

Yet Another *text* CAPtCha Solver

叶贵鑫

西北大学网络与信息安全实验室



Work in Collaboration with



Zhanyong Tang
Dingyi Fang
Xiaojiang Chen
Pengfei Xu



Zheng Wang



Zhanxing Zhu
Yansong Feng



西北大学肇始于1902年的陕西大学堂和京师大学堂速成科仕学馆。

1912年始称西北大学。

1923年改为国立西北大学。

1950年复名西北大学。



现为首批国家“世界一流学科建设高校”，国家“211工程”建设院校、教育部与陕西省共建高校。

Our Team



房鼎益



陈晓江



汤战勇

无线网络 & 系统安全



实验室近三年主要成果

• 2018年

SIGCOMM'18 1篇、CCS'18 1篇、MOBICOM'18 2篇、
AAAI'19 1篇、UbiCom'18 2篇、IJCAI'18 3篇、
PerCom'18 1篇、SECON 1篇

• 2017年

NDSS'17 1篇、MOBICOM'17 2篇、INFOCOM'17 1篇、
CoNext'17 1篇、ICDCS'17 1篇

• 2016年

MOBICOM'16 1篇、INFOCOM'16 2篇、ICNP'16 1篇

Captchas are WIDELY used

Log in Wikipedia

Incorrect username or password entered. Please try again.

Username

fanzixi

Password

Enter your password

Keep me logged in (for up to 365 days)

To protect the wiki against automated password cracking, we kindly ask you to enter the words that appear below in the box ([more info](#)):

CAPTCHA Security check

rendsvises

[Refresh](#)

Enter the text you see on the image

Log in

Microsoft

Recover your account

We can help you reset your password and security info. First, enter your Microsoft account and follow the instructions below.

yeguix@hotmail.com



New

Audio

Enter the characters you see

Cancel

Next

Sign In Alipay

Account Name:

Password:

[Forgot your password?](#)

Code:

FWK [Change](#)

The characters are case-insensitive.

Sign In

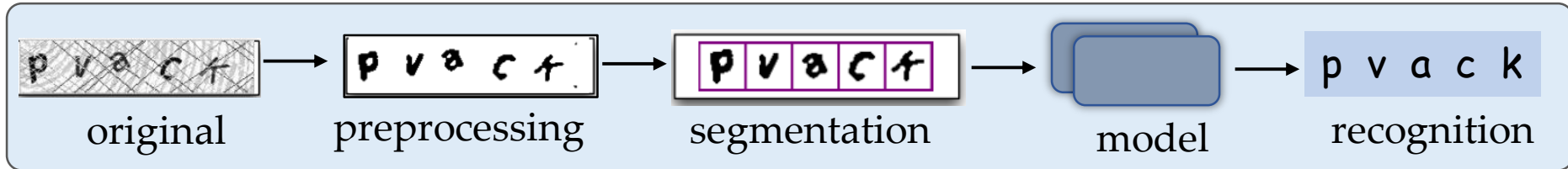
[Sign Up for Free](#)

32 of

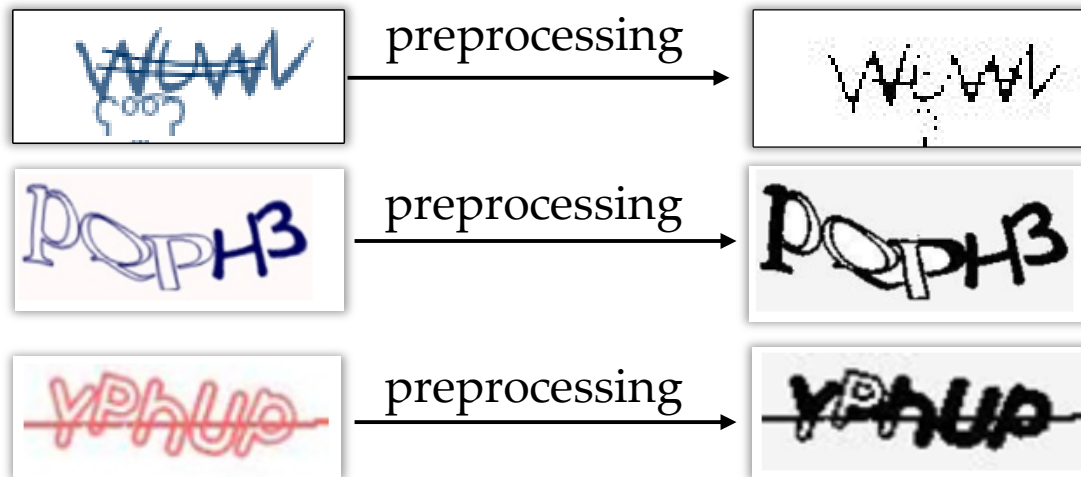
TOP-50

websites

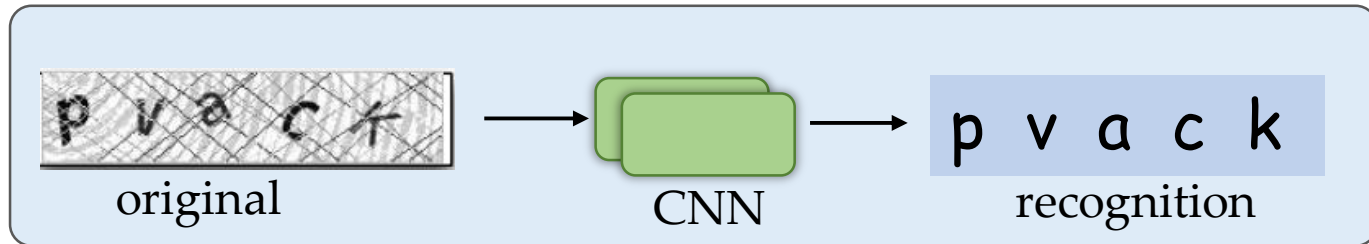
Model-based Approach



Extensive Human Involvements



Deep Learning based approach



Scheme	Example	Training Set	Acc.
reCaptcha		2.3 M	89.9%
PhotoOCR		7.9 M	1.9%
Baidu		50K	0.1%
Google		50K	0%

