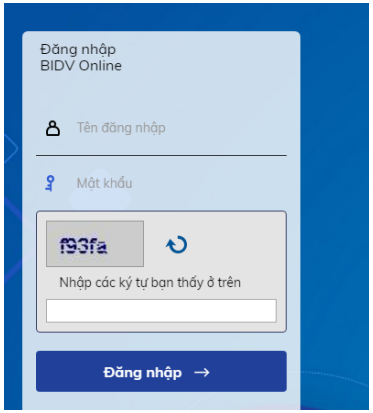
How a simple Tensorflow based program defeat CAPTCHA from almost Vietnamese banks

Nguyễn Hồng Điệp
Product Manager, ketoan.app

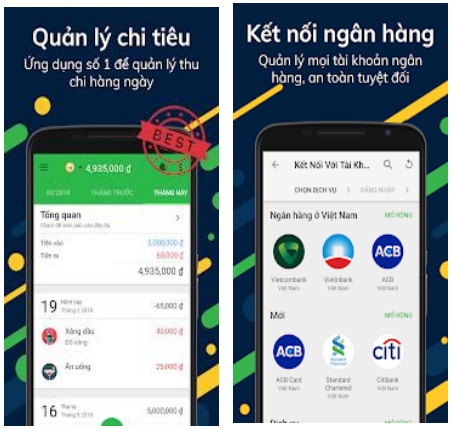
Captcha là gì?



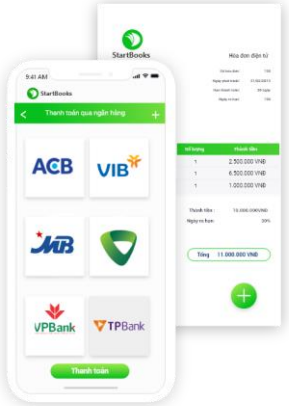
VIETCOMBANK **BIDV** **AGRIBANK**

Vượt captcha để làm gì?

