GLOBAL
INNOVATION
EXCHANGE

Baidu

Team Members: Darren Yang, Xincheng Li, Jay Xiong

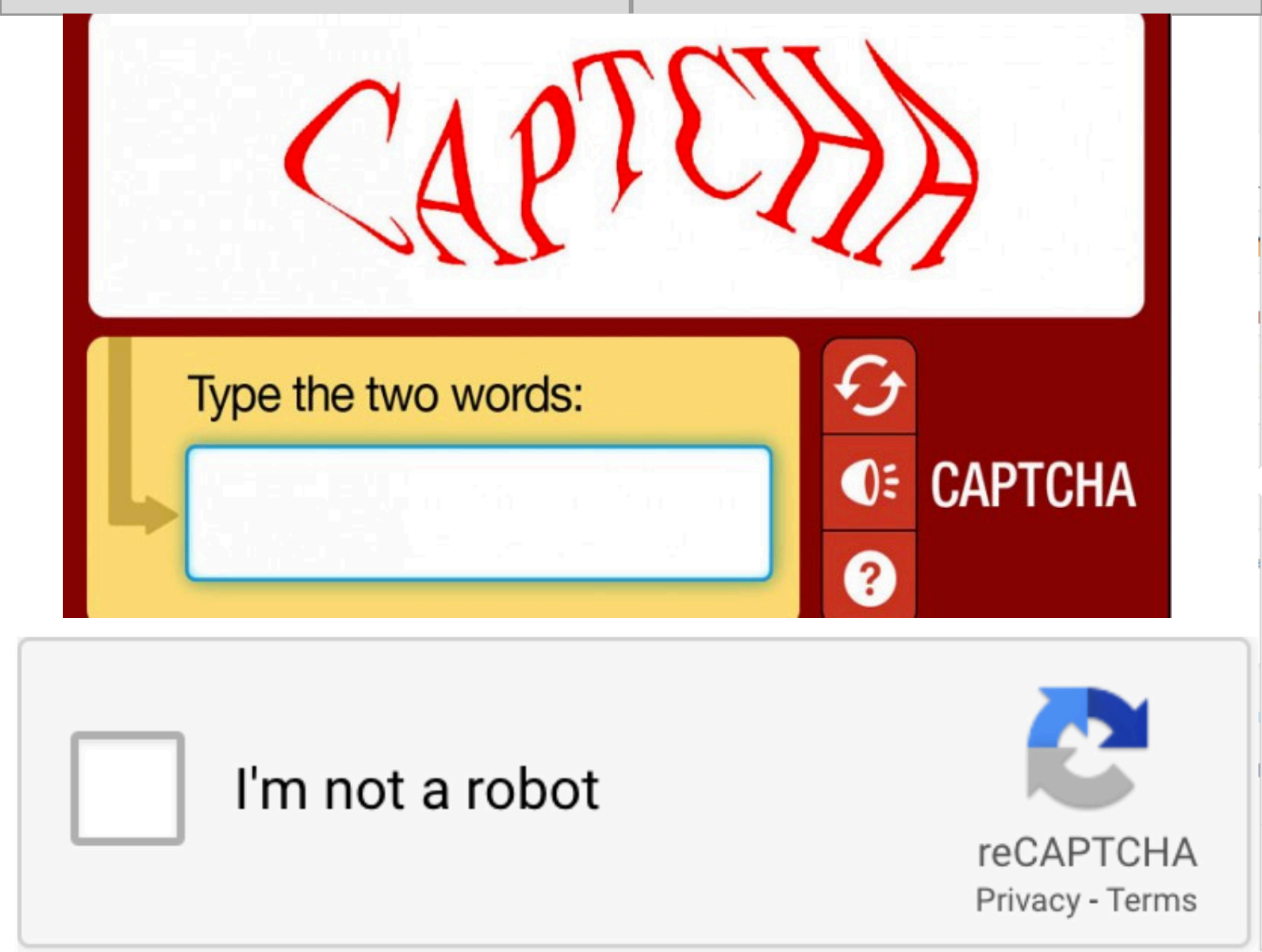
Bai du 百度

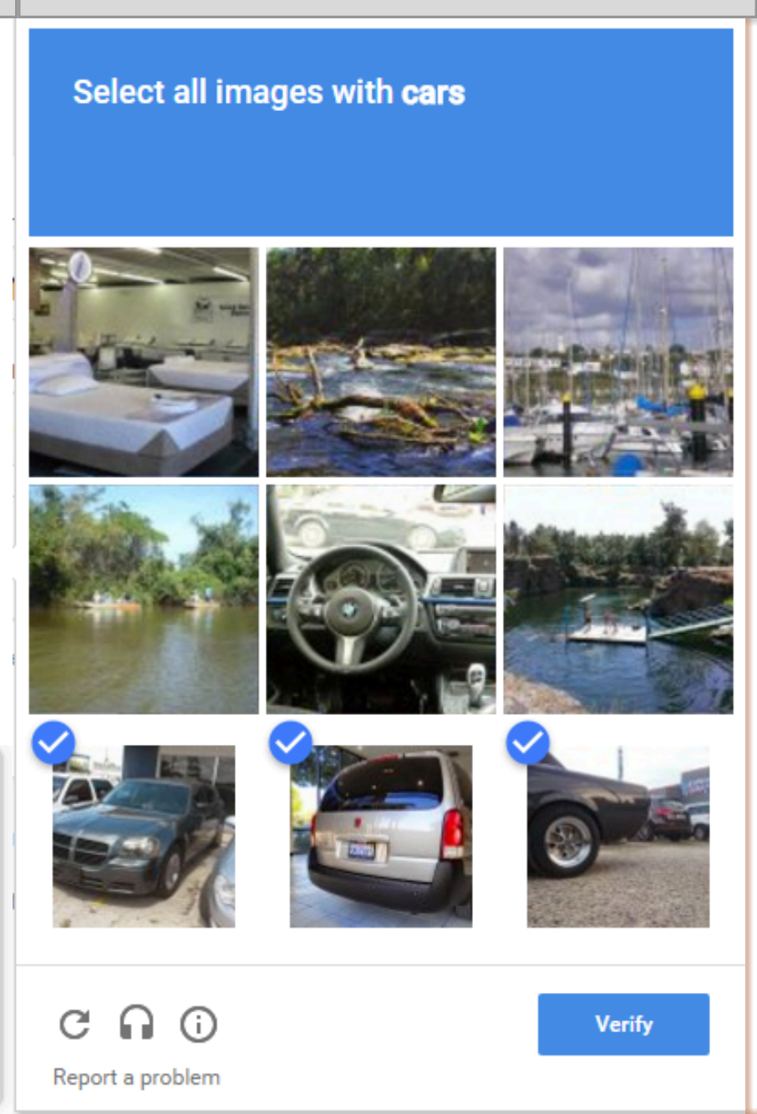
Problem

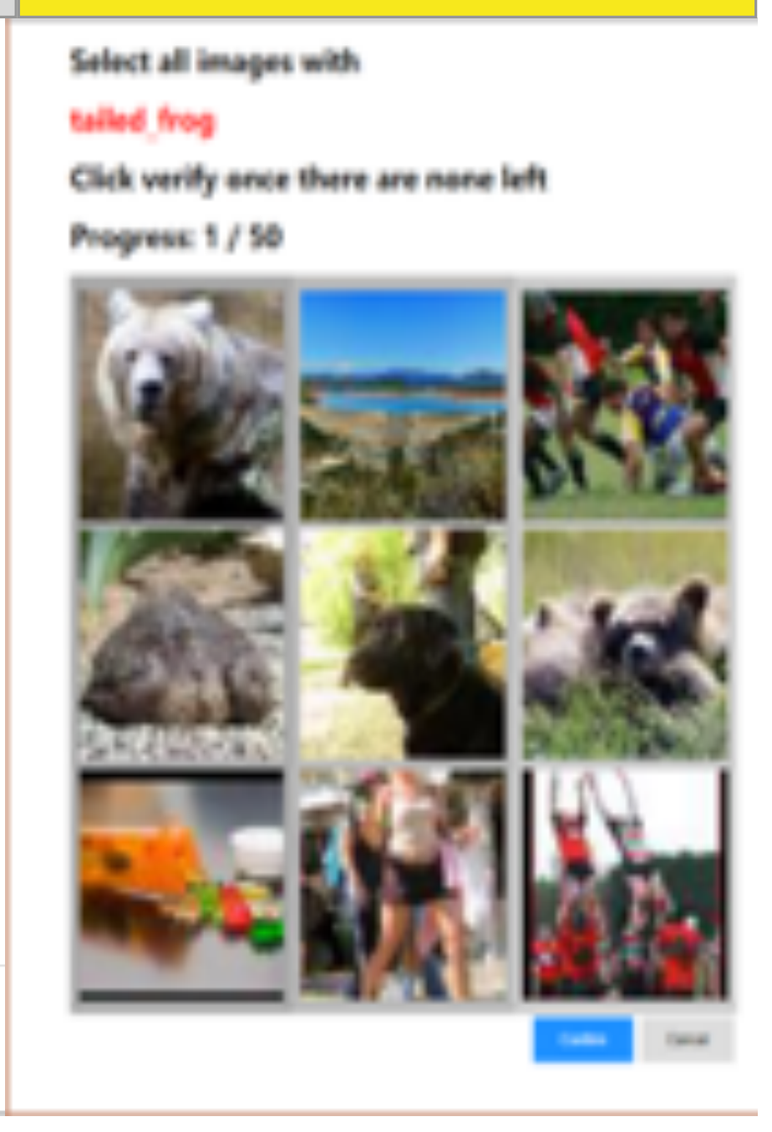
With increasing amount of valuable and privacy-sensitive digital assets uploaded on the social media, online booking website and et al, image CAPTCHAs are being widely used across the Internet to defend against abusive programs[1] and economics of cyber-crimes that rely on large-scale automation. However, via deep learning technology, a CAPTCHA breaker is capable of compromising most of the CAPTCHAs on the market[2,3]. What's more, the arm race between defender and CAPTCHA breaker results in CAPTCHAs too complex for humans making traditional CAPTCHA schemas only a hassle for the Internet users.