Getting A Large Number of Captchas is difficult



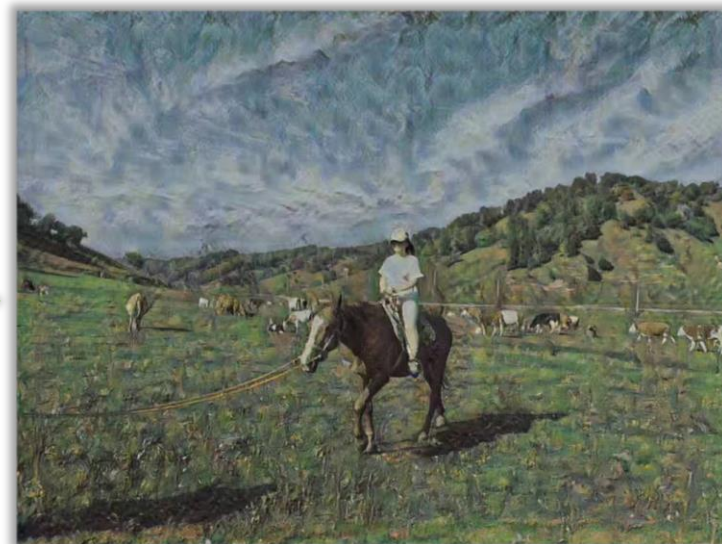
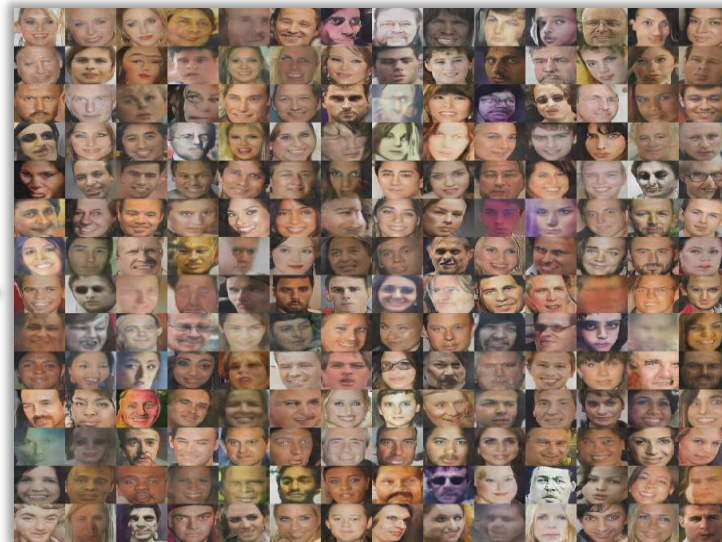
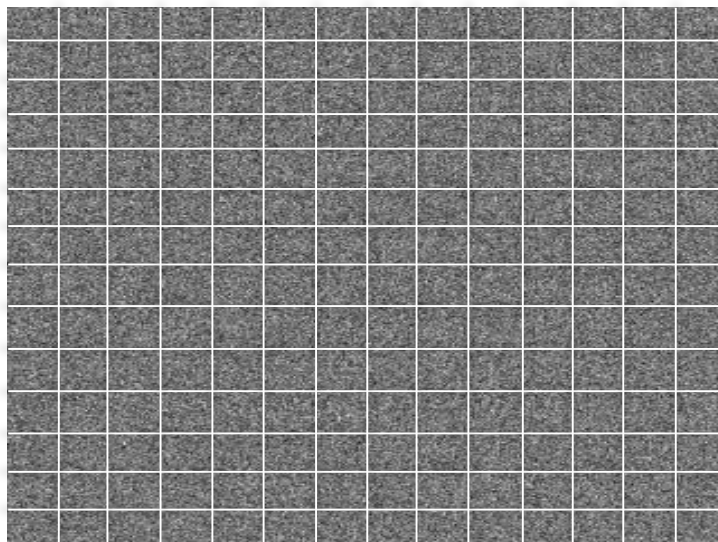
Anti-crawling mechanisms