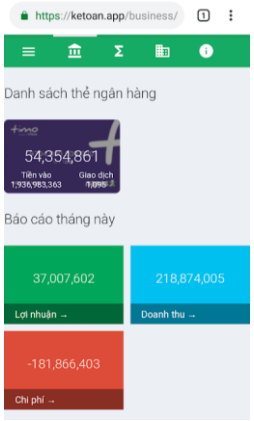
MONEY LOVER



MISA STARTBOOK



KETOAN.APP



Accounting Automation Software

Credit Rating

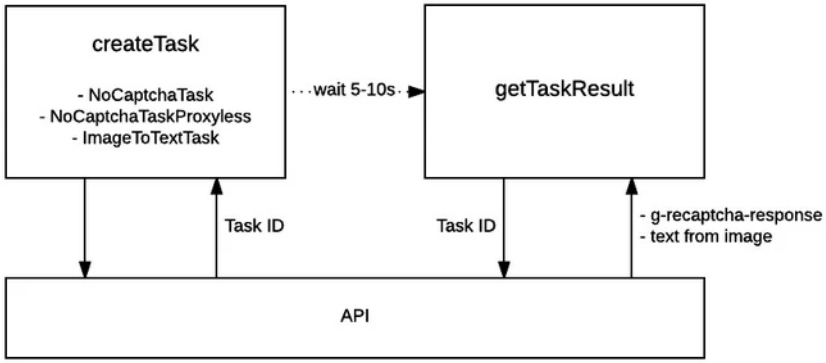
Bank Aggregator

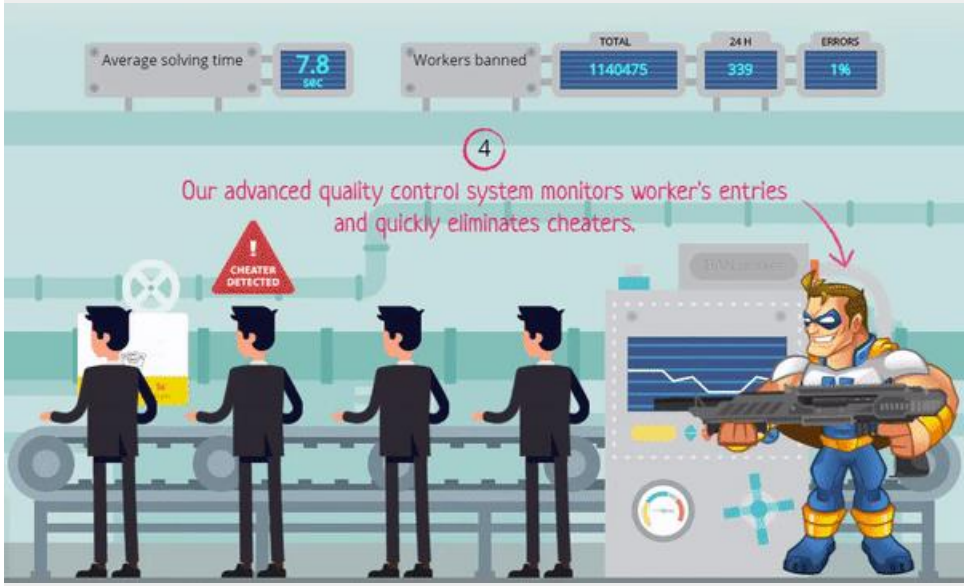
Bots

Captcha Bypass

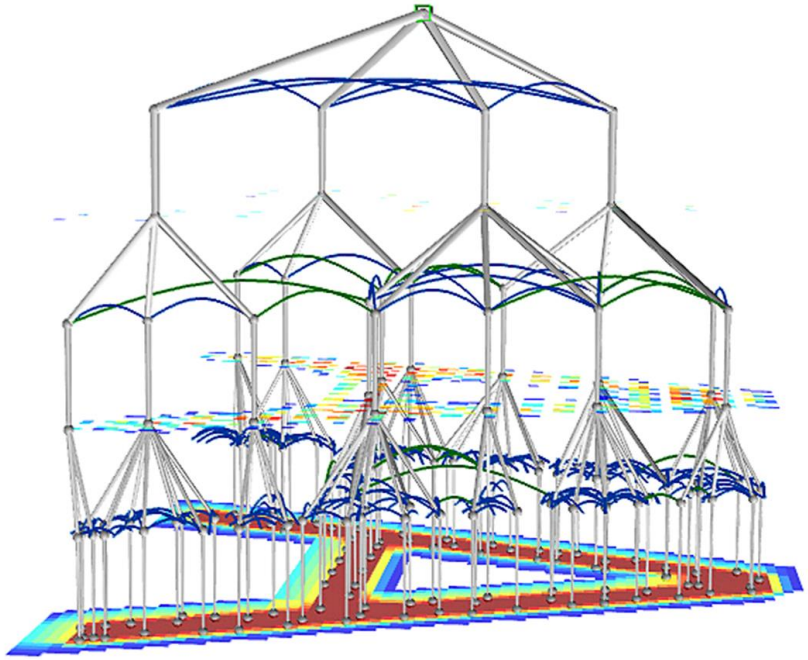
Có một nghề gõ captcha

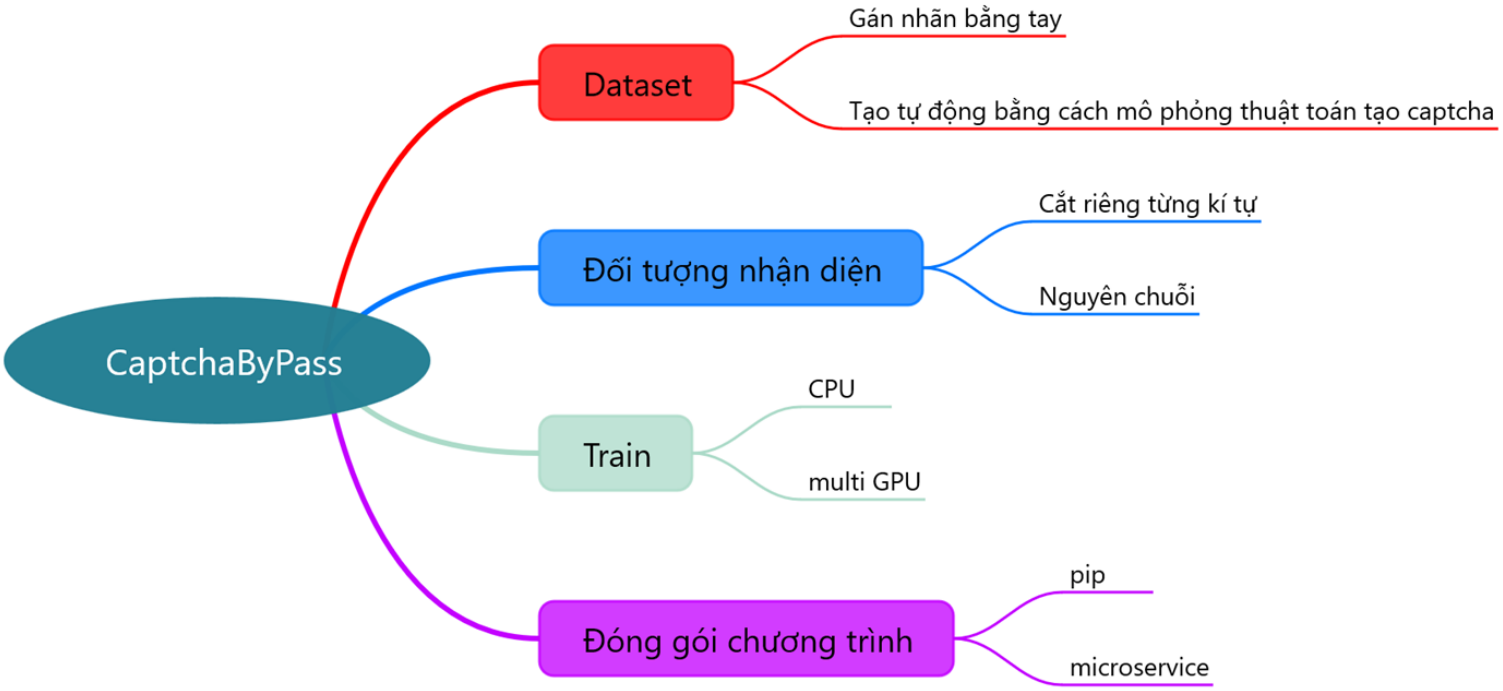
