

Competition 4: Attacking CAPTCHA

Shan-Hung Wu & DataLab
Machine Learning

Outline

- Why CAPTCHA?
- Tasks and datasets
- Evaluation
- Hints from TA

What is CAPTCHA?

- Completely Automated Public Turing test to tell Computers and Humans Apart
- Tells humans from bots

$$7 \times \begin{matrix} 5 \\ \text{[large stylized yellow/red number]} \end{matrix} = \boxed{\text{[empty square]}}$$



Levelers critics

Type the two words:

reCAPTCHA
stop spam.
read books.

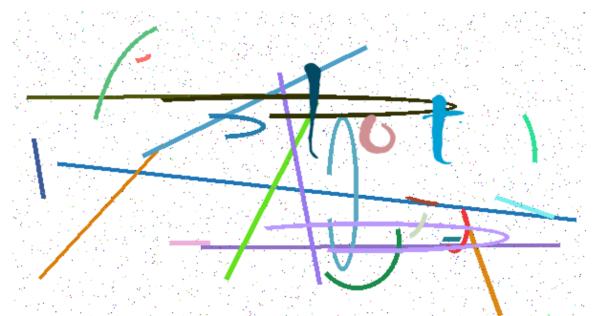
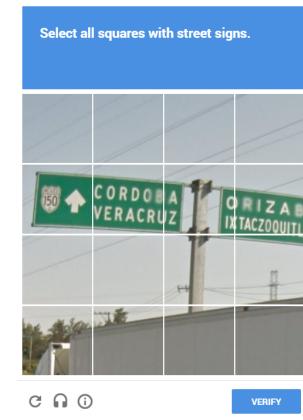


↻ ⓘ

Verify

I'm not a robot

reCAPTCHA
Privacy - Terms

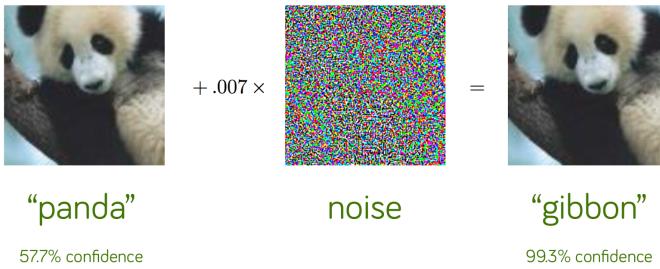


Why CAPTCHA?

- Spam accounts
- Brute-forcing passwords
- Booking bots
- Free web service collectors
- Fake comments or votes
- and more...

Does CAPTCHAs Secure?

- Google has used ML and AI to enhance CAPTCHAs
- Based on adversarial attacks



選取圖片中含有
汽車
的所有圖片
確定沒有遺漏後，請按一下 [驗證]。

驗證

Can AI/ML Used to Break CAPTCHAs?



CAPTCHA solving service

- ✓ Cheapest price on the market
Starting from 0.5USD per 1000 images, depending on your daily upload volume
- ✓ Pay as you go
Pay-per-captcha payment basis. Minimum refill is 1 USD, no recurring charges
- ✓ 99.99% uptime since 2007
Vast amount of workers and premium infrastructure allows us to provide highly reliable 24/7/365 service
- ✓ Solving Google Recaptcha since 2016
You may fully rely on our stable solution and forget about browser emulation

[Create Account](#)

 [Customers Area](#)

Outline

- Why CAPTCHA?
- Tasks and datasets
- Evaluation
- Hints from TA

Mission

- We provide CAPTCHA
- You break it!



Input/Output

- Input:
 - 1 question image
 - 9 candidate images
- Output:
 - 101010000

請從以下的九宮格中，選出與左圖相同的物品

1	0	1
0	1	0
0	0	0

Verify

Tasks



- Task1:
 - Distortion



- Task2:
 - Distortion
 - Style transfer



- Task3:
 - Distortion
 - Style transfer
 - Adversarial attack

Datasets

- For each task:
 - Training set (~20K images, ~50K CAPTCHAs)
 - Validation set (~4K images , ~10K CAPTCHAs)
 - Testing set (~24K images , ~60K CAPTCHAs)
- ***The objects in these sets are different!***
 - If you train a model, make sure it generalizes to ***unseen*** objects

Outline

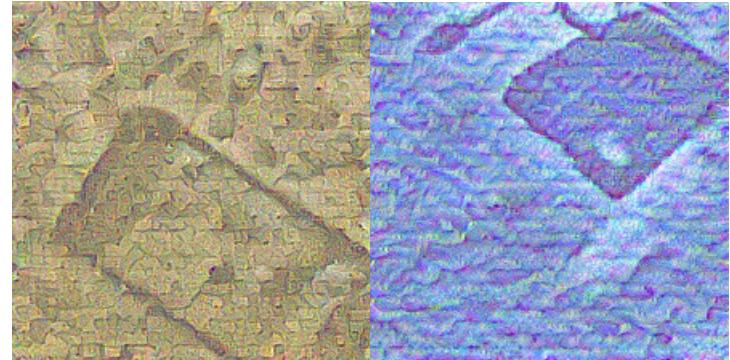
- Why CAPTCHA?
- Tasks and datasets
- **Evaluation**
- Hints from TA

Scoring

- Task 1: 30% (crack)
- Task 2: 15% (crack) + 15% (ranking)
- Task 3: 15% (crack) + 15% (ranking)
- Report: (10%)

Bonus

- Bonus1: 20% (crack)
 - Stronger adversarial attack
 - Other secrets...
- Bonus2: 10% (crack)
 - Input: audio clip
 - Output: #cues (1, 2, or 3)



Evaluation

- What does the “crack” mean?
 - >10% accuracy for image CAPTCHAs
 - >50% accuracy for audio CAPTCHAs

Your Report Must Include...

- For each task:
 - Code and model weights (if any)
 - For reproducibility
 - Methods and steps you've tried
 - How did they perform?
- Conclusion and lesson learned (10%)
- ***Missing a task means 0 score!***

Schedule

- Deadline: 2020/01/14 23:59
- Showoff: 2020/01/16 15:30

Outline

- Why CAPTCHA?
- Tasks and datasets
- Evaluation
- Hints from TA

Hints

- Read this [adversarial attack](#) paper
- You can try any methods, except
 - Manually solve CAPTCHAs in test set
 - Peek information in the test set

Q&A