Frequent changing
captchas styles



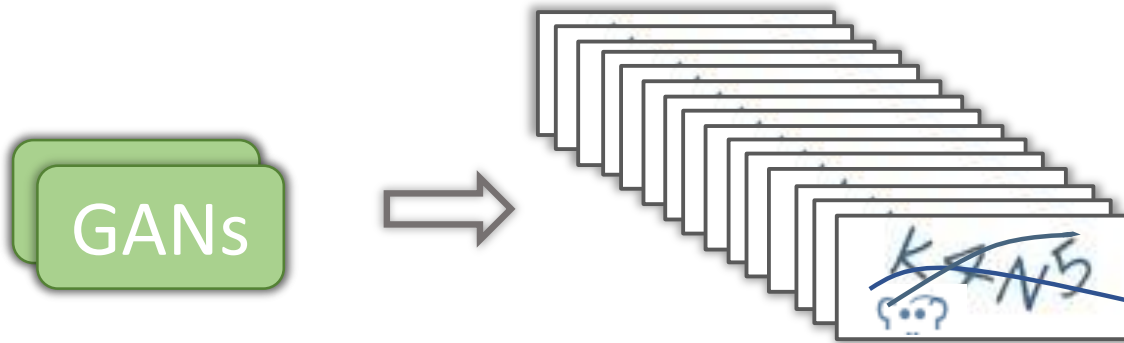
Inspired by GANs





Transfer to Captcha

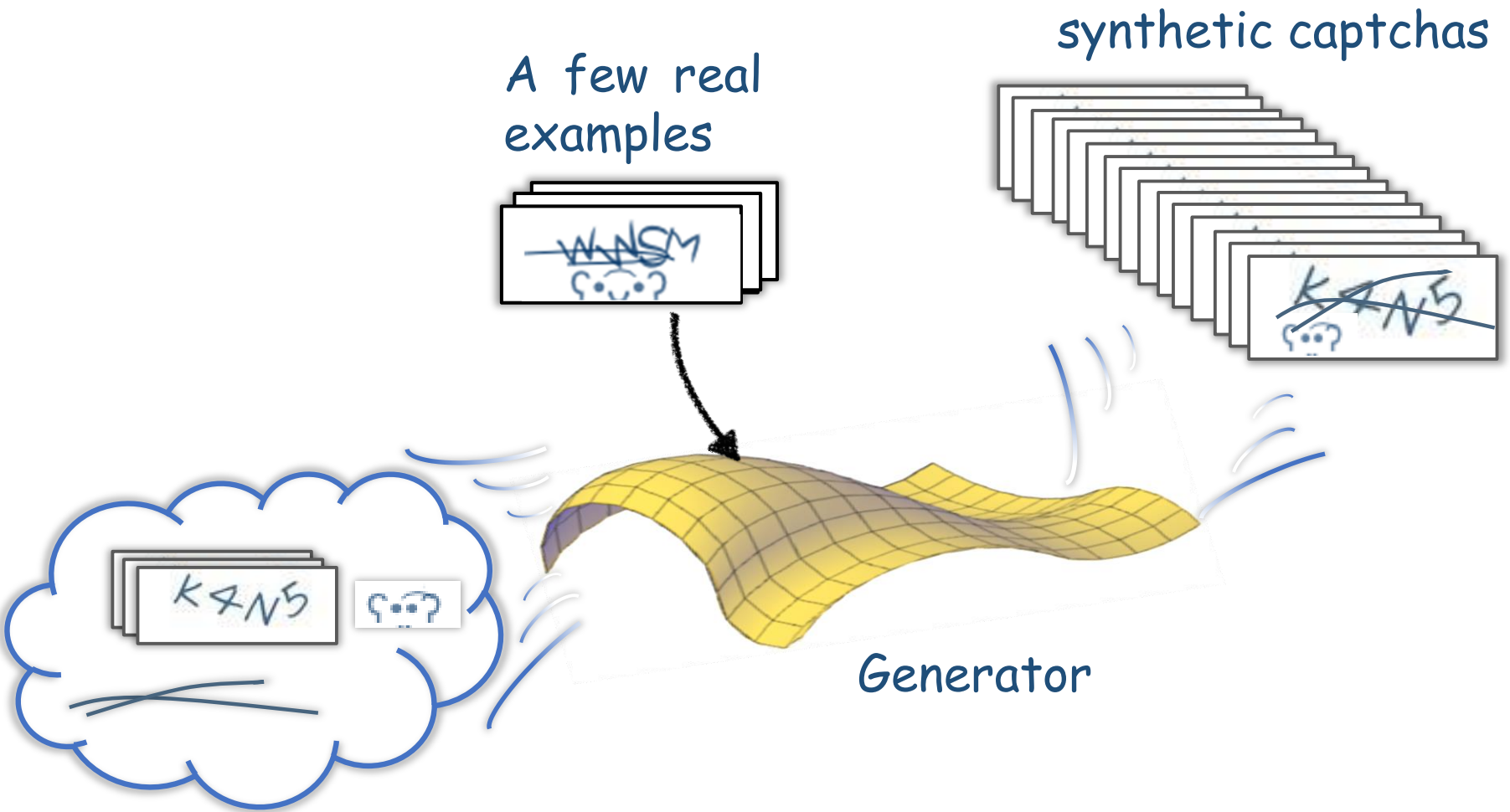
□ Synthesize Captchas



□ Remove Security Features

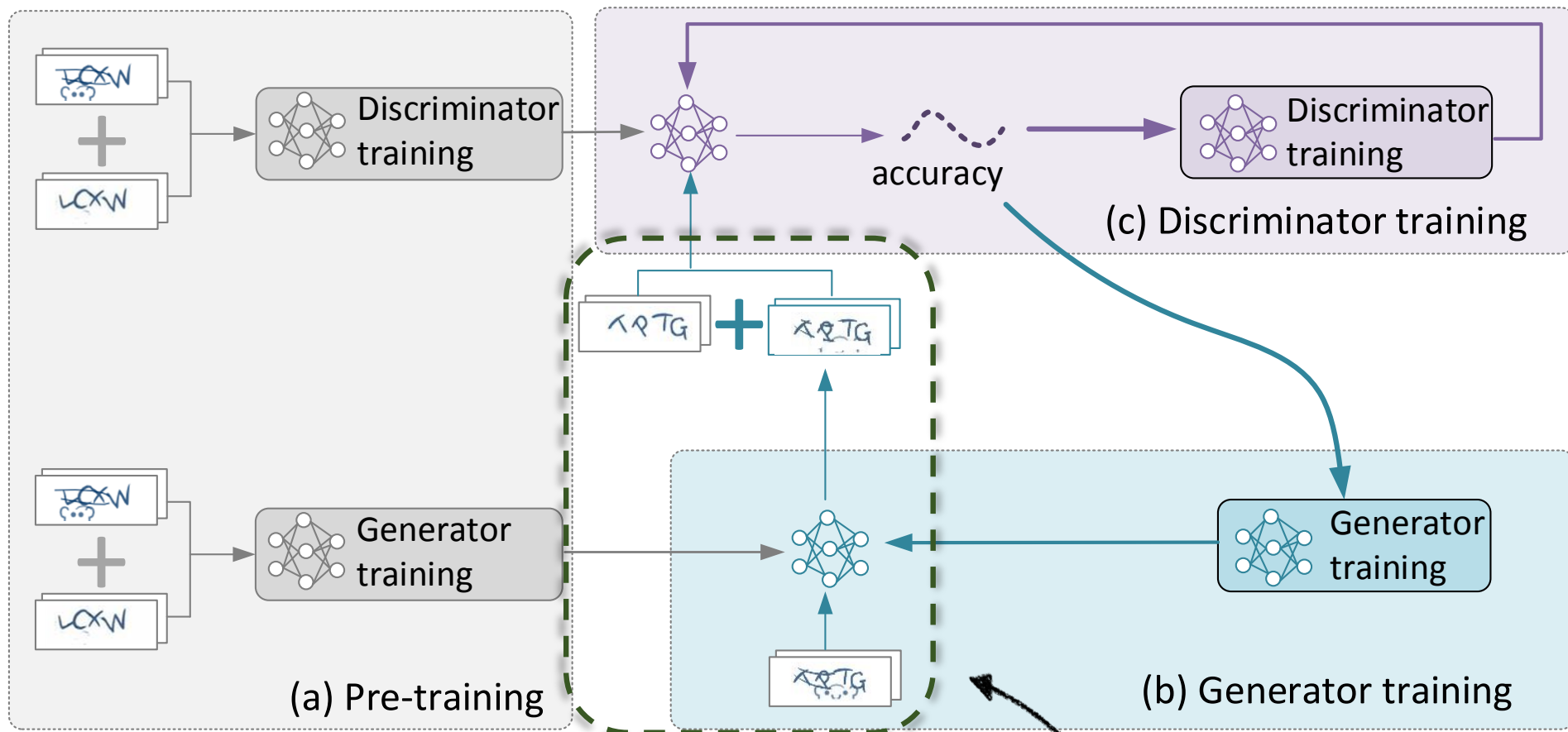


Our Approach



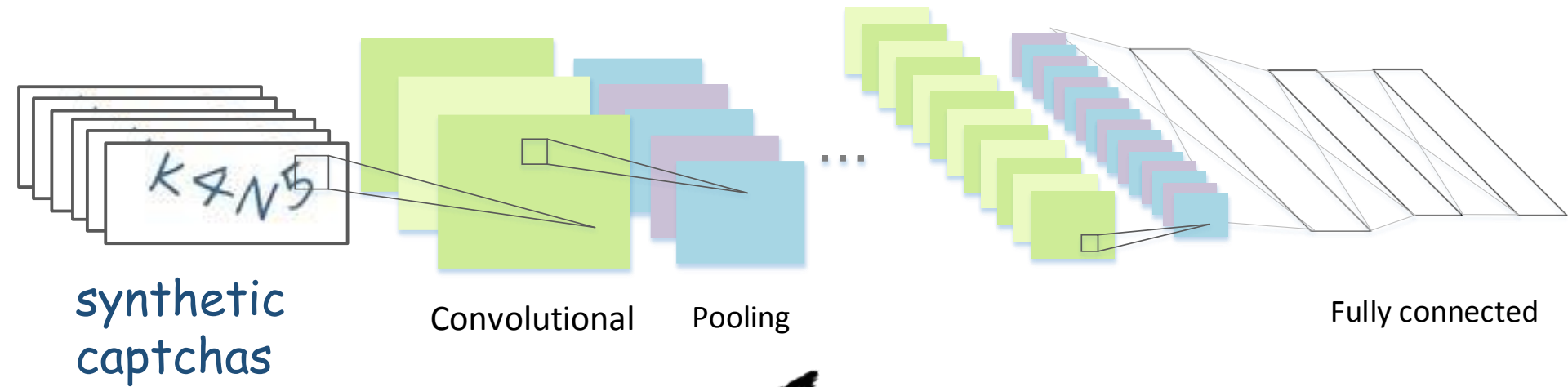
Automatically Generate Millions of synthetic captchas