Anticaptcha API





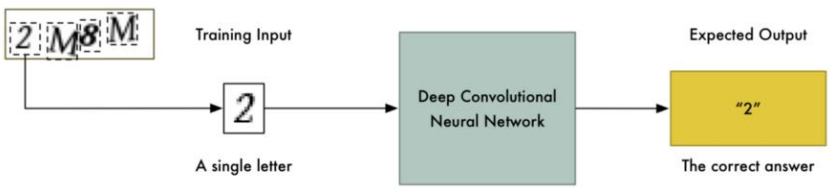
Vượt Captcha với AI



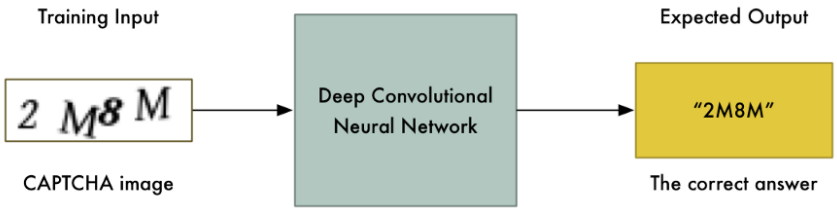

CaptchaByPass
Dataset
Gán nhãn bằng tay
Tạo tự động bằng cách mô phỏng thuật toán tạo captcha
Đối tượng nhận diện
Cắt riêng từng kí tự
Nguyên chuỗi
Train
CPU
multi GPU
Đóng gói chương trình
pip
microservice

