

**CHENNAI – PONDICHERRY**

**Research on Deep Learning Techniques in Breaking Text-based Captchas and Designing Image-based Captcha**

**Abstract**

The ability of hackers to infiltrate computer systems using computer attack programs and bots led to the development of Captchas or Completely Automated Public Turing Tests to Tell Computers and Humans Apart. The text Captcha is the most popular Captcha scheme given its ease of construction and user friendliness. However, the next generation of hackers and programmers has decreased the expected security of these mechanisms, leaving websites open to attack. Text Captchas are still widely used, because it is believed that the attack speeds are slow, typically two to five seconds per image, and this is not seen as a critical threat. In this paper, we introduce a simple, generic, and fast attack on text Captchas that effectively challenges that supposition. With deep learning techniques, our attack demonstrates a high success rate in breaking the Roman-character-based text Captchas deployed by the top 50 most popular international websites and three Chinese Captchas that use a larger character set. These targeted schemes cover almost all existing resistance mechanisms, demonstrating that our attack techniques are also applicable to other existing Captchas. Does this work then spell the beginning of the end for text-based Captcha? We believe so. A novel image-based Captcha named Style Area Captcha (SACaptcha) is proposed in this paper, which is based on semantic information understanding, pixel-level segmentation, and deep learning techniques. Having demonstrated that text Captchas are no longer secure, we hope that our proposal shows promise in the development of image-based Captchas using deep learning techniques.