Our Approach



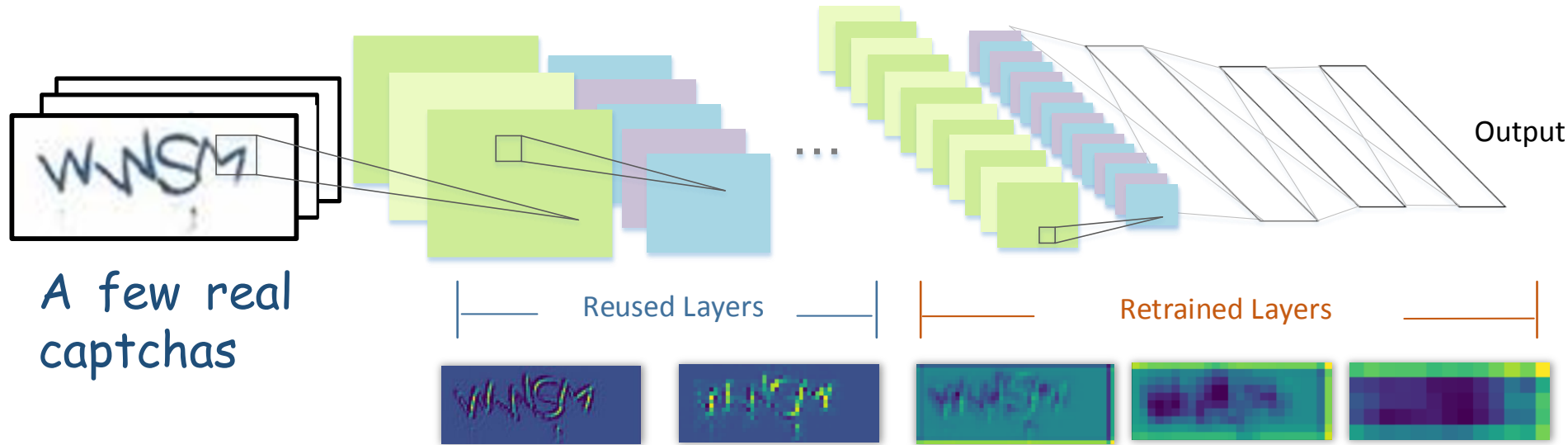
Build the Pre-process Model using the synthetic captchas

Our Approach



Base solver was built using synthetic captchas

Our Approach



Build the fine-tuned solver

Reason for Tuning **Base solver**

Synthetic domain **Target domain**

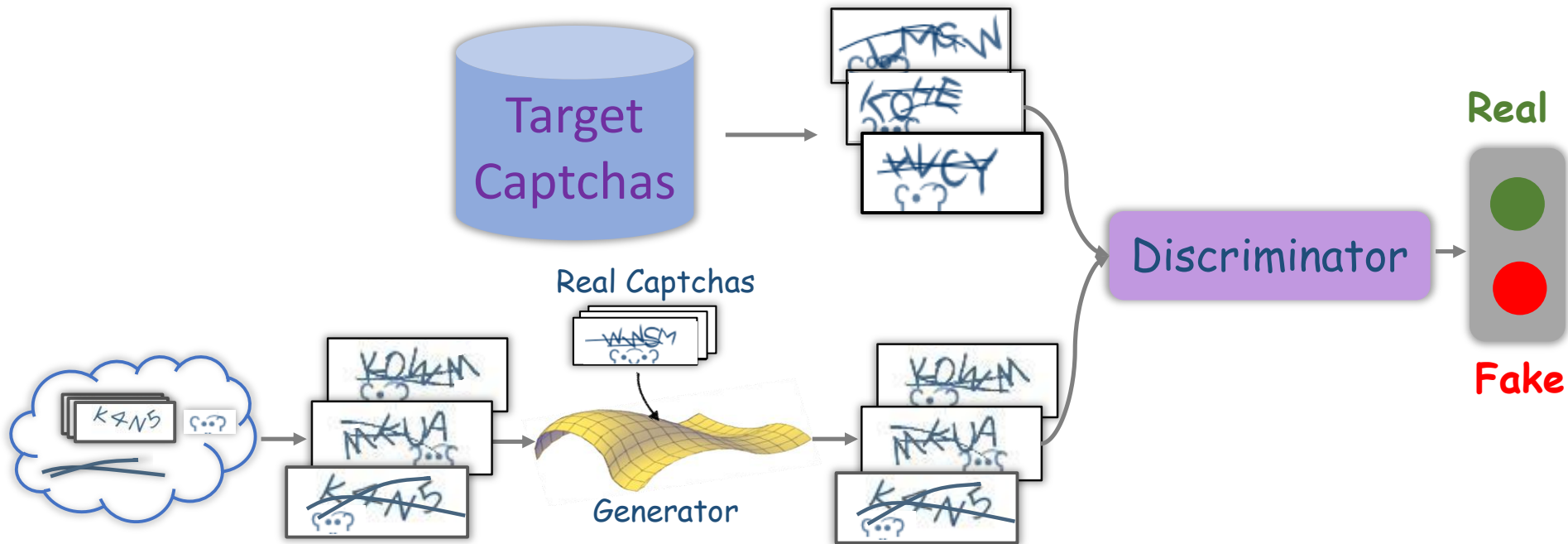


A diagram illustrating the difference between real and synthetic handwritten characters. It features a light blue rounded rectangle containing three rows of examples. Each row has a 'Real' example on the left and a 'Synthetic' example on the right. The 'Real' examples are in green italics, and the 'Synthetic' examples are in blue italics. The synthetic examples show subtle differences in stroke thickness and shape compared to the real ones. Three orange boxes with white text are placed between the columns: 'M and W' at the top, 'K' in the middle, and 'V' at the bottom.

Real	Synthetic
<i>M and W</i>	<i>M and W</i>
<i>K</i>	<i>K</i>
<i>V</i>	<i>V</i>

Synthetic captchas are subtle difference from the real ones.

Generative Adversarial Network based Synthesizer



Make sure synthetic captchas are similar to the target ones

Captcha Schemes Used

- **11** from 32 of the top-50 websites

ebay

Google

Microsoft



360
WWW.360.CN

Alipay™

weibo

sina 新浪网
sina.com.cn

Baidu 百度

JD.COM

搜 狐
SOHU.com

- Prior **22** schemes used by other works

Megaupload, Blizzard, Authorize, Captcha.net, NIH, Reddit, Digg, Slashdot, QQ, Taobao, reCAPTCHA(2011), reCAPTCHA(2013), Amazon, Microsoft(2016), Yahoo!(2016), Yahoo!(2014), Baidu(2016), Baidu(2013), Baidu(2011), CNN, Papal, Sina(2016)

All captchas have multiple complex security features



eBay



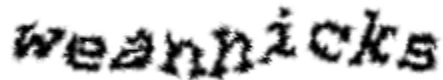
Google



Microsoft



Sohu



Wikipedia



JD



Weibo



Alipay



Sina



Baidu



Qihu360

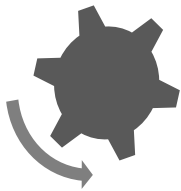
Excluded Characters

Scheme	Used Characters	Excluded Characters
Sohu	a-z, 0-9	0,1, i,l,o,z
JD	A-Z, 0-9	0,1,2,7,9, D,G,I,J,L,O,P,Q,Z
Microsoft	A-Z, a-z, 0-9	0,1,5, D,G,I,Q,U
Alipay	A-Z, 0-9	0,1, I,L,O
Qihu360	A-Z, a-z, 0-9	0, I,L,O,T, i,l,o,t,q