Segmentation

Theo từng kí tự

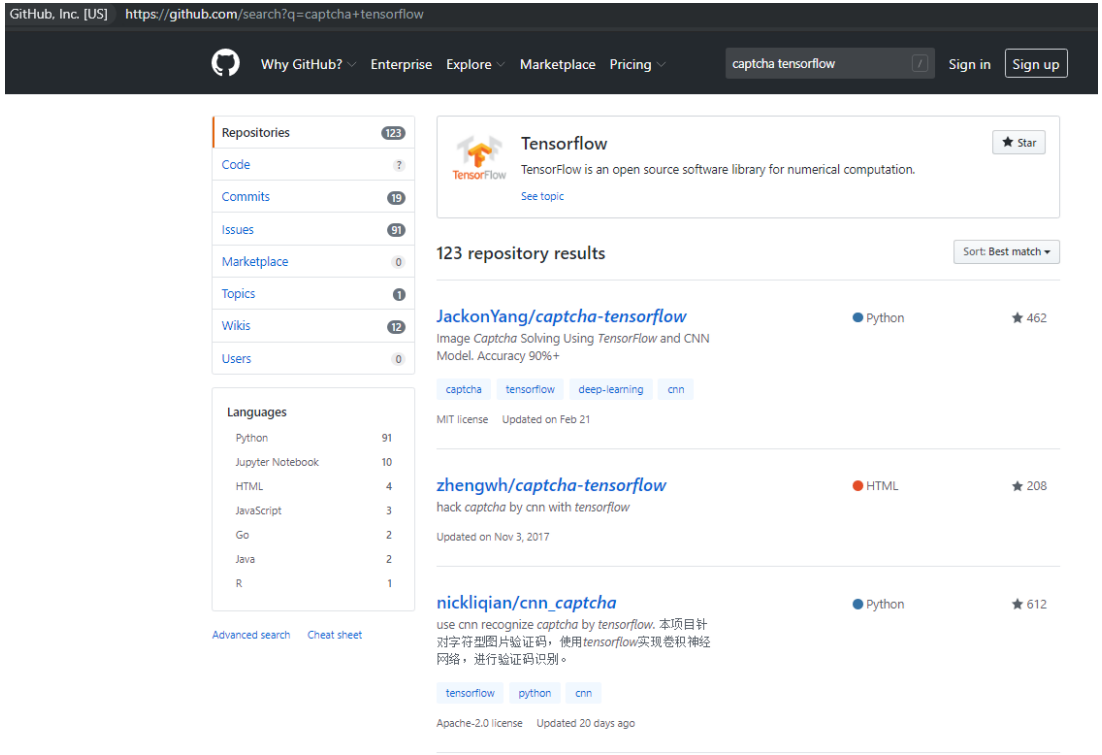


Theo nguyên cụm



Don't reinvent the wheel

Ngó qua github, khá nhiều opensource đã có sẵn.



Dễ dàng tìm được

3 dự án captcha bypass (Tensorflow)

chất lượng tốt

trên  .

vnbankpass

Vietcombank	
VietinBank	No captcha
AGRIBANK	
BIDV	
ACB	
VPBank	
TECHCOMBANK	No captcha

