UNIVERSITI TEKNOLOGI MARA

ENHANCED RECOGNITION METHODS FOR TEXT AND SLIDER CAPTCHA VULNERABILITY ASSESSMENT

WAN XING

Thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering)

Faculty of Electrical Engineering

September 2025

ABSTRACT

As vital components of human-computer interaction, CAPTCHAs are widely used across various industries, including video, education, finance, e-commerce, aviation, and public services. Research into CAPTCHA recognition is crucial for identifying network security vulnerabilities and advancing cybersecurity measures. Among the commonly used types, text and slider CAPTCHAs are particularly notable. While slider CAPTCHAs ask users to determine the location of a gap, text CAPTCHAs require users to recognize characters. In text CAPTCHAs, significant noise and correlations between characters pose challenges for recognition. To address these anti-attack mechanisms in text CAPTCHAs, a new text CAPTCHA framework has been proposed consisting of three parts: a data augmentation module, a font enhancement network, and a recognition network. Among them, the recognition network named Adaptive-CAPTCHA is improved based on Deep-CAPTCHA, consisting of Convolutional Recurrent Neural Network (CRNN), Adaptive Fusion Filtering Networks (AFFN), and residual connections, achieving an Average Attack Success Rate (AASR) of 99% on complex datasets (M-CAPTCHA) and near-perfect performance (99.9%) on simpler ones (P-CAPTCHA). Additionally, a Font Enhancement (FE) network based on Generative Adversarial Networks (GAN) has been introduced, which significantly undermines the interference in text CAPTCHAs. To address the limitation of traditional color enhancement algorithms that lack adaptive learning capabilities, three types of Variation Color Shift (VCS) algorithms have also been proposed for data augmentation. Experimental results show VCS notably improves recognition accuracy; for instance, it boosts the AASR of Adaptive-CAPTCHA on the challenging M-CAPTCHA dataset by approximately 10 percentage points compared with no color shift. For slider CAPTCHA detection, mean Relative Offset (mRO) has been proposed as a specific metric for slider CAPTCHA recognition, and Offset-based Intersection over Union (OIoU) loss is developed to improve the loss function, effectively reducing the mRO to below 1% on the Geetest dataset. The Fixed Quantity Prediction Non-Maximum Suppression (FQP-NMS) method, along with lightweight backbones and attention mechanisms, is proposed to optimize recognition performance and improve the architecture of YOLOv5, achieving a mean Average Precision (mAP) of 0.994 on the SliderCAPTCHA dataset. By comparing them with the benchmark models in terms of recognition accuracy, precision, computational complexity, storage space, and other metrics, the superiority of our proposed algorithm is proved. Furthermore, we have delineated the path for subsequent study and enhancement.

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to my main supervisor, Dr. Fazlina Ahmat Ruslan, for her invaluable guidance and support throughout this journey. I greatly appreciate her thoughtful mentorship, which has created a relaxed research environment and allowed me to fully enjoy my PhD experience, especially with her timely assistance. I also extend my thanks to my associate supervisor, Prof. Juliana Johari, for her insightful advice and encouragement. Her patience and precise guidance have been instrumental to my progress.

I would like to express my sincere gratitude to Leshan Vocational and Technical College for their financial support. My thanks also extend to the Leshan Special Robot Engineering Technology Research Center for providing invaluable experimental computing resources that facilitated the completion of extensive algorithmic experiments. Additionally, I am grateful to my colleagues and friends for their suggestions, advice, and encouragement throughout this project.

I extend my deepest gratitude to my wife, Yawen Luo, for her unwavering support and care, which allowed me to devote myself entirely to my research. I am especially thankful for her companionship during our extensive travels across the Sichuan, Guizhou, Guangxi, and Yunnan provinces, where we collected crucial landscape data and organized CAPTCHA datasets together. Her support has created a warm and loving home, enabling me to fully concentrate on my studies.

This thesis is dedicated to my beloved parents, Peixuan Wan and who have nurtured me and created a heaven of love throughout my life. I deeply miss my late father, whose meticulous care, boundless love, and cherished companionship have transformed into countless joyful memories and an enduring legacy. He always encouraged me to complete my doctoral degree, and this thesis serves as a profound source of solace and remembrance for him.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 Research Background

The rapid growth of the Internet has enabled both companies and individuals to leverage it for business, communication, and entertainment, significantly altering lifestyles [1]. While the online environment offers numerous advantages, it also presents certain security challenges. One widely adopted protective measure is the use of CAPTCHA. In the early 21st century, the concept of CAPTCHA was introduced by Luis von Ahn et al. [2]. This public security tool primarily functions to differentiate between human users and automated bots, effectively blocking malicious software from gaining access [3]. A CAPTCHA mechanism is deemed secure if the probability of a computer's successful breach is less than 10% [4]. As the first line of defense in information systems, CAPTCHA provides a direct and effective layer of security. Today, most websites implement CAPTCHAs to protect against online threats [5]. Recently, advancements in deep learning technology have offered valuable insights and frameworks for both attacking and protecting CAPTCHAs [6]. As a result, CAPTCHA recognition has emerged as a significant research focus in both academia and industry. The security of CAPTCHAs has been actively studied by many world-class universities and scientific research institutions [7]. There are about ten categories of CAPTCHAs with different designs, and the most prevalent type of CAPTCHA is text-based [8]. As the first CAPTCHA method, text-based CAPTCHAs are simple to generate and deploy as images. Their minimal storage requirements and rapid loading speeds have led to widespread adoption by many websites, making them the most commonly used CAPTCHA mechanism [9]. The text-based CAPTCHA has accounted for more than 50% of the market share, surpassing other types of CAPTCHA [1].

To protect against cyberattacks, text-based CAPTCHAs contain lots of security schemes [8]. However, their security has diminished due to advancements in Optical Character Recognition (OCR) [10]. To enhance the recognition difficulty, variations in fonts, distortions, noise, colors, shades, positions, and backgrounds are utilized [11]. Each CAPTCHA usually adopts multiple schemes to increase anti-attack ability [12]. Figure 1.1 presents examples of text-based CAPTCHAs. Each CAPTCHA utilizes a