Competitor Analysis of CAPTCHAs

	Text	Behavior	Image	Our AdCaptcha
For Computer	Easy	Moderate	Hard	Hard
For Human	Moderate	Easy	Hard	Easy



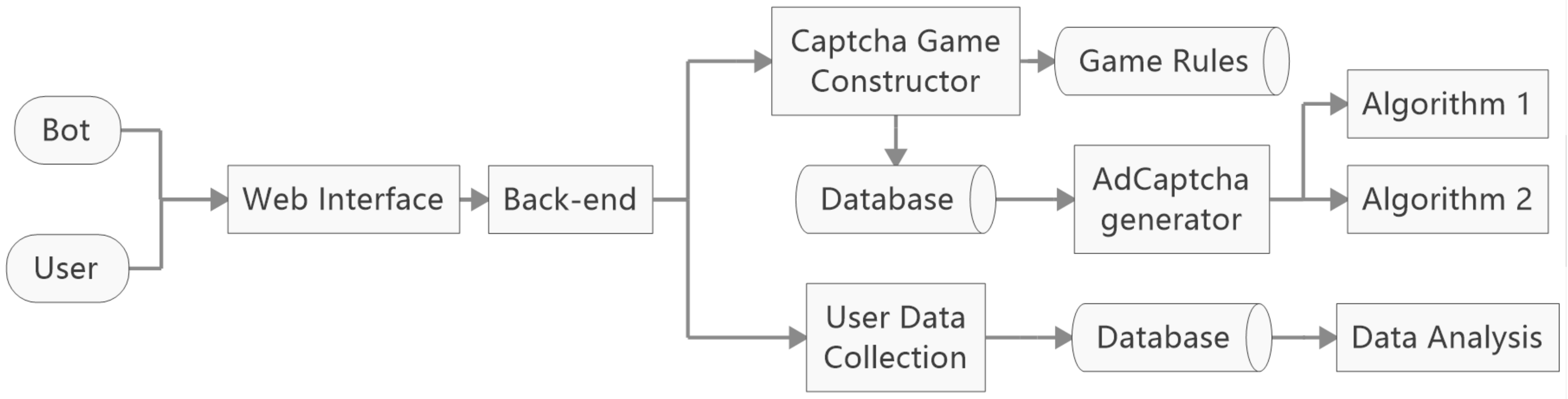



Select all images with
tailed_frog
Click verify once there are none left
Progress: 1 / 50

Solution

To make CAPTCHAs more effective in screening bots and more friendly to human users, we proposed AdCaptcha (Adversarial CAPTCHA). AdCaptcha's schema comes from deep learning attacking algorithms which differs from traditional ones whose security schemas are produced by a set of fixed rules. The technology could exploit the maximum effectiveness of a security schema that could be imposed on a CAPTCHA image. We used C&W[4] attacking algorithm to build Adversarial images on ImageNet[5] data set, and we prototyped a system with API that works on both human and bot for HCI testing.

