

Forum:[Hack Forums](#)

Thread: [\[Showoff\] Facebook Captcha OCR](#)

Ninja - June 7, 2015, 6:48 pm

So I was working on a Facebook cracker, but a dictionary attack was impossible due to the captchas that show up after 10 attempts. I wrote a function to remove the distortion from the images and then trained Tesseract a bit so it can read the processed images better. The accuracy was around %20, but since Facebook still accepts the captcha response if 1 character is incorrect or missing, the total accuracy is around %40.

Example of an input image:

Obviously Tesseract would not be able to read the text from the image before processing.

Example of a processed image:

Tesseract successfully read this image as Xp7bad.

The method I used to process the images is pretty simple. I made a 'selection' of all black colored areas and shrunk them, and then resized them. By shrinking enough all lines are removed, but not the letters as they are bigger. And by enlarging the selections again the letters come back to their original size so they can be read easily.

Combining this with Tor networking so I can get 'unlimited' high anonymity, high speed proxies from the Tor network, I can make one fast, 'proxyless' Facebook cracker. If I can't sell this cracker to anyone for a reasonable price I will probably release the code, including the OCR and image processing code in this section for free.

Also made a captcha OCR library for Amazon. No processing was needed since they are already easily readable. Just a bit of training and the accuracy was around %95. Amazon is proxyless as well. ?

Abhax - June 7, 2015, 6:52 pm

(Jun 7, 2015 06:48 PM)Tyranitar Wrote: [\[CITE\]](#) So I was working on a Facebook cracker, but a dictionary attack was impossible due to the captchas that show up after 10 attempts. I wrote a function to remove the distortion from the images and then trained Tesseract a bit so it can read the processed images better. The accuracy was around %20, but since Facebook still accepts the captcha response if 1 character is incorrect or missing, the total accuracy is around %40.

Example of an input image:

Obviously Tesseract would not be able to read the text from the image before processing.

Example of a processed image:

Tesseract successfully read this image as Xp7bad.

The method I used to process the images is I made a 'selection' of all black colored areas and shrunk them, and then resized them. By shrinking enough all lines are removed, and by enlarging again the letters come back to their original size so they can be read easily.

Combining this with Tor networking so I can get 'unlimited' high anonymity, high speed proxies from the Tor network, I can make one fast, 'proxiless' Facebook cracker.

Also made a captcha OCR library for Amazon. No processing was needed since they are already easily readable. Just a bit of training and the accuracy was around %95. Amazon is proxiless as well. ?

Awesome! I can't wait to see the outcome of this project!

Ninja - June 7, 2015, 6:53 pm

(Jun 7, 2015 06:52 PM)Abhax Wrote: [\[CITE\]](#) Awesome! I can't wait to see the outcome of this project! This is the outcome. All I can do is train it so it is more accurate. Or are you talking about the cracker? If all works well I can (hopefully) bruteforce accounts, as the only limitation I have encountered so far is this bloody captcha.

NETExplorer - June 7, 2015, 7:10 pm

I'm surprised a tarantula could make this, let alone an obese one
GGOP

Abhax - June 7, 2015, 7:11 pm

I mean the actual cracker lol.

Only3 - June 7, 2015, 8:44 pm

(Jun 7, 2015 07:10 PM)PatPositron Wrote: [\[CITE\]](#) I'm surprised a tarantula could make this, let alone an obese one
GGOP
its tarantalula or u mean taranalula ?

NETExplorer - June 8, 2015, 8:40 am

(Jun 7, 2015 08:44 PM)Only3 Wrote: [\[CITE\]](#) its tarantalula or u mean taranalula ?
Pretty sure its tarantula

Ninja - June 8, 2015, 9:14 am

(Jun 8, 2015 08:40 AM)PatPositron Wrote: [\[CITE\]](#) Pretty sure its tarantula

Damnit, should have switched my Facebook to private. Now everyone can see my photos on Google. :(

Nicktron - June 8, 2015, 10:26 am

(Jun 8, 2015 09:14 AM)Tyranitar Wrote: [\[CITE\]](#) Damnit, should have switched my Facebook to private. Now everyone can see my photos on Google. :(
At least you have acheived the rank: Family, not forever alone like me. :(

Ninja - June 9, 2015, 10:51 am

(Jun 8, 2015 10:26 AM)Nicktron Wrote: [\[CITE\]](#) At least you have acheived the rank: Family, not forever alone like me. :(
Its okay Nick, one day you will achieve this rank.

Nicktron - June 9, 2015, 10:59 am

(Jun 9, 2015 10:51 AM)Tyranitar Wrote: [\[CITE\]](#) Its okay Nick, one day you will achieve this rank.
:) thanks #hopeforthenope

24/7 - June 9, 2015, 12:22 pm

Impressive work, especially the 95% amazon accuracy ?

