Research on the Security of Visual Reasoning CAPTCHA

Yipeng Gao[1] Haichang Gao[1]* Sainan Luo[1] Yang Zi[1] Shudong Zhang[1]
Wenjie Mao[1] Ping Wang[1] Yulong Shen[1] Jeff Yan[2]

[1] *School of Computer Science and Technology, Xidian University*

[2] *Department of Computer and Information Science, Linköping University*

西安電子科技大學
XIDIAN UNIVERSITY

OUTLINE

01 Background

AI Problems Underlying Existing CAPTCHA Schemes

Text-based CAPTCHA

Character Recognition

Character Segmentation


(a)


(b)


(c)


(d)


(e)


(f)


(g)


(h)


(i)


(j)


(k)

01 Background

AI Problems Underlying Existing CAPTCHA Schemes

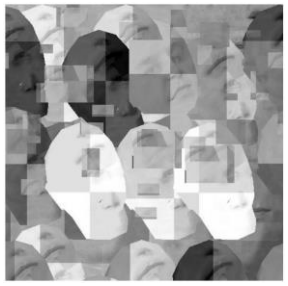
Image-based CAPTCHA

Object Recognition

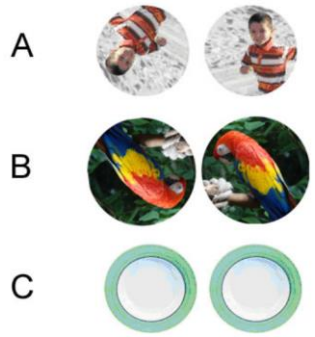
Face Recognition

Image Perception

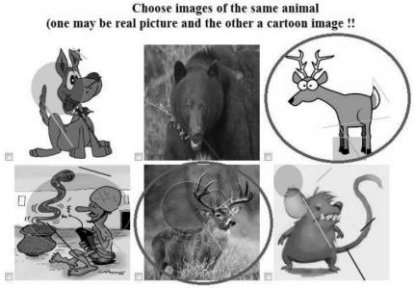
ASIRRA CAPTCHA



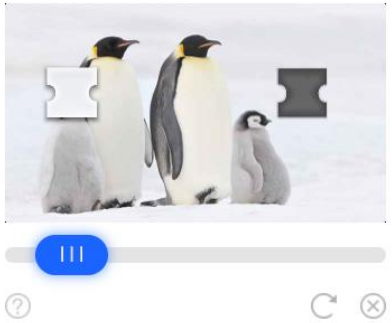
ARTiFACIAL CAPTCHA



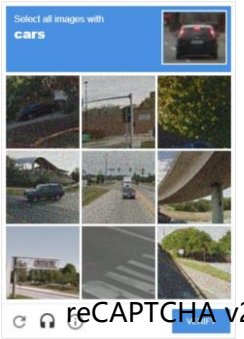
What' s Up CAPTCHA



SEMAGE CAPTCHA



Slider CAPTCHA

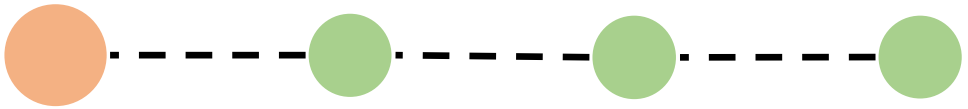


reCAPTCHA v2

01 Background

AI Problems Underlying Existing CAPTCHA Schemes

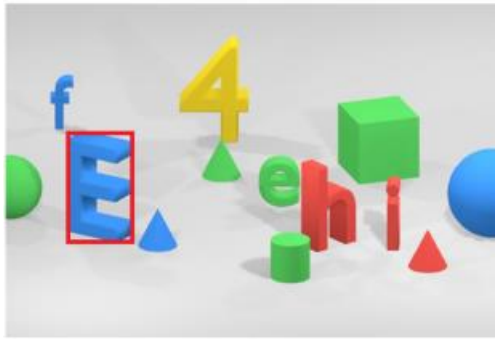
Visual Reasoning CAPTCHA



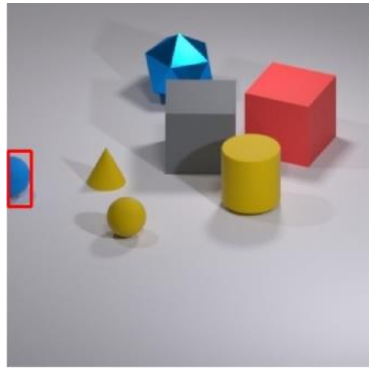
Adversarial Perturbation

Semantic Comprehension

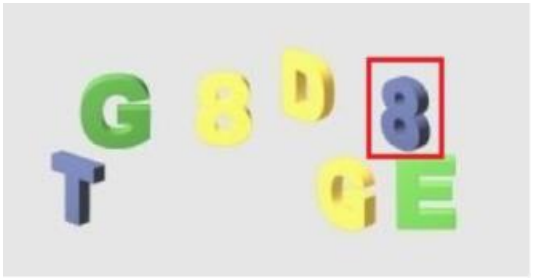
Behavior Detection