System Architecture



References:

[1] "Towards Evaluating the Security of Real-World Deployed ... - NESA." https://nesa.zju.edu.cn/download/Towards%20Evaluating%20the%20Security%20of%20Real-World%20Deployed%20Image.pdf. Accessed 2 Dec. 2018.

[2] "Yet Another Text Captcha Solver: A Generative Adversarial Network ...." http://www.lancaster.ac.uk/staff/wangz3/publication/ccs18/. Accessed 2 Dec. 2018

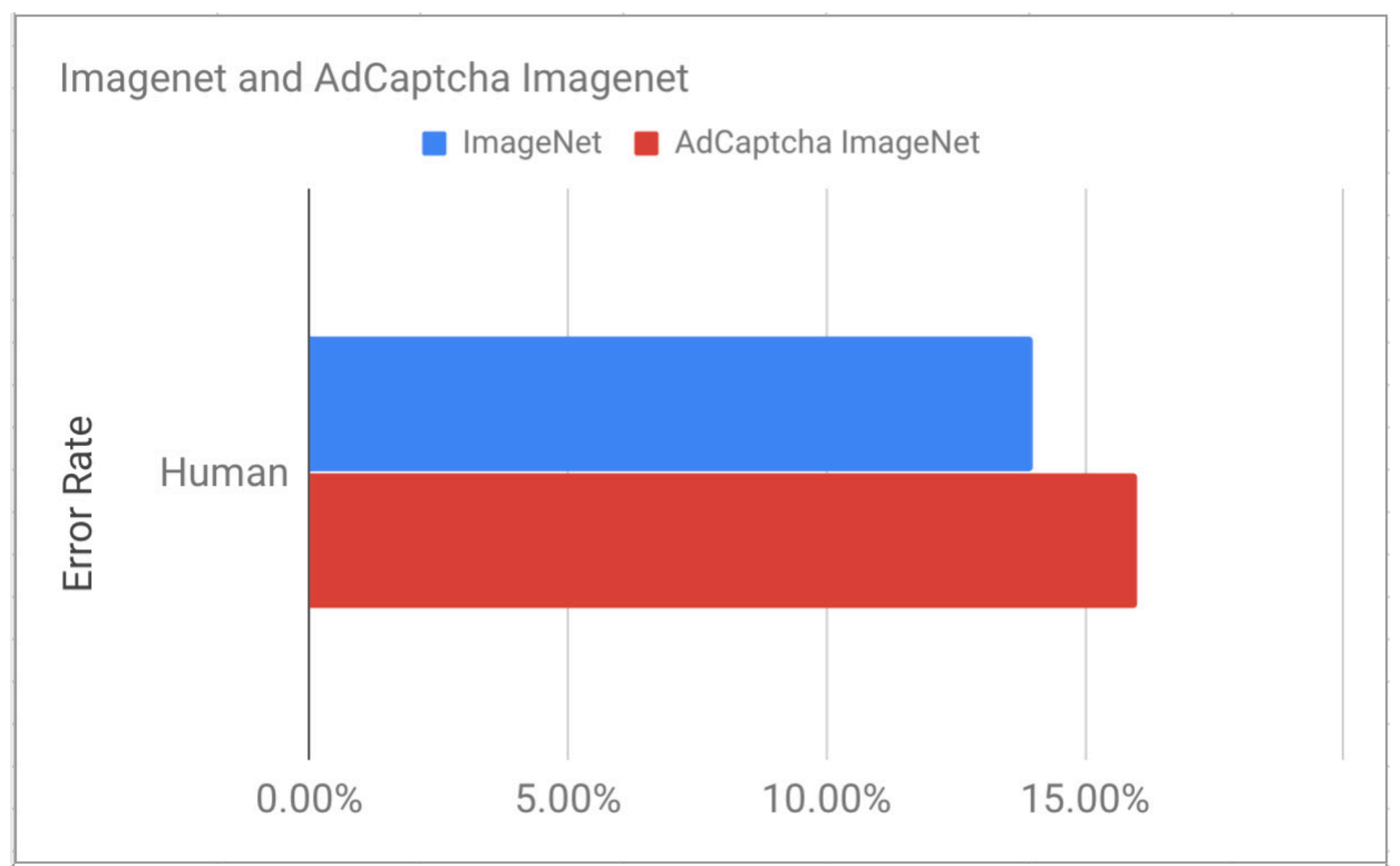
[3] "I'm not a human: Breaking the Google reCAPTCHA - Black Hat." https://www.blackhat.com/docs/asia-16/materials/asia-16-Sivakorn-Im-Not-a-Human-Breaking-the-Google-reCAPTCHA-wp.pdf. Accessed 2 Dec. 2018.

