1 Abstract

CAPTCHAs are used by websites to defense against attacks like malicious programs, automated registrations, email spam, dictionary attacks, and search engine bots. To achieve higher security, different types of CAPTCHA have been invented in the past few years, including text-based, image-based, video based and game-based. However, with the rapid development of machine learning techniques in recent years, most of the existing CAPTCHAs become vulnerable and can be solved easily by well-trained machine learning models. In this paper, we show that it is possible to restrict the image recognition ability of machine learning models by using adversarial examples in image-based CAPTCHA.We create adversarial examples by integrating multiple adversarial attacks into one model and use this model to add perturbations to the original legitimate images. The experiments show that the adversarial examples generated by our model can significantly weaken the ability of machine learning models whilst maintain the usability of images. Our analysis and experiments demonstrate that our adversarial examples have the ability to largely increase the security of existing image-based CAPTCHAs.