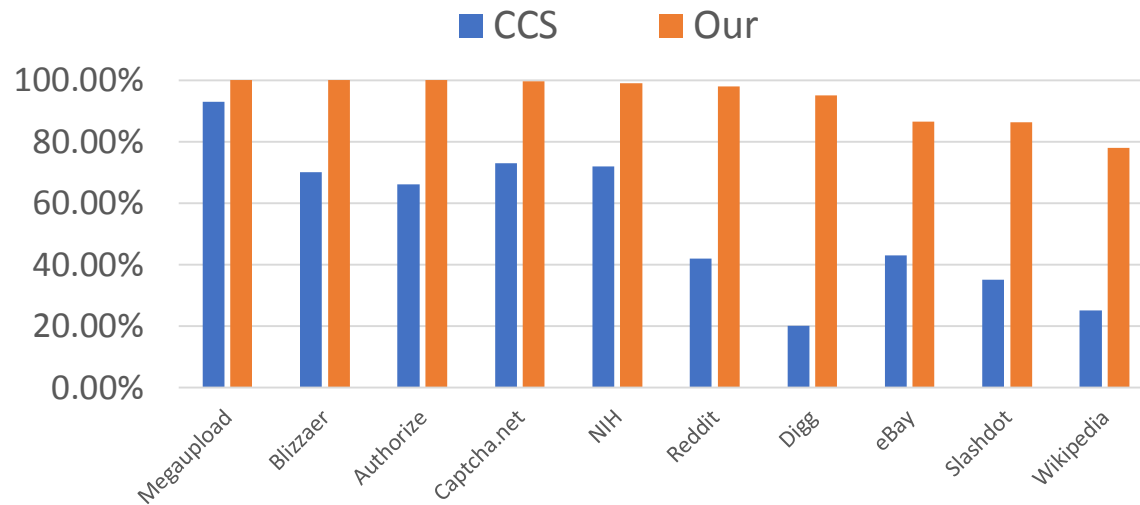
High Success Rate

#	Scheme	Before tuning	After Tuning
1	Sohu	83%	92%
2	Ebay	52%	86.6%
3	JD	60%	86%
4	Wikipedia	7%	78%
5	Microsoft	36.6%	69.6%
6	Alipay	23%	61%
7	Qihu360	48.5%	56%
8	Sina	40.6%	52.6%
9	Weibo	4.7%	44%
10	Baidu	6%	34%
11	Google	0%	3%

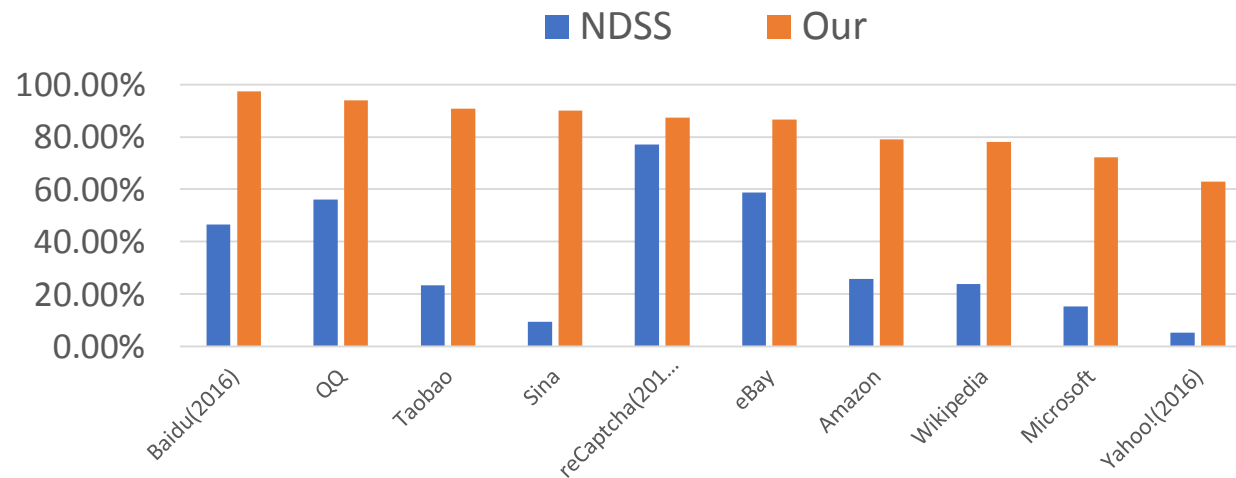
Better Performance



Comparing to CCS'11



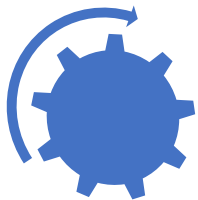
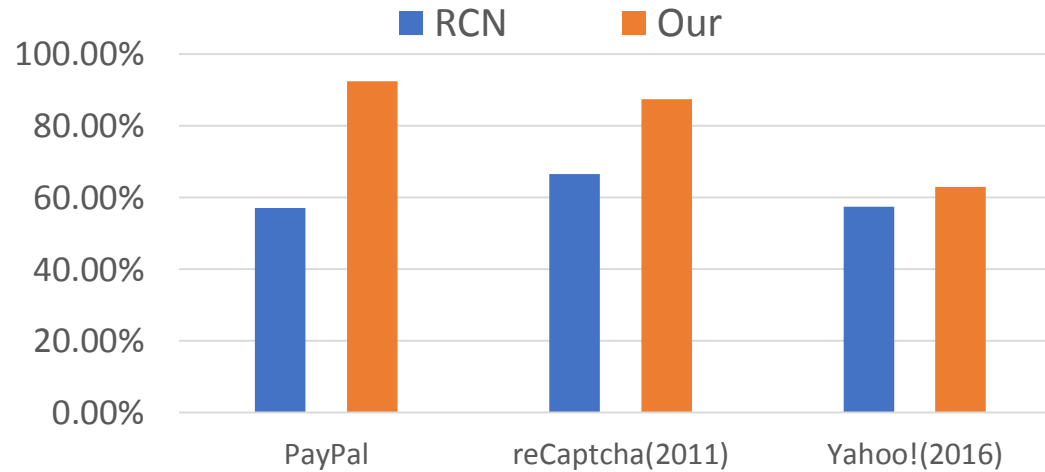
Comparing to NDSS'16