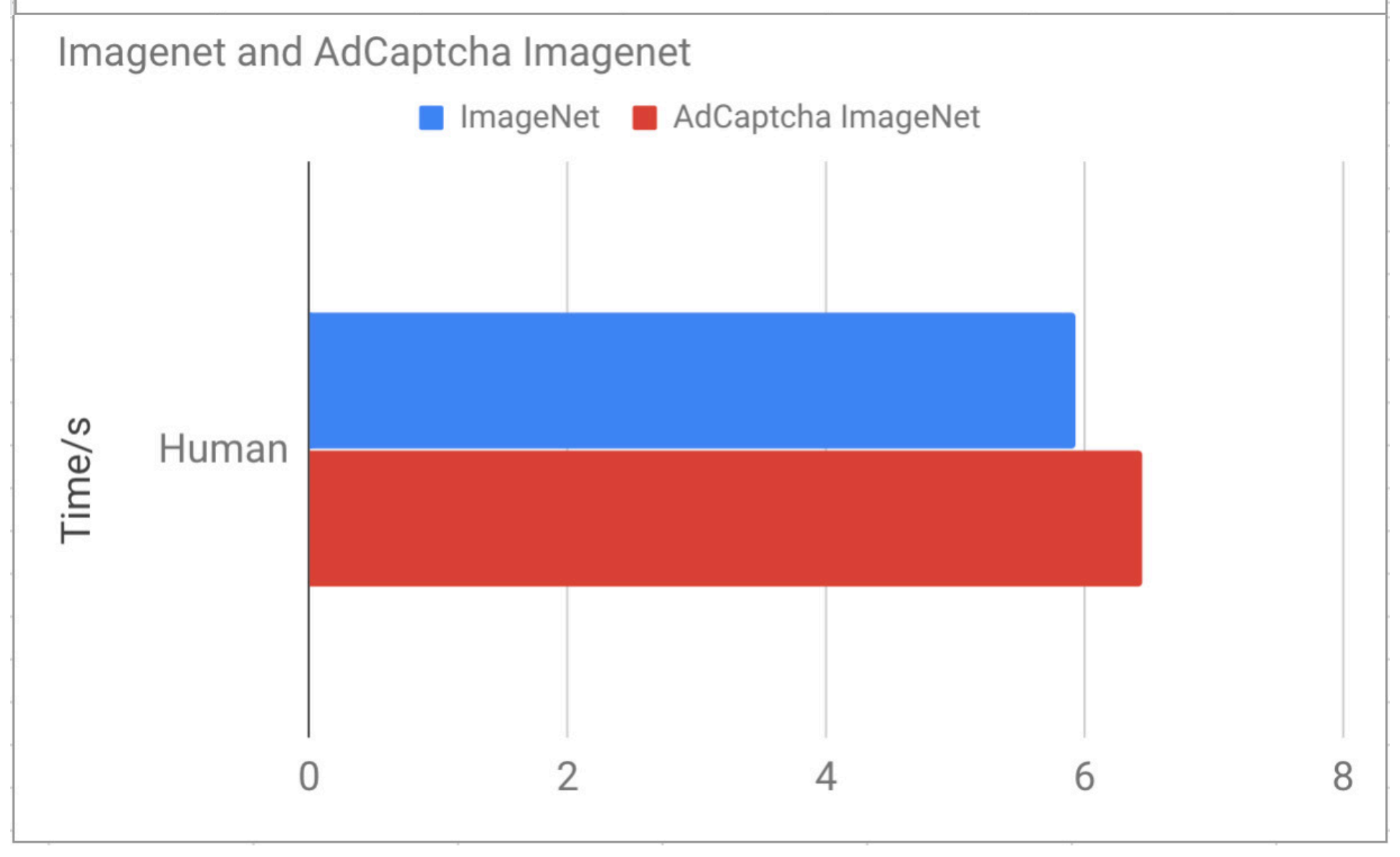
[4] "Towards Evaluating the Robustness of Neural Networks." Accessed December 4, 2018. https://arxiv.org/abs/1608.04644.

[5] "ImageNet Large Scale Visual Recognition Competition (ILSVRC)." Accessed December 4, 2018. http://image-net.org/challenges/LSVRC/.