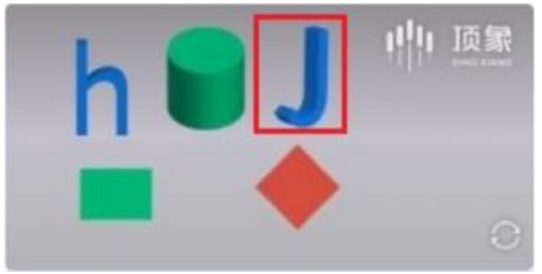
Tencent VTT



Geetest

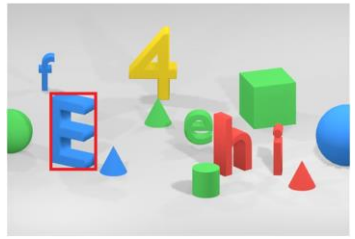


NetEase

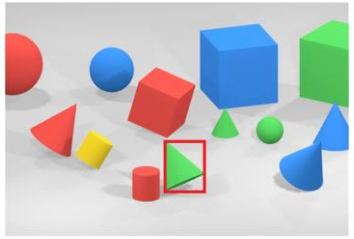


DingXiang

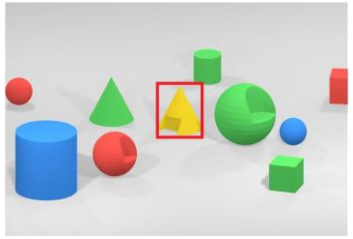
02 Visual Reasoning CAPTCHAs



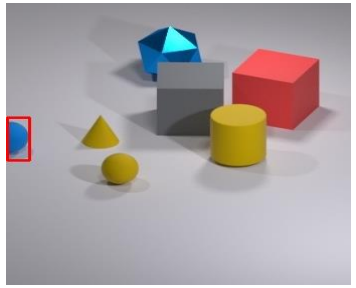
Q1: Please click the uppercase of the green letter



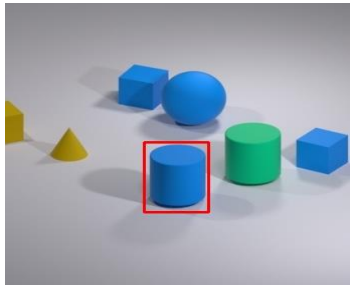
Q2: Please click the object with a different tilt direction



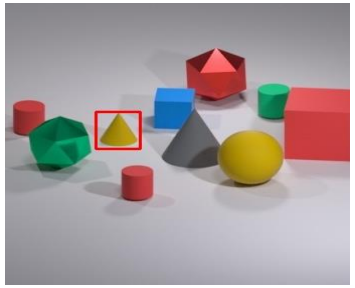
Q3: Please click the object with a notch on the left



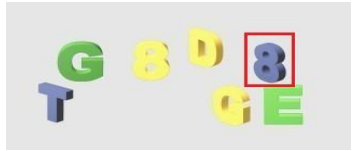
Q1: Please click the tiny sphere that is behind the big cylinder



Q2: Please click the other blue thing that is the same shape as the big green thing



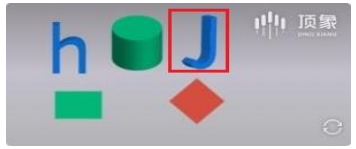
Q3: Please click the other conoid that is the same color as the big sphere



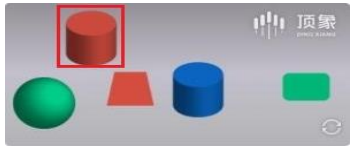
Q1: Please click the number 8 that is side facing to you

Q2: Please click the letter h with the same color as the number 6

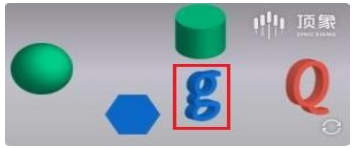
Q3: Please click the letter J with the same direction as the letter n



Q1: Please click the letter that is on the right of the cylinder



Q2: Please click the object with the same color as the trapezoid



Q3: Please click the letter closest to the sphere

Main object category in the existing visual reasoning schemes.