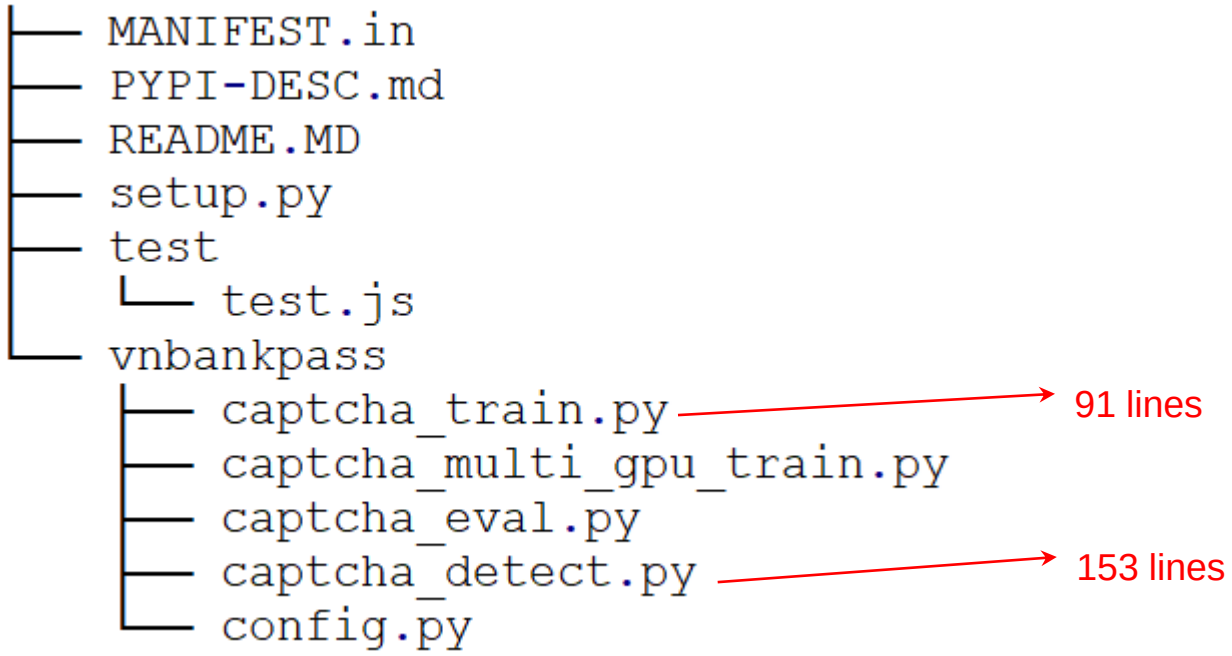
```
BANKS = {
    "vietcombank" : {
        "IMAGE_HEIGHT" : 25,
        "IMAGE_WIDTH" : 87,
        "CHAR_SETS" : '0123456789ABCDEF',
        "CLASSES_NUM" : 16,
        "CHARS_NUM" : 5,
    },
    "agribank" : {
        "IMAGE_HEIGHT" : 35,
        "IMAGE_WIDTH" : 75,
        "CHAR_SETS" : '0123456789abcdef',
        "CLASSES_NUM" : 16,
        "CHARS_NUM" : 5,
    },
    "bidv" : {
        "IMAGE_HEIGHT" : 35,
        "IMAGE_WIDTH" : 75,
        "CHAR_SETS" : '0123456789abcdef',
        "CLASSES_NUM" : 16,
        "CHARS_NUM" : 5,
    },
    "acb" : {
        "IMAGE_HEIGHT" : 27,
        "IMAGE_WIDTH" : 100,
        "CHAR_SETS" : '0123456789ABCDEFGHIJKLMNOPQRSTUVWXYZ',
        "CLASSES_NUM" : 36,
        "CHARS_NUM" : 5,
    },
    "vpbank" : {
        "IMAGE_HEIGHT" : 30,
        "IMAGE_WIDTH" : 150,
        "CHAR_SETS" : '0123456789',
```