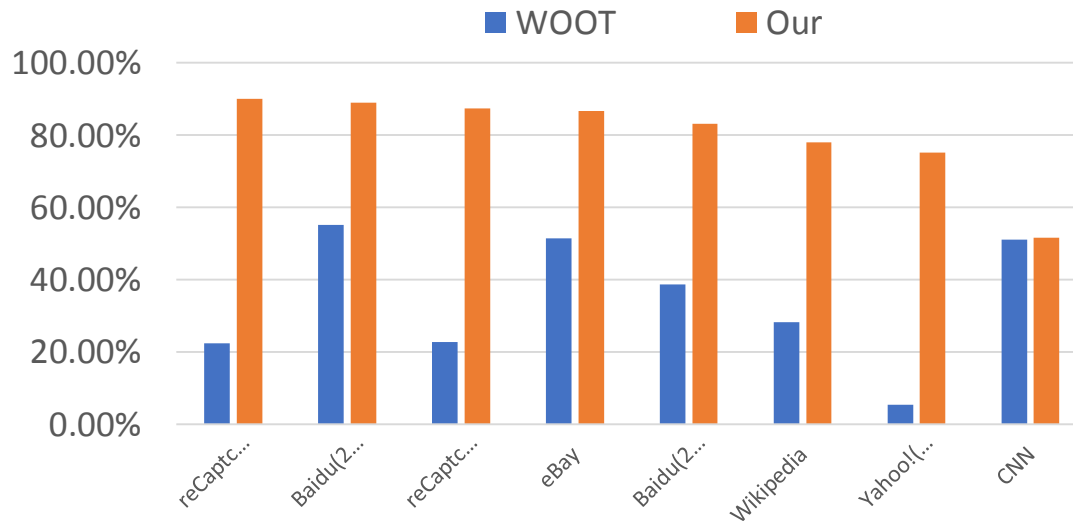
Better Performance



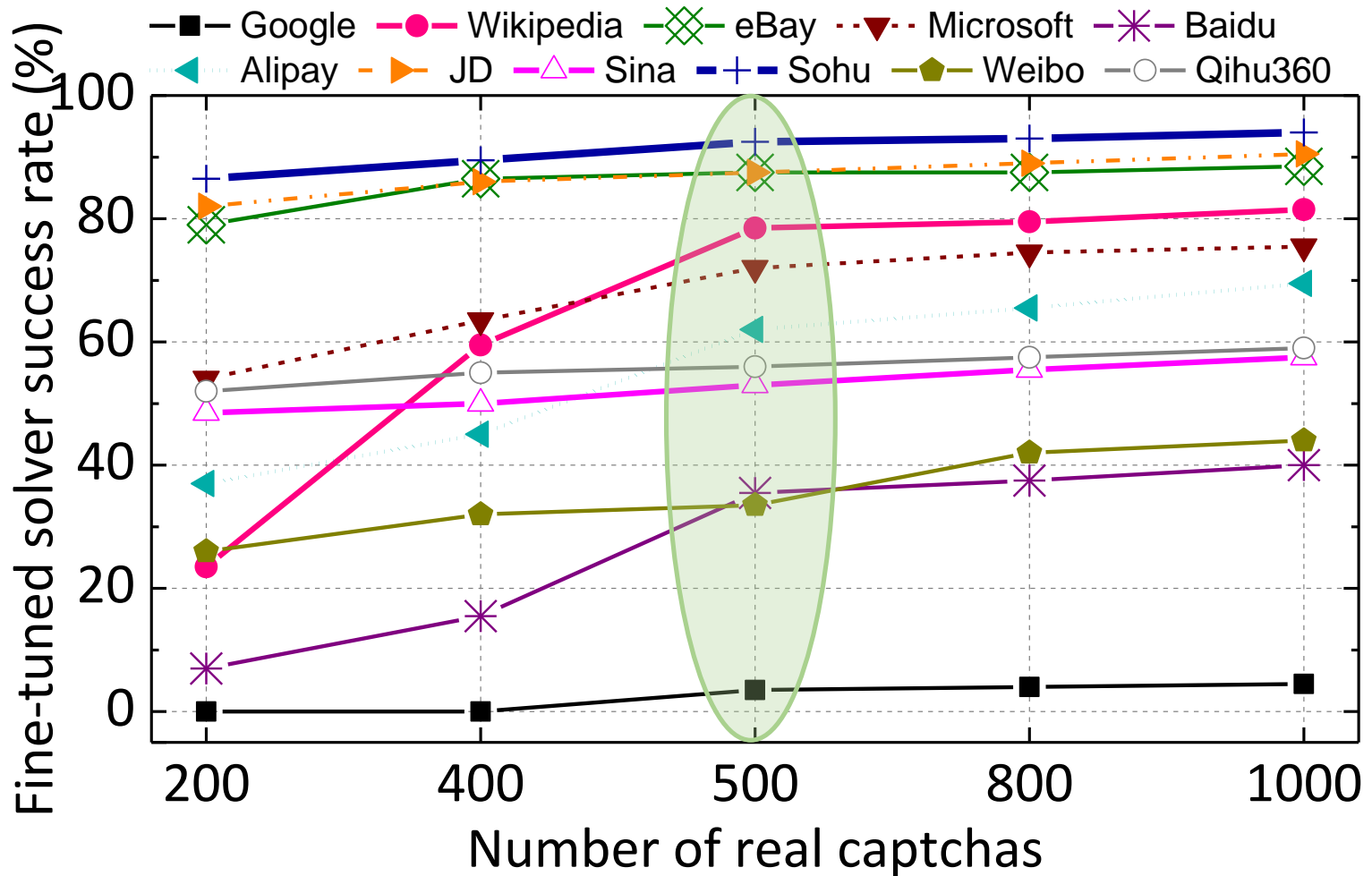
Comparing to Science'17







Comparing to Usenix WOOT'14

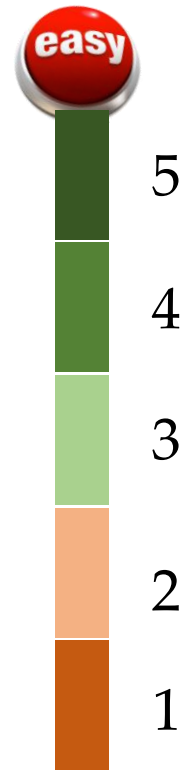


500 real captchas are enough

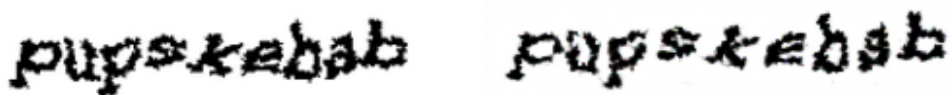


User Study

No.	Example	Success Rate		Usability Score
		Human	Our approach	
1	tah8	95.25%	100%	4
2		90.25%	88%	2.75
3		91%	96%	2.8
4		89.25%	86%	2.7
5	g/oborick	79.75%	77%	2.8
6		68.75%	40%	2.1



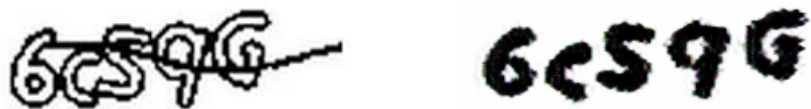
GAN-based Segmentation Approach



(a) Wikipedia captcha and its pre-processing version



(b) Microsoft captcha and its pre-processing version



(c) Sina captcha and its pre-processing version



(d) Baidu captcha and its pre-processing version

Demo

captcha_solver - [E:\Research Interests\我的论文\CCS-Captcha\通讯稿\演示相关\captcha_solver] - ...\solver.py - PyCharm Community Edition 2016.3.2

