Robust CAPTCHAs towards Malicious OCR

# Introduction

Turing test was originally proposed to examine whether machine’s behavior is indistinguishable from a human. The most popular and practical Turing test is CAPTCHA, which is to discriminate algorithm from human by offering recognitionalike questions. The recent development of deep learning has significantly advanced the capability of algorithm in solving CAPTCHA questions, forcing CAPTCHA designers to increase question complexity. Instead of designing questions difficult for both algorithm and human, this study attempts to employ the limitations of algorithm to design robust CAPTCHA questions easily solvable to human. Specifically, our data analysis observes that human and algorithm demonstrates different vulnerability to visual distortions: adversarial perturbation is significantly annoying to algorithm yet friendly to human. We are motivated to employ adversarially perturbed images for robust CAPTCHA design in the context of character-based questions. Four modules of multi-target attack, ensemble adversarial training, image preprocessing differentiable approximation, and expectation are proposed to address the characteristics of character-based CAPTCHA cracking. Qualitative and quantitative experimental results demonstrate the effectiveness of the proposed solution. We hope this study can lead to the discussions around adversarial attack/defense in CAPTCHA design and also inspire the future attempts in employing algorithm limitation for practical usage.

In summary, the contributions of this study are two-fold:

1. We have discovered the different vulnerability between human and algorithm on visual distortions. Based on the observations, adversarial perturbation is employed to improve the robustness of character-based CAPTCHA.
2. Corresponding to the characteristics of typical OCR cracking solutions, we proposed a novel methodology addressing issues including sequential recognition, indifferentiable image preprocessing, stochastic image transformation and black-box cracking.

# Framework

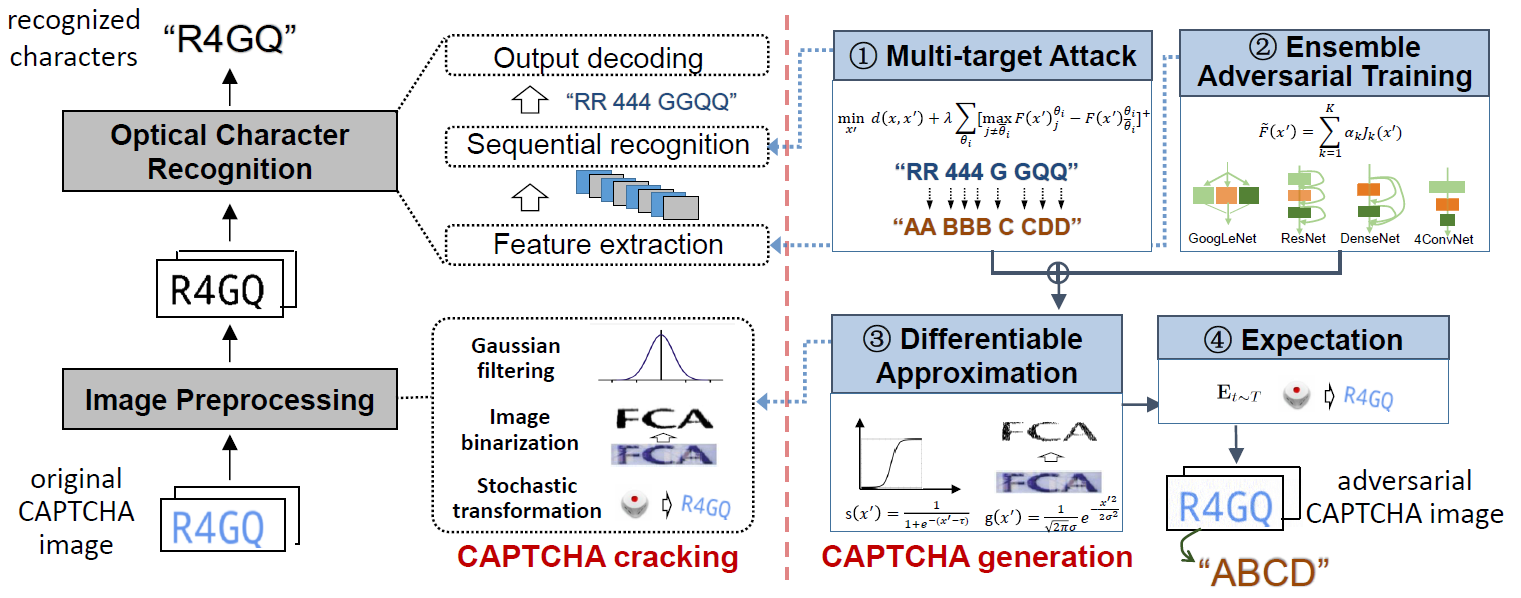


Fig.1 The proposed robust CAPTCHA designing framework. The left represents the process of CAPTCHA cracking, including sequential recognition, feature extraction, image binarization (Gaussian filtering) and stochastic transformation. The right represents our solution of CAPTCHA generation, including the corresponding multi-target attack, ensemble adversarial training, differentiable approximation and expectation, respectively.

# Paper

https://ieeexplore.ieee.org/abstract/document/9158388