vnbankpass

Để train cho một loại captcha mới

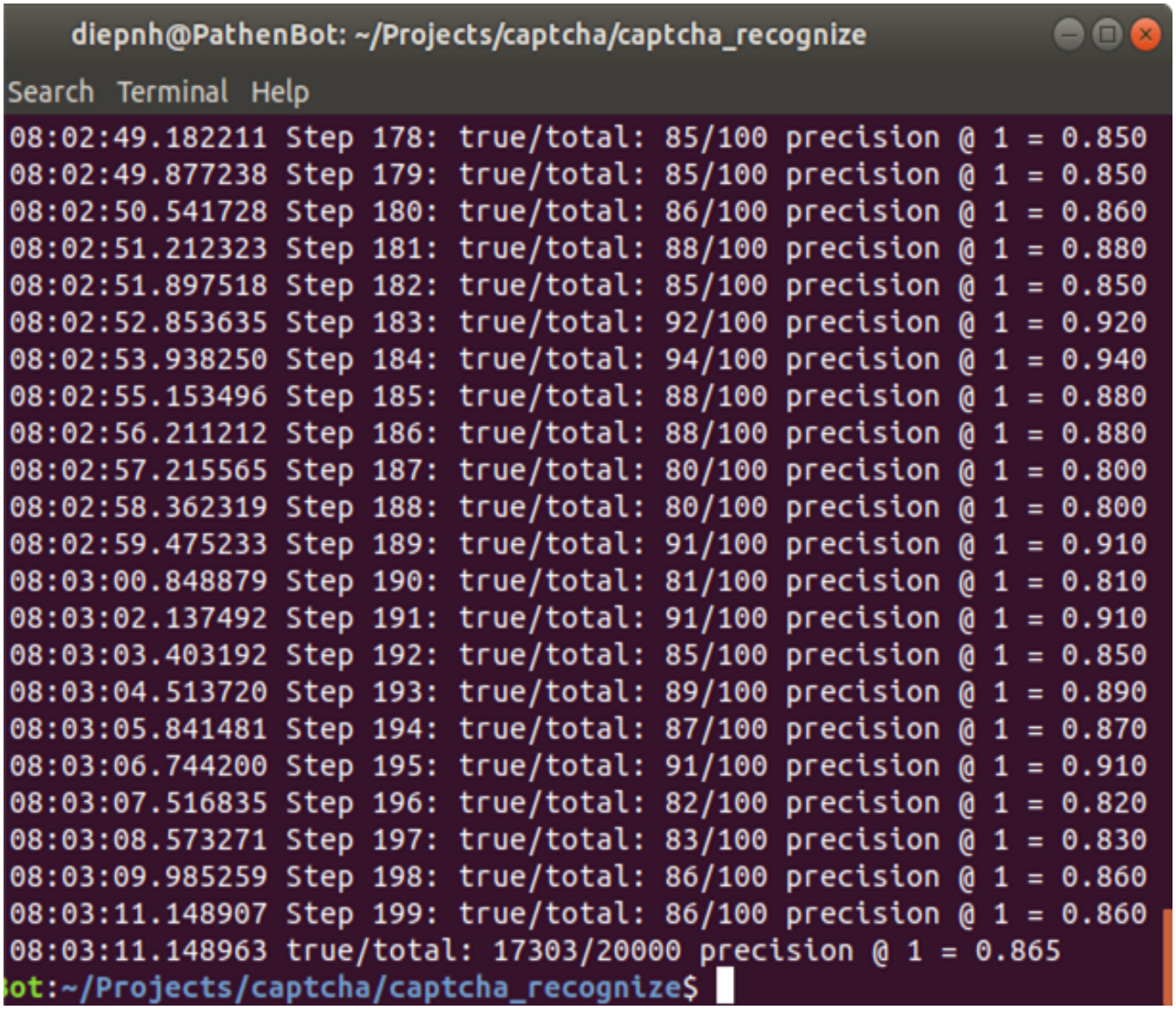
5,000	2 giờ	90%
dataset *(10USD)*	training time *(Single CPU Desktop)*	độ chính xác