File Edit View Navigate Code Refactor Run Tools VCS Window Help

captcha_solver \ solver.py

solver

solver.py x solver-wiki.py x action_chains.py x base_options.py x test_options.py x util.py x

```
1 from selenium import webdriver
2 from selenium.webdriver.common.by import By
3 from selenium.webdriver.common.action_chains import ActionChains
4 from selenium.webdriver.common.keys import Keys
5 import win32api
6 import win32con
7 import time
8
9 from options.test_options import TestOptions
10 import os
11 from models.models import creat_model
12 from util import util
13
14 import shutil
15
16
17 VK_CODE = {'enter':0x0D, 'down_arrow':0x28}
18
19
20 # Press down
21 def KeyDown(keyName):
22     win32api.keybd_event(VK_CODE[keyName], 0, 0, 0)
23
24
25 # Up
26 def KeyUp(keyName):
27     win32api.keybd_event(VK_CODE[keyName], 0, win32con.KEYEVENTF_KEYUP, 0)
28
29 download_img = "C:/Users/gxye/Downloads/checkcode.png"
30 data_dir = "E:/attack_datasets"
31
32 # load model and weight
33 opt = TestOptions().parse()
34 model = creat_model(opt)
```

- Copy Reference Ctrl+Alt+Shift+C
- Paste Ctrl+V
- Paste from History... Ctrl+Shift+V
- Paste Simple Ctrl+Alt+Shift+V
- Column Selection Mode Alt+Shift+Insert
- Find Usages Alt+F7
- Refactor
- Folding
- Go To
- Generate... Alt+Insert
- Run 'solver' Ctrl+Shift+F10
- Debug 'solver'
- Save 'solver'
- Local History
- Execute Line in Console Alt+Shift+E
- Compare with Clipboard
- File Encoding
- Create Gist...

TODO Python Console Terminal

Event Log

Generate constructor, getter or setter method, etc.

10:10 CRLF+ GBK+



COUNTERMEASURES



original captcha



P 3 K P R

ground truth



adversarial example



R 7 X 0 T 

incorrect prediction

Other Findings

MNIST



Single character



Conclusions

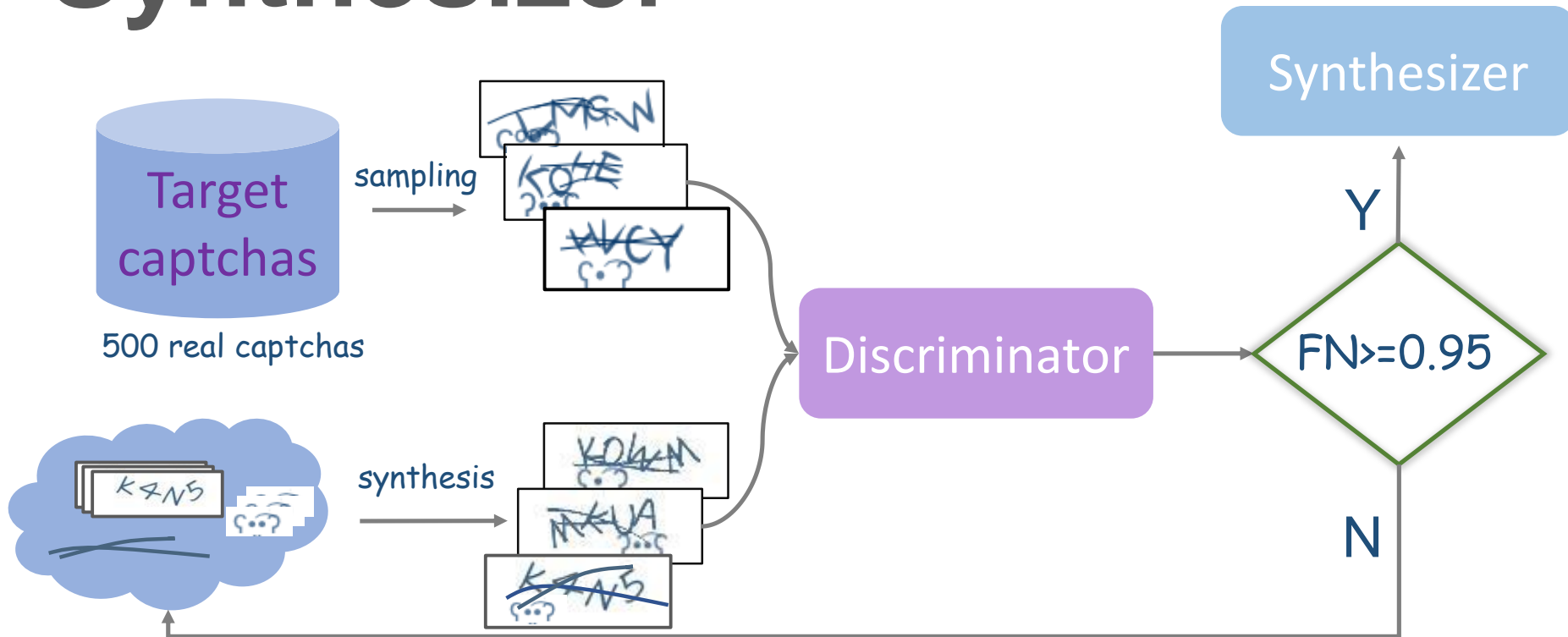
Text based captchas are vulnerable under generative adversarial network based solver.

Our approach needs less human involvements and requires a smaller number of real captchas.

*Thank
You*

Backup Slides

Training the Captcha Synthesizer



Synthesize the captchas whose style is similar to the real ones.