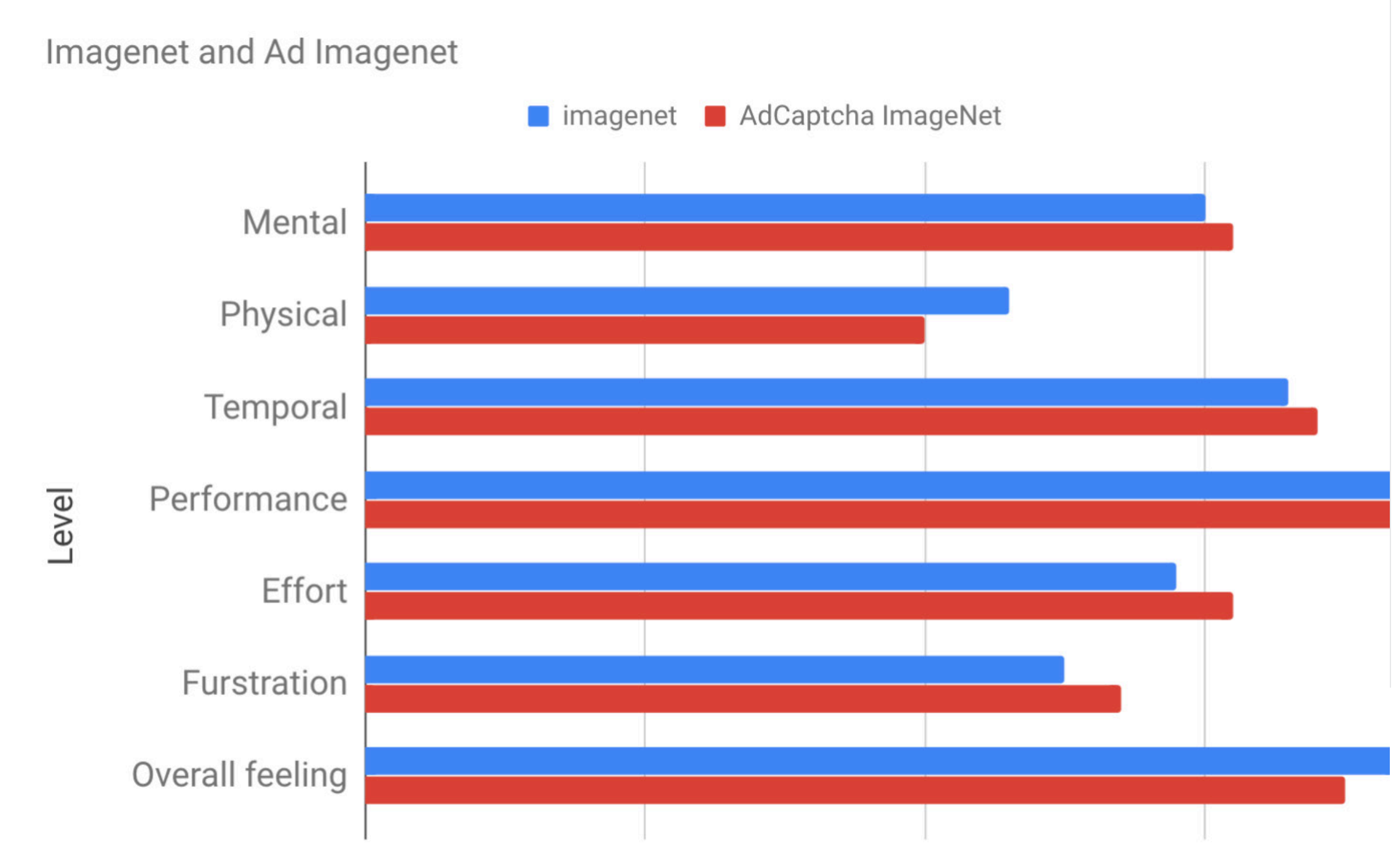
Evaluation

We conducted two groups of tests on Google Vision API and human participants. Independent variables are generated adversarial images and original images from imagenet dataset. Captcha recognition error rate and time are used for metrics. In addition, NASA TLS is used for difficulty measurement on human test. Results show that for Google Vision API, adversarial images have higher error rate, while for human original and adversarial images don't have significant difference on error rate, time and difficulties. This means adversarial captcha can defend captcha cracker better, and is as easy as normal images for human.





User Testing Result


Imagenet and Ad Imagenet
imagenet
AdCaptcha ImageNet
Level
Mental
Physical
Temporal
Performance
Effort
Furstration
Overall feeling