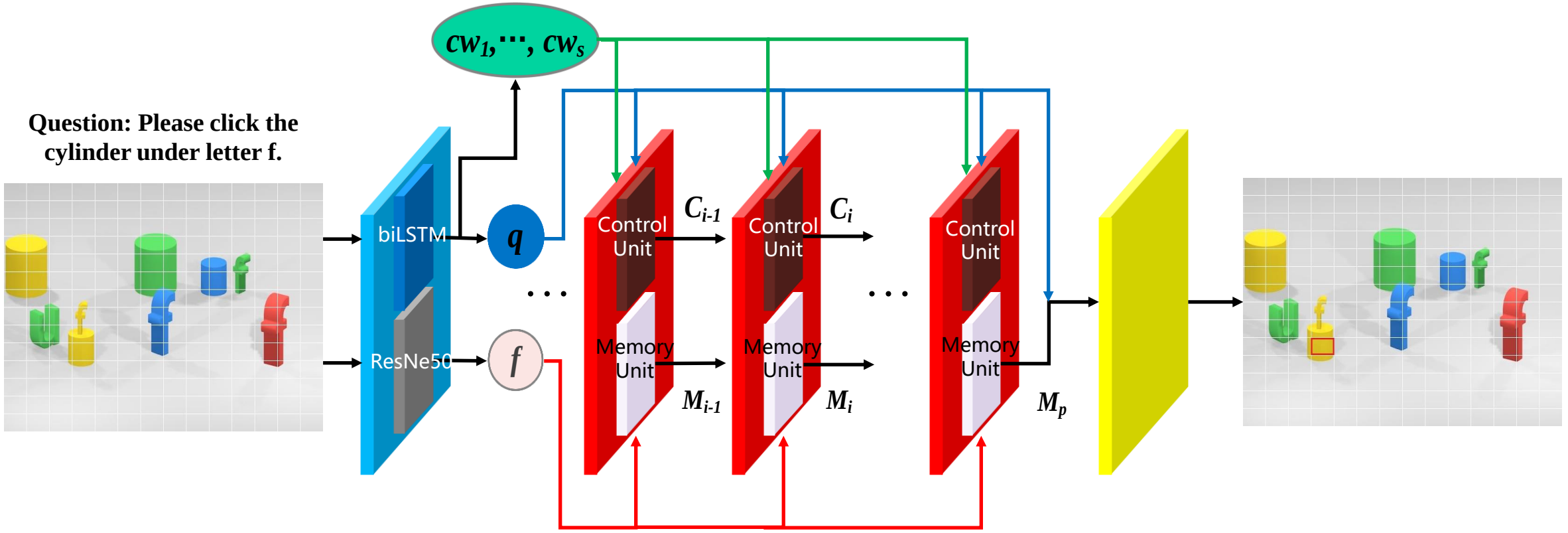
	VTT	Geetest	NetEase	Dingxiang
Regular geometries	√	√	√	√
Chinese characters	√			
English letters	√		√	√
Digits	√		√	

Motivation：Is visual reasoning CAPTCHAs safe enough?

- From the perspective of AI problems
  - Considering visual reasoning tasks
- From the perspective of CAPTCHA
  - Considering traditional cracking methods

03 Holistic Method

- **Input Module** is designed to extract semantic features and global visual features.
- **Reasoning Module** is designed to determines which parts of the text instruction and the global visual feature vector are the most relevant to each reasoning step, follows the working principle of the MAC cell.
- **Output Module** is designed to predict the probability distribution over all candidate grid cells.



03 Holistic Method

Experiment results

Proportions and success rates of different answer questions (in VTT).

Answer object	Proportion	Success rate
Regular geometries	35.5%	78.5%
Chinese characters	30.2%	32.9%
English letters	18.2%	83.6%
Digits	16.1%	76.2%
Total	100.0%	67.3%

Attack results for different visual reasoning CAPTCHAs.