Google Captchas

tuteniske

dringstal

redingsti

delysopp

tenicande

ismidev

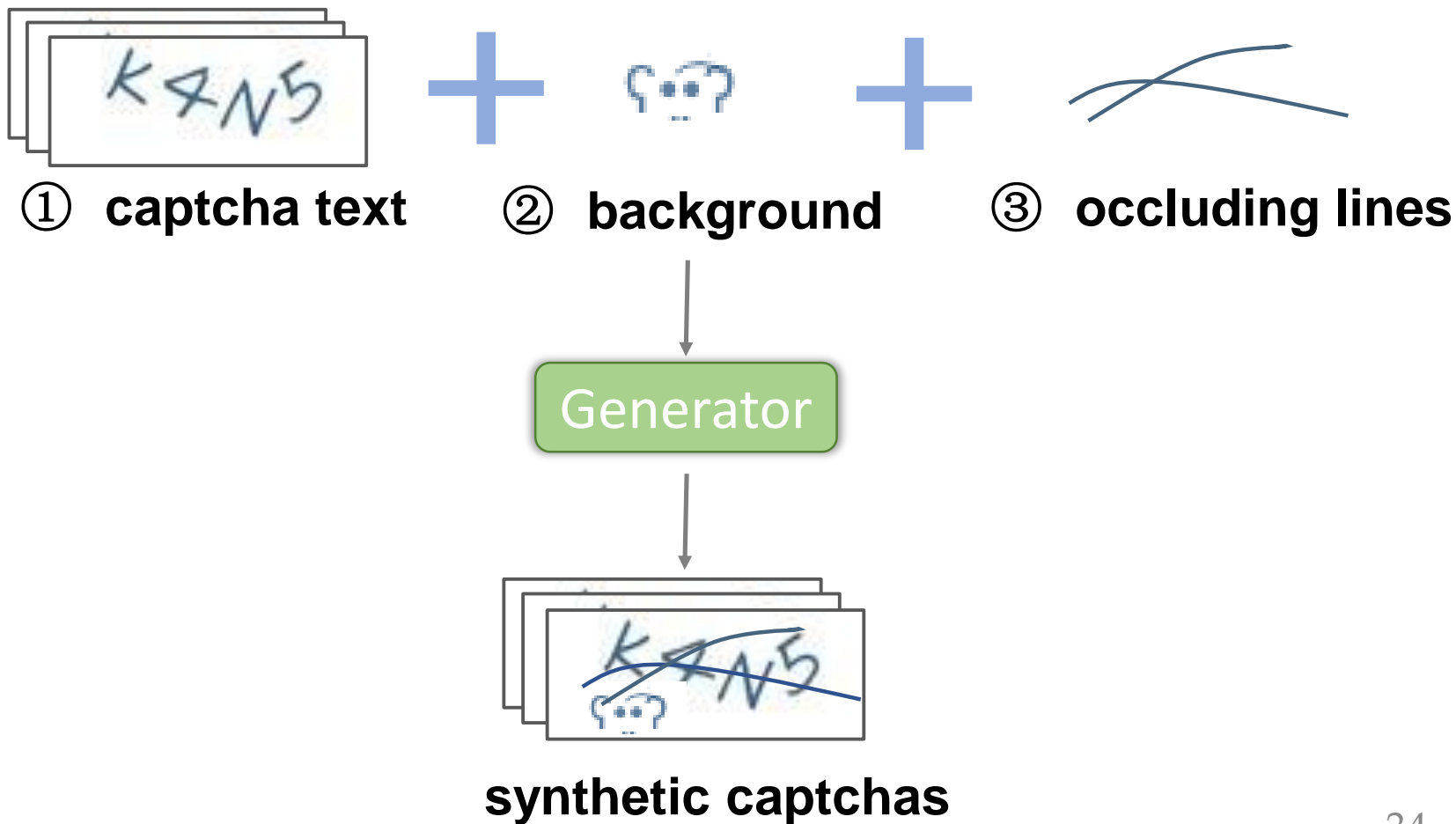
impediscer

nicahers

nationst

Extremely Complex & More than 8 fonts

Generate captchas using security features

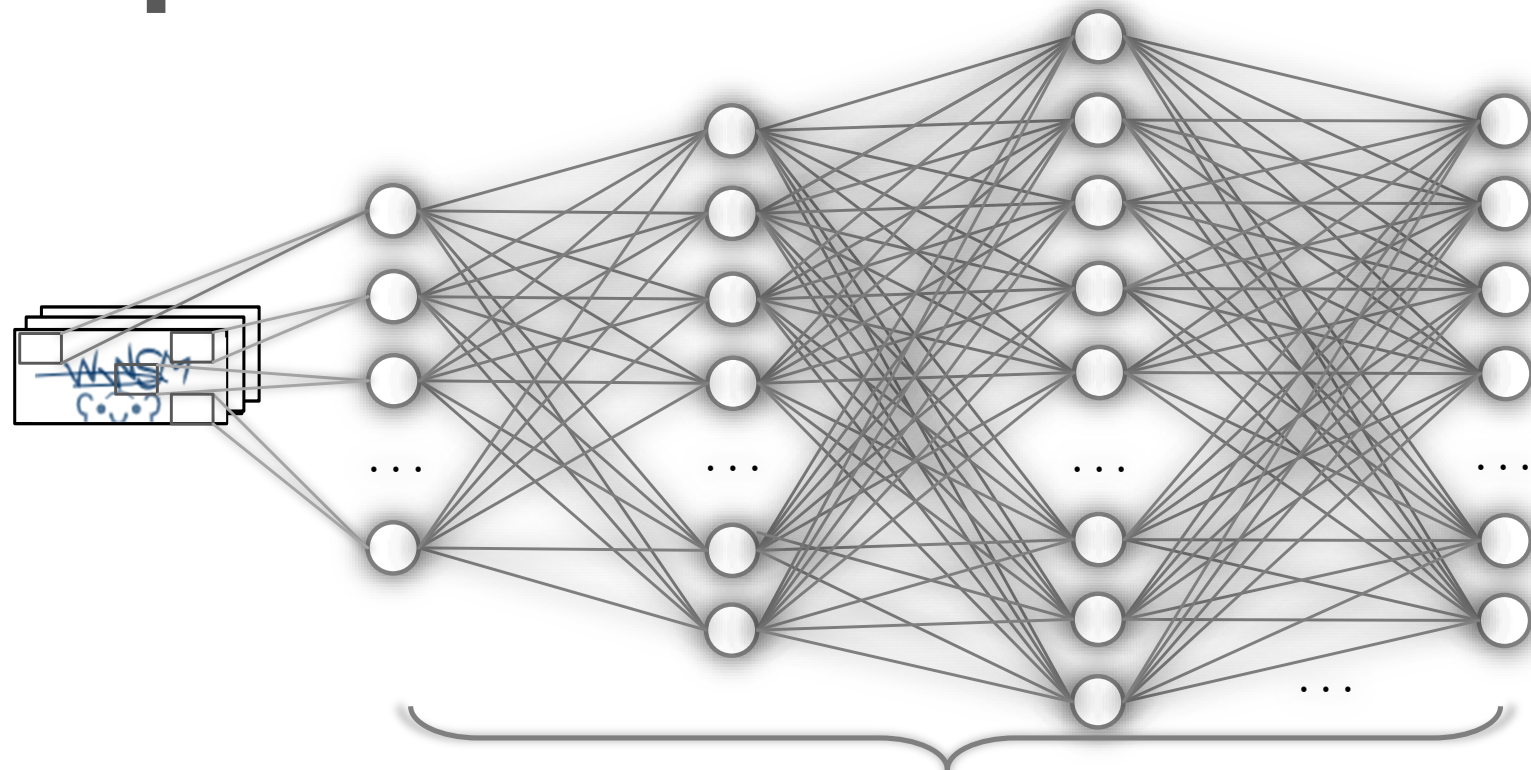


Security features of captcha



Security Feature	On/Off	#Options	Value Range
Noisy background(s)	On	5	[10, img.width]
Occluding lines	On	2	{Line, Sin, Quadratic, Bezieer}
Char. Overlapping	On	-	[-3, 10]
Character set	On	4	[A – Z]
Font style(s)	On	1	Solid
Font color(s)	On	1	RGB (65, 103, 141)
Distortion	On	-	{[0.1, 0.2], [0.2, 0.3]}
Rotation	On	-	[-30, 30]
Waving	Off	-	-

Captcha Solver Network



[32, 64, 128, 256, 512, 3072] nodes

LeNet-5 + 2x Conv. + 3x Pooling

$$y = f(x)$$

text → captcha