by the CRRVA model, with "I" referring to the incorrectly recognized characters and "G" representing the ground truth

6 Discussion

This study presents an innovative approach tailored specifically for the intricate task of CAPTCHA recognition, representing a significant advancement in the field. The incorporation of the RVA attention module into our model boosts the model's ability to capture relevant local visual features and identify CAPTCHA characters. Through meticulous examination and rigorous testing, our proposed algorithm demonstrates exceptional performance across a range of challenging CAPTCHA defense mechanisms. The comprehensive evaluation of our algorithm reveals its robustness in overcoming complex obstacles inherent in CAPTCHA designs, such as character overlapping, noise lines, rotations, distortions, and diverse color backgrounds. Moreover, our algorithm showcases its adaptability in handling multiple CAPTCHA character categories, further enhancing its versatility and applicability. Of particular note is the resilience demonstrated by our algorithm when confronted with the formidable defense mechanisms of the Gregwar CAPTCHA scheme. This scheme, renowned for its stringent security measures, poses significant challenges to traditional CAPTCHA recognition methods. However, our algorithm rises to the occasion, exhibiting remarkable accuracy and efficacy in deciphering even the most intricate Gregwar CAPTCHAs. Furthermore, our approach eliminates the need for the cumbersome process of segmenting CAPTCHA images into individual characters, streamlining the recognition process and significantly improving accuracy and efficiency. By pushing the boundaries of CAPTCHA recognition technology, our work not only advances state-of-the-art technology but also holds significant implications for enhancing online security measures. The robustness and versatility of our algorithm pave the way for more secure online platforms, safeguarding against automated threats and ensuring the integrity of digital interactions.

Notably, during the implementation of visual attention mechanisms in CAPTCHA recognition, we encountered several challenges. One of the primary challenges was ensuring the model's ability to effectively utilize character context to recognize overlapping characters. This aspect is crucial for the model to focus on the most relevant features for accurate character recognition. To address this, we enhanced the attention mechanism by incorporating a sigmoid layer that significantly amplifies the role of character context in the attention process. Specifically, we added a linear layer followed by a sigmoid function, allowing us to reweight the attention based on the current character context. This refinement involved multiplying the adjusted attention output with the standard attention mechanism, giving more weight to features closely related to the character context. This strategic enhancement has improved the model's ability to recognize overlapping characters and better understand the character context, leading to higher accuracy in CAPTCHA

recognition. By emphasizing the character context in the attention process, we ensured that the model focuses more precisely on the most relevant features, thereby effectively overcoming the challenges.

In comparison to other models like the CRABI, which utilizes Attached Binary Image (ABI) technology, our CRRVA model offers several advantages. The CRABI model, while having a straightforward structure, presents notable deficiencies. It requires a preprocessing step where multiple copies of input images must be created and processed sequentially for recognition. This approach complicates the input handling process and significantly prolongs the training time. Additionally, the multilabel CNN approach becomes less effective as the model size increases in proportion to the number of characters, leading to significant scalability issues. On the other hand, the CRNN model involves numerous hyperparameters that require careful tuning, adding complexity to the model's design. The convolutional layers within CRNN demand specific adjustments to work effectively, further complicating the overall architecture. In contrast, our CRRVA model offers several key advantages. It does not rely on the number of characters in the image, thus sidestepping the limitations faced by other models that require preprocessing steps, longer training time, or complicated architecture. Also, the number of hyperparameters in our model is considered relatively few. The architecture of our CNN is inherently flexible and seamlessly integrates with the RNN layer, allowing for efficient processing without the need for extensive preprocessing steps. A significant strength of our approach is its novelty; we have pioneered the use of an image captioning model for CAPTCHA recognition, aiming to enhance the performance and accuracy of CAPTCHA recognition systems.

Our approach can be extended to cybersecurity domains that rely on sequential data. The refined attention mechanisms, especially in processing such data, have proven effective and hold significant potential for enhancing cybersecurity measures. Additionally, this approach could be advantageous in other applications, such as car plate recognition, where improved attention mechanisms can lead to more accurate and reliable results.