	VTT	Geetest	NetEase	Dingxiang
Success rate	67.3%	66.7%	77.8%	86.5%

03 Holistic Method

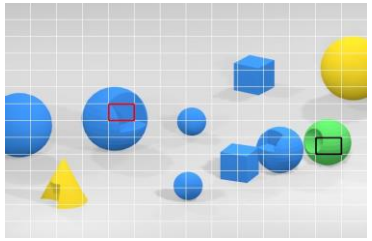
Robustness analysis

- Robustness to higher visual logical complexity
  - Extend the number of reference objects to 2 and 3
    - "Please click the blue cube that is on the right of the *blue cone*"
    - "Please click the green cone that is on the right side of the *green cone* left of the *red cube*"

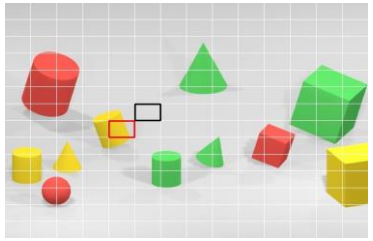
- Robustness to new object categories
  - Regard Chinese character classes as individual categories
    - Train without Chinese samples used in the base experiment
    - Train with Chinese samples involved in 100 classes and the instructions that all based on common attributes rather than abstract attributes of Chinese characters.

03 Holistic Method

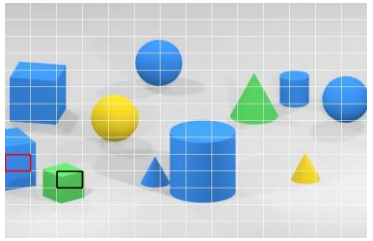
Failure case analysis



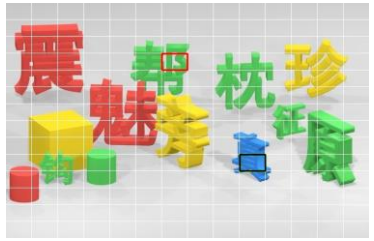
Q1: Please click the object with a notch on the right



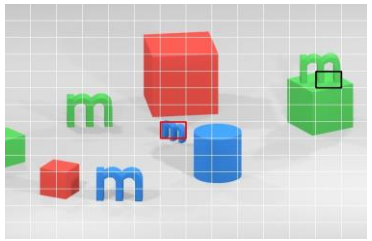
Q1: Please click the object tilting to the left



Q1: Please click the closest blue cube to you

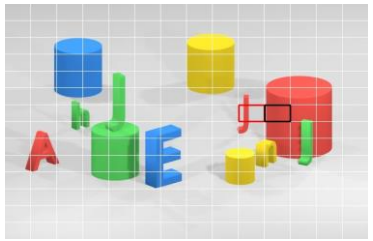


Q1: Please click the Chinese character with pronunciation ˊbangˊ



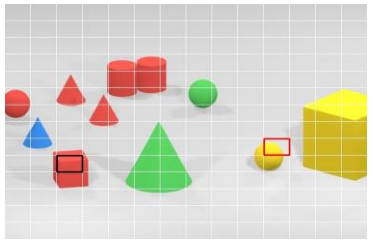
Q2: Please click the letter that is side facing to you