Ninja - June 9, 2015, 1:23 pm

(Jun 9, 2015 12:22 PM)24/7 Wrote: [\[CITE\]](#) Impressive work, especially the 95% amazon accuracy ?
They are really easy:

bluntly - June 9, 2015, 1:27 pm

Acutally a pretty nice piece of coding you have there :D I wouldn't know how to do that because im currently still learning VB.NET :D I have a question...Why do you code? I do it for fun but mostly for having that great feeling that everything actually makes sense and you actually DID that :D

Ninja - June 9, 2015, 1:48 pm

(Jun 9, 2015 01:27 PM)StrangerDanger Wrote: [\[CITE\]](#) Acutally a pretty nice piece of coding you have there :D I wouldn't know how to do that because im currently still learning VB.NET :D I have a question...Why do you code? I do it for fun but mostly for having that great feeling that everything actually makes sense and you actually DID that :D

I do it for the same reasons. But also one more reason: money. Call me greedy but people pay good money for such a project.

SIM-1676753 - June 9, 2015, 5:32 pm

(Jun 9, 2015 01:48 PM)Tyranitar Wrote: [\[CITE\]](#) I do it for the same reasons. But also one more reason: money. Call me greedy but people pay good money for such a project.
Can't think of the last time I sold code. ?

Ninja - June 11, 2015, 7:09 pm

(Jun 9, 2015 05:32 PM)Mr. Trvp Wrote: [\[CITE\]](#) Can't think of the last time I sold code. ?
Me neither until a while ago when I started doing projects again. I had this time period of a few months to a year when I did nothing really...

SIM-1676753 - June 11, 2015, 8:08 pm

(Jun 11, 2015 07:09 PM)Tyranitar Wrote: [\[CITE\]](#) Me neither until a while ago when I started doing projects again. I had this time period of a few months to a year when I did nothing really...
Sounds like you're having fun you giant tarantula.

NETExplorer - June 11, 2015, 9:04 pm

(Jun 11, 2015 08:08 PM)Mr. Trvp Wrote: [\[CITE\]](#) Sounds like you're having fun you giant tarantula.
giant, overweight tarantula

SIM-1676753 - June 11, 2015, 9:06 pm

(Jun 11, 2015 09:04 PM)PatPositron Wrote: [\[CITE\]](#) giant, overweight tarantula
Don't correct me you fat imbecile.

idb. - June 12, 2015, 4:29 am

Should post the Amazon OCR. I could use that and hate writing OCR code. :P

Ninja - June 12, 2015, 6:45 am

(Jun 12, 2015 04:29 AM)idb. Wrote: [\[CITE\]](#) Should post the Amazon OCR. I could use that and hate writing OCR code. :P

No processing needed, just use the Tesseract wrapper and make a whitelist for alphanumeric characters. Accuracy should be around %80 already. Train it and you can get to %95-%99.

Nicktron - June 12, 2015, 7:30 am

What is "OCR" short for???

Ninja - June 12, 2015, 7:39 am

(Jun 12, 2015 07:30 AM)Nicktron Wrote: [\[CITE\]](#) What is "OCR" short for???

If you had Google you would know it is short for "Optical Character Recognition".

?Sakura? - June 12, 2015, 7:42 am

This is actually extremely impressive, Nice job

idb. - June 12, 2015, 10:49 am

(Jun 12, 2015 06:45 AM)Tyranitar Wrote: [\[CITE\]](#)

No processing needed, just use the Tesseract wrapper and make a whitelist for alphanumeric characters. Accuracy should be around %80 already. Train it and you can get to %95-%99.

How's it doing against the secondary captcha with the lines going through them and the effects applied?

Nicktron - June 12, 2015, 11:54 am

(Jun 12, 2015 07:39 AM)Tyranitar Wrote: [\[CITE\]](#) If you had Google you would know it is short for "Optical Character Recognition".

I thought this was google?? O.o

Ninja - June 12, 2015, 5:33 pm

(Jun 12, 2015 10:49 AM)idb. Wrote: [\[CITE\]](#) How's it doing against the secondary captcha with the lines going through them and the effects applied?

To remove the lines you can use the same method as I did. Since this background has a gradient, my program doesn't know what to do. This is the outcome of processing your image with it:

With some tweaking it's definitely possible.

(Jun 12, 2015 11:54 AM)Nicktron Wrote: [\[CITE\]](#) I thought this was google?? O.o This forum is certainly treated as such, sadly.

Ninja - June 13, 2015, 6:43 pm

(Jun 12, 2015 07:42 AM)?Sakura? Wrote: [\[CITE\]](#) This is actually extremely impressive, Nice job
Thanks bro. Appreciate the nice words.

-Buck- - June 13, 2015, 6:44 pm

Good thinking. I'd never think to shrink and enlarge but it makes complete sense. Nicely done.