Code structure



CNN Model

```
images = tf.reshape(images, [-1, IMAGE_HEIGHT, IMAGE_WIDTH, 1])

with tf.variable_scope('conv1') as scope:
  kernel = _weight_variable('weights', shape=[3,3,1,64])
  biases = _bias_variable('biases',[64])
  pre_activation = tf.nn.bias_add(_conv2d(images, kernel),biases)
  conv1 = tf.nn.relu(pre_activation, name=scope.name)

pool1 = _max_pool_2x2(conv1, name='pool1')

with tf.variable_scope('conv2') as scope:
  kernel = _weight_variable('weights', shape=[3,3,64,64])
  biases = _bias_variable('biases',[64])
  pre_activation = tf.nn.bias_add(_conv2d(pool1, kernel),biases)
  conv2 = tf.nn.relu(pre_activation, name=scope.name)

pool2 = _max_pool_2x2(conv2, name='pool2')

with tf.variable_scope('conv3') as scope:
  kernel = _weight_variable('weights', shape=[3,3,64,64])
  biases = _bias_variable('biases',[64])
  pre_activation = tf.nn.bias_add(_conv2d(pool2, kernel),biases)
  conv3 = tf.nn.relu(pre_activation, name=scope.name)

pool3 = _max_pool_2x2(conv3, name='pool3')

with tf.variable_scope('conv4') as scope:
  kernel = _weight_variable('weights', shape=[3,3,64,64])
  biases = _bias_variable('biases',[64])
  pre_activation = tf.nn.bias_add(_conv2d(pool3, kernel),biases)
  conv4 = tf.nn.relu(pre_activation, name=scope.name)

pool4 = _max_pool_2x2(conv4, name='pool4')

with tf.variable_scope('local1') as scope:
  batch_size = images.get_shape()[0].value
  reshape = tf.reshape(pool4, [batch_size,-1])
  dim = reshape.get_shape()[1].value
  weights = _weight_variable('weights', shape=[dim,1024])
```


diepnh@PathenBot: ~/Projects/captcha/captcha_recognize
Search Terminal Help
08:02:49.182211 Step 178: true/total: 85/100 precision @ 1 = 0.850
08:02:49.877238 Step 179: true/total: 85/100 precision @ 1 = 0.850
08:02:50.541728 Step 180: true/total: 86/100 precision @ 1 = 0.860
08:02:51.212323 Step 181: true/total: 88/100 precision @ 1 = 0.880
08:02:51.897518 Step 182: true/total: 85/100 precision @ 1 = 0.850
08:02:52.853635 Step 183: true/total: 92/100 precision @ 1 = 0.920
08:02:53.938250 Step 184: true/total: 94/100 precision @ 1 = 0.940
08:02:55.153496 Step 185: true/total: 88/100 precision @ 1 = 0.880
08:02:56.211212 Step 186: true/total: 88/100 precision @ 1 = 0.880
08:02:57.215565 Step 187: true/total: 80/100 precision @ 1 = 0.800
08:02:58.362319 Step 188: true/total: 80/100 precision @ 1 = 0.800
08:02:59.475233 Step 189: true/total: 91/100 precision @ 1 = 0.910
08:03:00.848879 Step 190: true/total: 81/100 precision @ 1 = 0.810
08:03:02.137492 Step 191: true/total: 91/100 precision @ 1 = 0.910
08:03:03.403192 Step 192: true/total: 85/100 precision @ 1 = 0.850
08:03:04.513720 Step 193: true/total: 89/100 precision @ 1 = 0.890
08:03:05.841481 Step 194: true/total: 87/100 precision @ 1 = 0.870
08:03:06.744200 Step 195: true/total: 91/100 precision @ 1 = 0.910
08:03:07.516835 Step 196: true/total: 82/100 precision @ 1 = 0.820
08:03:08.573271 Step 197: true/total: 83/100 precision @ 1 = 0.830
08:03:09.985259 Step 198: true/total: 86/100 precision @ 1 = 0.860
08:03:11.148907 Step 199: true/total: 86/100 precision @ 1 = 0.860
08:03:11.148963 true/total: 17303/20000 precision @ 1 = 0.865
ot:~/Projects/captcha/captcha_recognize$

Áp dụng AI vào product của bạn.