a. Classification error



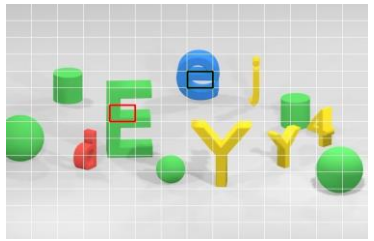
Q2: Please click the letter facing you

b. Grid prediction error



Q2: Please click the object with the same color as the biggest cube

c. Semantic parsing error



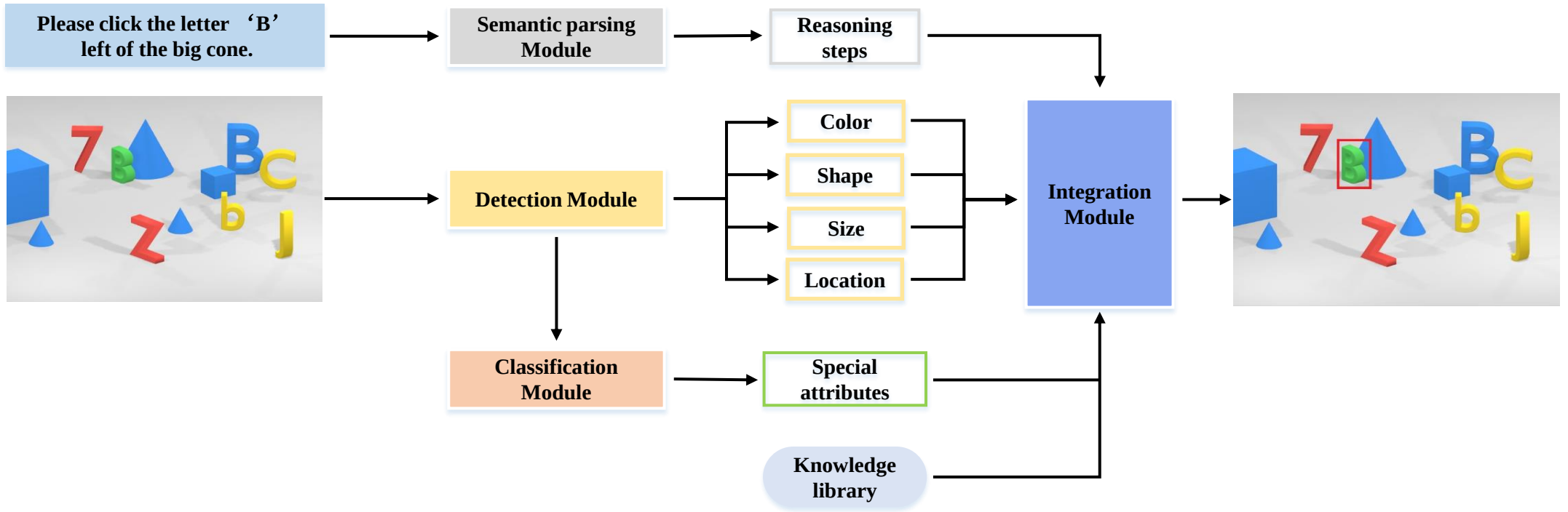
Q2: Please click the uppercase of the blue letter

d. Abstract attributes error

- Classification error
  - The model learns the features corresponding to some subtle attributes that appear only in relation to specific challenges.
- Grid prediction error
  - The model incorrectly outputs a grid cell that is close but not identical to the answer grid cell.
- Semantic parsing error
  - The model fails to extract the logical relationships expressed in the natural language instructions.
- Abstract attribute error
  - The numbers of classes of synonyms or antonyms, pronunciations, components, and other attributes are larger on Chinese-based CPATCHAs.

04 Modular Method

- **Semantic parsing Module** breaks down the raw text instruction into the corresponding reasoning procedures.
- **Detection Module** is to locate the positions of all foreground objects.
- **Classification Module** is to recognize subtle visual attributes such as notches, fractures, tilt directions and character categories.
- **Integration Module** is to carry out a sequence of program-based filtration operations to obtain the predicted answers.



04 Modular Method

Results of modular method

Results of modular attack of visual reasoning CAPTCHAs.

Answer object	Semantic Parsing Module	Detection Module	Classification Module	Attack Success Rate
Regular geometries	100%	93.0%	90.0%	99.0%
Chinese characters	100%	96.6%	82.7%	80.0%
English letters	100%	98.5%	93.8%	83.7%
Digits	100%	99.0%	96.3%	94.7%
Overall accuracy	100%	95.0%	88.8%	88.0%

	Semantic Parsing Module	Detection Module	Attack Success Rate
Geetest	100%	95.7%	90.8%
NetEase	100%	93.5%	86.2%
Dingxiang	100%	95.2%	98.6%

05 Guideline and Future Direction

Using a larger category set

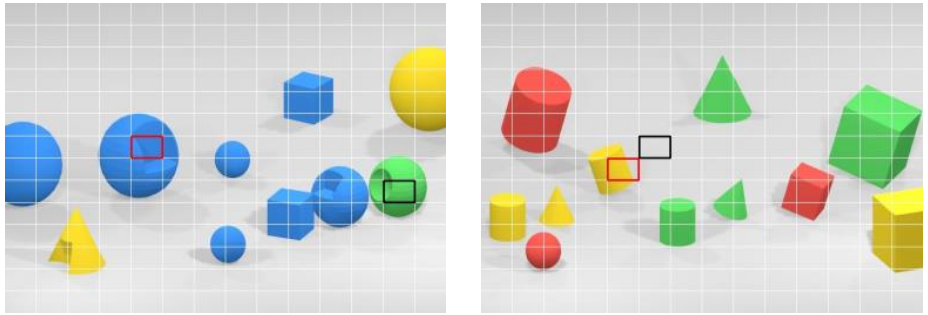
- Using more categories in CAPTCHA design results in a larger theoretical solution space that a malicious bot must search and thus provides better security.