7 Conclusions

In summary, our study presents a groundbreaking approach to CAPTCHA recognition by integrating a specialized RNN model derived from the UpDown architecture in image captioning. Augmenting the model with CNNs enables the extraction of both local and global features, enhancing its capacity to decipher intricate details within CAPTCHA images. Further enhancement through refined visual attention mechanisms and dual layers of LSTM networks significantly bolsters performance. Across various datasets such as BoC, Weibo, Gregwar, and Captcha 0.3, our system achieves remarkable success rates without resorting to segmentation. Beyond theoretical advancements, our research holds practical significance in fortifying online security. The simplicity and adaptability inherent in our model position it as a promising solution in CAPTCHA recognition, leveraging techniques from image captioning. This work represents a notable stride forward in CAPTCHA security, shedding light on the potential applications of image captioning in enhancing internet security. For future work, we plan to incorporate advanced techniques such as Transformer models, self-attention, and multi-head attention, along with a two-layer attention mechanism. These enhancements are expected to significantly improve the model's performance and adaptability.

Acknowledgement: The authors are thankful to the Deanship of Graduate Studies and Scientific Research at University of Bisha for supporting this work through the Fast-Track Research Support Program.

Funding Statement: This work is supported by the National Natural Science Foundation of China (Nos. U22A2034, 62177047), High Caliber Foreign Experts Introduction Plan funded by MOST, and Central South University Research Programme of Advanced Interdisciplinary Studies (No. 2023QYJC020).

Author Contributions: Zaid Derea was responsible for investigation, conceptualization, and software development. Xiaoyan Kui contributed to the review and editing of the manuscript. Beiji Zou provided supervision throughout the project. Monir Abdullah participated in the investigation. Alaa Thobhani handled conceptualization, validation, and visualization, while Amr Abdussalam contributed resources. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Due to privacy concerns and proprietary restrictions, the dataset cannot be shared openly. The data were sourced from publicly available platforms, but the specific dataset remains confidential to protect participant privacy and ownership rights.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Sinha S, Surve MI. CAPTCHA recognition and analysis using custom based CNN model-capsecure. In: 2023 International Conference on Emerging Techniques in Computational Intelligence (ICETCI); 2023; IEEE. p. 244–50.
2. Kumar M, Jindal M, Kumar M. An efficient technique for breaking of coloured Hindi CAPTCHA. Soft Comput. 2023;27:11661–86. doi:10.1007/s00500-023-07844-3.
3. Wang P, Gao H, Guo X, Xiao C, Qi F, Yan Z. An experimental investigation of text-based CAPTCHA attacks and their robustness. ACM Comput Surv. 2023;55(9):1–38. doi:10.1145/3559754.
4. Hussain R, Gao H, Shaikh RA, Soomro SP. Recognition based segmentation of connected characters in text based CAPTCHAs. In: 2016 8th IEEE International Conference on Communication Software and Networks (ICCSN); 2016; IEEE. Vol. 2656, p. 673–6.
5. Von Ahn L, Blum M, Hopper NJ, Langford J. CAPTCHA. Using hard AI problems for security. In: Biham E (ed.), Advances in Cryptology—EUROCRYPT 2003. Lecture Notes in Computer Science. Berlin/Heidelberg, Germany: Springer; 2003. Vol. 2656. p. 294–311.
6. Von Ahn L, Blum M, Langford J. Telling humans and computers apart automatically. Commun ACM. 2004;47(2):56–60. doi:10.1145/966389.966390.
7. Kumar M, Jindal M, Kumar M. A systematic survey on CAPTCHA recognition: types, creation and breaking techniques. Arch Comput Methods Eng. 2022;29(2):1107–36. doi:10.1007/s11831-021-09608-4.
8. Tang M, Gao H, Zhang Y, Liu Y, Zhang P, Wang P. Research on deep learning techniques in breaking text-based captchas and designing image-based captcha. IEEE Trans Inf Forensics Secur. 2018;13(10):2522–37. doi:10.1109/TIFS.2018.2821096.
9. Wang P, Gao H, Rao Q, Luo S, Yuan Z, Shi Z. A security analysis of captchas with large character sets. IEEE Trans Depend Secure Comput. 2020;18(6):2953–68.
10. Ye G, Tang Z, Fang D, Zhu Z, Feng Y, Xu P et al. Yet another text captcha solver: a generative adversarial network based approach. In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security; 2018. p. 332–48.
11. Gao H, Wang W, Qi J, Wang X, Liu X, Yan J. The robustness of hollow CAPTCHAs. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security; 2013. p. 1075–86.
12. Malik S, Soundararajan R. Llrnet: a multiscale subband learning approach for low light image restoration. In: 2019 IEEE International Conference on Image Processing (ICIP); 2019; IEEE. p. 779–83.
13. Jin Z, Iqbal MZ, Bobkov D, Zou W, Li X, Steinbach E. A flexible deep CNN framework for image restoration. IEEE Trans Multimedia. 2019;22(4):1055–68.
14. Dong W, Wang P, Yin W, Shi G, Wu F, Lu X. Denoising prior driven deep neural network for image restoration. IEEE Trans Pattern Anal Mach Intell. 2018;41(10):2305–18.
15. Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017. p. 7263–71.

