

[登录](#)

TensorFlow识别字母扭曲干扰型验证码-开放源码与98%模型

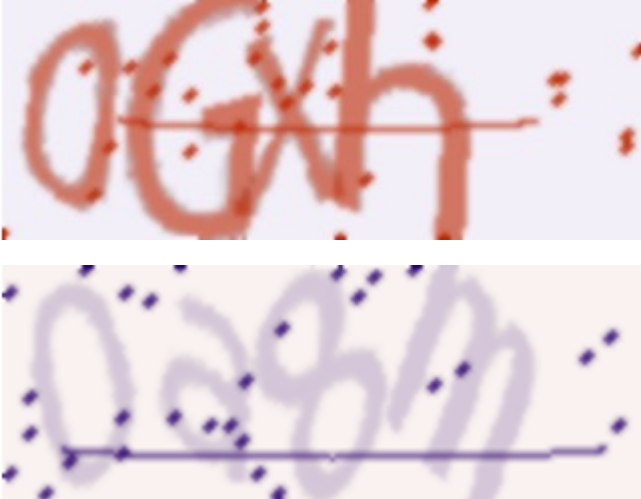
[evil77](#) / 2017-11-02 16:33:00 / 浏览数 6107 [技术文章](#) [技术文章 顶\(0\)](#) [踩\(0\)](#)

本项目源码及训练完成的模型均开源当前识别率98%。

转载请附：博文网址：[urlteam](#)，github 网址：[tensorflow_cnn](#)

新开一个专门存储TensorFlow项目的仓库逐步更新欢迎star：[tensorflow](#)

主流验证码偏向于用扭曲，倾斜，干扰例如下图：



因为字符距离近，没法采用先切割为单个字符然后进行局部识别的方式，so。

使用TensorFlow+cnn。进行卷积识别，该方法无需切割验证码，最终结果为训练4天(单台i5机器)达到98准确率

项目综述：

相关论文：

- [Multi-digit Number Recognition from Street View Imagery using Deep CNN](#)
- [CAPTCHA Recognition with Active Deep Learning](#)
- <http://matthewearl.github.io/2016/05/06/cnn-anpr/>

使用深度学习+训练数据+大量计算力，我们可以在几天内训练一个可以破解验证码的模型，不需要分割验证码，而是把验证码做为一个整体进行识别。

自己做一个验证码生成器，然后训练CNN模型破解自己做的验证码生成器。感觉的字符验证码机制可以废了，单纯的增加验证码难度只会让人更难识别，使用[CNN](#)+[RNN](#)，

CNN需要大量的样本进行训练。如果使用数字+大小写字母CNN网络有462个输出，只使用数字CNN网络有410个输出。因此需要一个脚本自动生成训练集。

最初cnn学习自: <http://blog.topspeedsnail.com/archives/10858>

成功率（可能波动，均亲身实践）：

1. 达到50%成功率需要2000个批次，总计20w张图片。
2. 达到70%成功率需要4000个批次，总计40w张图片。
3. 达到94%成功率需要40000个批次，总计400w张图片。
4. 达到98%成功率需要100000个批次，总计1000w张图片。

loss曲线为：



100

test_2 python tensorflow_cnn.py

运行命令

(\xe9\xad\x8c\xe8\xaf\x81\xe7\xa0\x81\xe5\x9b\xbe\xe5\x83\x8fchannel:', (60, 160, 3))
(\xe9\xad\x8c\xe8\xaf\x81\xe7\xa0\x81\xe5\x9b\xbe\xe5\x83\x8fchannel:', (60, 160, 3))
W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library w
asn't compiled to use SSE4.1 instructions, but these are available on your m
achine and could speed up CPU computations.
W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library w
asn't compiled to use SSE4.2 instructions, but these are available on your m
achine and could speed up CPU computations.
W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library w
asn't compiled to use AVX instructions, but these are available on your mach
ine and could speed up CPU computations.
W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library w
asn't compiled to use AVX2 instructions, but these are available on your mac
hine and could speed up CPU computations.
W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library w
asn't compiled to use FMA instructions, but these are available on your mach
ine and could speed up CPU computations.
(0, 0.69947368)
(0, 0.022500001)
(1, 0.59916627)
(2, 0.39934894)
(3, 0.69947368)

100的整倍数为当前准确率

当前loss值

SzFW.jpg T6f3.jpg Tb85.jpg Tkhi.jpg u0IL.jpg

实时图片生成

UCch.jpg ucTd.jpg vJ2g.jpg W11k.jpg w75y.jpg

W15X.jpg wu9D.jpg YMkf.jpg yQtA.jpg YT7P.jpg

YZmK.jpg Z4fq.jpg Z9UB.jpg ZO79.jpg zxTR.jpg

测试模型：

test_2 vim tensorflow_cnn.py

test_2 python tensorflow_cnn.py

(\xe9\xad\x8c\xe8\xaf\x81\xe7\xa0\x81\xe5\x9b\xbe\xe5\x83\x8fchannel:', (60, 160, 3))
(\xe9\xad\x8c\xe8\xaf\x81\xe7\xa0\x81\xe5\x9b\xbe\xe5\x83\x8fchannel:', (60, 160, 3))
W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to
use SSE4.1 instructions, but these are available on your machine and could speed up CPU comp
utations.
W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to
use SSE4.2 instructions, but these are available on your machine and could speed up CPU comp
utations.
W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to
use AVX instructions, but these are available on your machine and could speed up CPU comput
ations.
W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to
use AVX2 instructions, but these are available on your machine and could speed up CPU comput
ations.
W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to
use FMA instructions, but these are available on your machine and could speed up CPU comput
ations.
正确：wUtG 预测：wUtG

checkpoint
crack_capcha.mo
del-100...f-00001
crack_capcha.mo
del-100700.index
crack_capcha.mo
del-100700.meta
gen_capcha.py
captcha.pyc
rm_jpg.sh
tensorflow_cnn.py
fMHB.jpg
wUtG.jpg

本项目由ur1team维护，欢迎star

相关的验证码破解系列可以在这里找到：[github](#)

逐步更新TensorFlow系列项目：[github](#)

点击收藏 | 0 关注 | 1

[上一篇：【菜鸟腾飞】Mysql数据库渗透及...](#) [下一篇：图图图](#)

1. 7 条回复

shades 2017-11-02 16:42:17

看来昨天的那个轮子可以继续优化了

0 回复Ta



[cover](#) 2017-11-02 16:43:42

好东西

0 回复Ta



[紫霞仙子](#) 2017-11-02 17:19:01

我记得芒果还是数据流之前也做了套，应该多加参加下。

0 回复Ta



[stardustsky](#) 2017-11-03 10:46:46

建议你先做下预处理，你这样的爆破方式不通用不说，如果不是你自己生成的验证码而是从网站脱下来的，你想想给1kw张图打标签的工程量，这样纯暴力的训练意义不

0 回复Ta



[hades](#) 2017-11-06 09:19:26

[@stardustsky](#) 如果有好的方式可以聊聊

0 回复Ta



[stardustsky](#) 2018-01-26 14:10:10

[@hades](#) 对于预处理之前我在知乎上写过一篇，你可以看看 <https://www.zhihu.com/question/32209043/answer/228776686> 打标签的话这个没办法

0 回复Ta



[茶凉人尽散](#) 2018-12-09 17:13:40

模型的链接失效了，博主可以再发一次吗

0 回复Ta

[登录](#) 后跟贴

先知社区

[现在登录](#)

热门节点

[技术文章](#)

[社区小黑板](#)

目录

[RSS](#) [关于社区](#) [友情链接](#) [社区小黑板](#)