The attack success rates of adding more categories.

	50 classes	100 classes
Attack success rate	77.7%	69.7%

Using more variations

- Variation refers to objects in the same category that appear subtly different but remain the same in their main outline and basic features.
  - raising the difficulty for a model in recognizing the object category
  - recognizing these attributes themselves is even more challenging for a model than the classification task.



Q1: Please click the object with a notch on the right

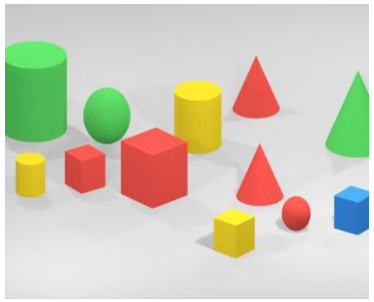
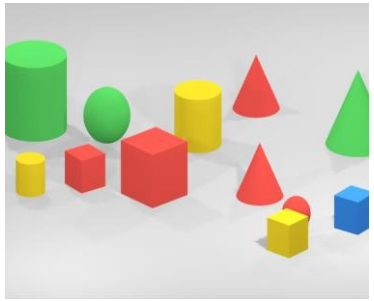
Q1: Please click the object tilting to the left

Failure cases

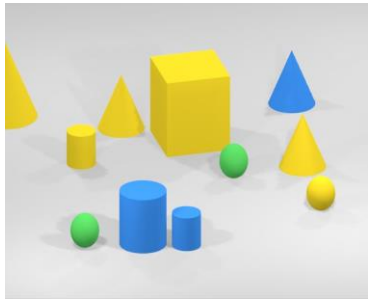
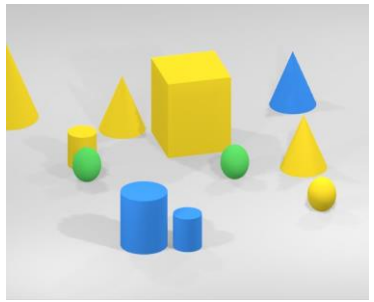
05 Guideline and Future Direction

Making some occlusion

- Occlusion refers to the case in which the view of an object is partially blocked by another object. Making some occlusion will enhance the security of CAPTCHAs.

Please click the red sphere

Please click the yellow cylinder

The attack success rate and human pass rate under different occlusion settings.

	No Occlusion	Occlusion
Attack Success Rate	86.0%	69.7%
Human Pass Rate	93.9%	92.9%

05 Guideline and Future Direction

Commonsense knowledge

- Abstract concepts can be regarded as a type of commonsense knowledge. The inability of the holistic model to address abstract concepts resulted in 81.9% of its failures on VTT tests based on Chinese characters. And the modular method can solve only a limited subset of challenges based on abstract concepts.

Error distribution(%) for the holistic method.

Answer object	Abstract Attribute Error
Regular geometries	0%
Chinese characters	81.9%
English letters	45.7%
Arabic numerals	38.3%

- The body of commonsense knowledge held by humans is nearly infinite. All these experimental results show that solving problems based on commonsense knowledge is indeed a complex task for current machine learning and deep learning algorithms.
- The high abstractness and infinite scope of commonsense knowledge greatly increase the problem complexity for a machine.

06 Conclusion

- Explored the hard AI problems underlying current existing CAPTCHAs and found that conventional CAPTCHA schemes have been proven to be insecure.
- Comprehensively studied the security of four visual reasoning schemes that proved the latest effort to use novel, hard AI problems (visual reasoning) for CAPTCHAs has not yet succeeded.
- Further summarized three guidelines for future vision-related CAPTCHA design.
- The adoption of commonsense knowledge in CAPTCHA design has promising prospects.

Thank you

西安电子科技大学
XIDIAN UNIVERSITY

hchgao@xidian.edu.cn