16. Liu Y. An improved faster R-CNN for object detection. In: 2018 11th International Symposium on Computational Intelligence and Design (ISCID); 2018. Vol. 2, p. 119–23.
17. Abdussalam A, Sun S, Fu M, Sun H, Khan I. License plate segmentation method using deep learning techniques. In: Signal and Information Processing, Networking and Computers: Proceedings of the 4th International Conference on Signal and Information Processing, Networking and Computers (ICSINC); 2019; Springer. p. 58–65.
18. Abdussalam A, Sun S, Fu M, Ullah Y, Ali S. Robust model for chinese license plate character recognition using deep learning techniques. In: Communications, Signal Processing, and Systems: Proceedings of the 2018 CSPS Volume III: Systems; 2020; Springer. p. 121–7.
19. Anderson P, He X, Buehler C, Teney D, Johnson M, Gould S et al. Bottom-up and top-down attention for image captioning and visual question answering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018. p. 6077–86.
20. Al-Qatf M, Hawbani A, Wang X, Abdusallam A, Alsamhi S, Alhabib M, et al. RVAIC: Refined visual attention for improved image captioning. J Intell Fuzzy Syst. 2024;46(2):3447–59. doi:10.3233/JIFS-233004.
21. Atri A, Bansal A, Khari M, Vimal S. De-CAPTCHA: a novel DFS based approach to solve CAPTCHA schemes. Comput Electr Eng. 2022;97:107593.
22. Mori G, Malik J. Recognizing objects in adversarial clutter: breaking a visual CAPTCHA. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2003; Madison, WI, USA.
23. Chellapilla K, Simard P. Using machine learning to break visual human interaction proofs (HIPs). Adv Neural Inf Process Syst. 2004;17.
24. Yan J, El Ahmad AS. A low-cost attack on a microsoft CAPTCHA. In: Proceedings of the 15th ACM Conference on Computer and Communications Security; 2008. p. 543–54.
25. El Ahmad AS, Yan J, Marshall L. The robustness of a new CAPTCHA. In: Proceedings of the Third European Workshop on System Security; 2010. p. 36–41.
26. Gao H, Tang M, Liu Y, Zhang P, Liu X. Research on the security of microsoft's two-layer captcha. IEEE Trans Inf Forensics Secur. 2017;12(7):1671–85.
27. Thobhani A, Gao M, Hawbani A, Ali STM, Abdussalam A. CAPTCHA recognition using deep learning with attached binary images. Electronics. 2020;9(9):1522.
28. Derea Z, Zou B, Al-Shargabi AA, Thobhani A, Abdussalam A. Deep learning based CAPTCHA recognition network with grouping strategy. Sensors. 2023;23(23):9487. doi:10.3390/s23239487.
29. Khatavkar V, Velankar M, Petkar S. Segmentation-free Connectionist Temporal Classification loss based OCR Model for Text Captcha Classification. arXiv: 240205417. 2024.
30. Chang G, Gao H, Pei G, Luo S, Zhang Y, Cheng N, et al. The robustness of behavior-verification-based slider CAPTCHAs. J Inf Secur Appl. 2024;81:103711. doi:10.1016/j.jisa.2024.103711.
31. Huang Y, Chen J, Ouyang W, Wan W, Xue Y. Image captioning with end-to-end attribute detection and subsequent attributes prediction. IEEE Trans Image Process. 2020;29:4013–26. doi:10.1109/TIP.2020.2969330.
32. Al-Qatf M, Wang X, Hawbani A, Abdusallam A, Alsamhi SH. Image captioning with novel topics guidance and retrieval-based topics re-weighting. IEEE Trans Multimedia; 2022;25:5989–99. doi:10.1109/TMM.2022.3202690.
33. Hossen MB, Ye Z, Abdussalam A, Hossain MI. GVA: Guided visual attention approach for automatic image caption generation. Multimed Syst. 2024;30(1):50. doi:10.1007/s00530-023-01249-w.