

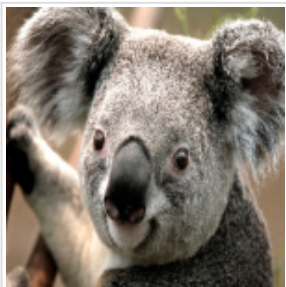
lujiandong1的专栏

目录视图

摘要视图

RSS 订阅

个人资料



BYR_jiandong



访问：132443次

积分：3525

等级：BLOG > 5

排名：第8044名

原创：220篇 转载：41篇

译文：0篇 评论：22条

【有奖投票】玩转Dragonboard 410c 的正确姿势 CSDN日报20170406 ——《代码很烂，所以离职。》 Python数据分析与机器学习 博客搬家，有礼相送

caffe 教程 Fine-tuning CaffeNet for Style Recognition on "Flickr Style" Data 遇到的问题及解决方案

标签：caffe

2016-01-11 10:55

1938人阅读

评论(1)

4X 收藏

分享

分类：caffe教程及遇到问题的解决方案 (1)

关闭

版权声明：本文为博主原创文章，未经博主允许不得转载。

在使用这个教程时，主要遇到了两个问题：

1、数据下不下来。

[python]



文章搜索

文章分类

[ACM DP问题](#) (7)[ACM其他文章](#) (2)[STL](#) (5)[人生感悟](#) (1)[C++基础](#) (20)[C#](#) (1)[ACM 贪心算法](#) (2)[ACM 哈希](#) (3)[二分法](#) (3)[STL 空间配置器](#) (1)[poj Trie树](#) (1)[poj 调试经验](#) (2)[POJ KMP](#) (2)[poj 暴力法](#) (1)[C++调试](#) (3)[More Effective C++读书笔记](#) (3)[Effective C++读书笔记](#) (25)[机器学习](#) (33)[北邮人论坛](#) (2)[杂项](#) (6)[数学](#) (2)[opencv](#) (3)[操作系统基本概念](#) (10)[Linux基础学习](#) (10)[程序优化](#) (3)[深入理解计算机系统](#) (5)

```
01. python examples/finetune_flickr_style/assemble_data.py --workers=1 --images=2000 --seed 831486
```

运行上述指令时，程序莫名其妙就不动了，也不下载文件，程序也没有挂掉，好像进入了死锁状态。

查看源程序：assemble_data.py，可以看出assemble_data.py用了大量多线程，多进程。我的解决方案就是改源程序，不使用进程来下载了。并且，对下载进行了超时限定，超过6s就认为超时，进而不下载。

assemble_data.py中使用多线程，多进程的源代码如下：

```
[python]

01. pool = multiprocessing.Pool(processes=num_workers)
02. map_args = zip(df['image_url'], df['image_filename'])
03. results = pool.map(download_image, map_args)
```

我修改后的源码如下：

```
[python]

01. #!/usr/bin/env python3
02. """
03. Form a subset of the Flickr Style data, download images to
04. Caffe ImagesDataLayer training file.
05. """
06. import os
07. import urllib
08. import hashlib
09. import argparse
10. import numpy as np
11. import pandas as pd
12. from skimage import io
13. import multiprocessing
```

关闭



[Python 基础 \(25\)](#)[SVM \(1\)](#)[推荐系统 \(1\)](#)[机器学习讲座笔记 \(1\)](#)[Kaggle学习笔记 \(10\)](#)[特征工程 \(2\)](#)[caffe教程及遇到问题的解决方案 \(2\)](#)[leetcode \(7\)](#)[自然语言处理 \(10\)](#)[哈工大SCIR 神经网络和深度学习转载 \(11\)](#)[tensorflow调研 \(37\)](#)

文章存档

[2017年03月 \(5\)](#)[2017年02月 \(9\)](#)[2017年01月 \(2\)](#)[2016年12月 \(8\)](#)[2016年11月 \(31\)](#)[展开](#)

阅读排行

[安装scikit-learn , win7 64 \(8805\)](#)[SVM的两个参数 C 和 ga \(6803\)](#)[Python 列表的清空 \(3361\)](#)[import sys sys.path.append \(3062\)](#)[error LNK2019: 无法解析 \(3018\)](#)[连续特征离散化达到更好 \(2567\)](#)[配置caffe的python接口及 \(2194\)](#)

```
14. import socket
15.
16.
17. # Flickr returns a special image if the request is unavailable.
18. MISSING_IMAGE_SHA1 = '6a92790b1c2a301c6e7ddef645dca1f53ea97ac2'
19.
20. example_dirname = os.path.abspath(os.path.dirname(__file__))
21. caffe_dirname = os.path.abspath(os.path.join(example_dirname, '../..'))
22. training_dirname = os.path.join(caffe_dirname, 'data/flickr_style')
23.
24.
25. def download_image(args_tuple):
26.     "For use with multiprocessing map. Returns filename on fail."
27.     try:
28.         url, filename = args_tuple
29.         if not os.path.exists(filename):
30.             urllib.urlretrieve(url, filename)
31.             with open(filename) as f:
32.                 assert hashlib.sha1(f.read()).hexdigest() != MISSING_IMAGE_SHA1
33.             test_read_image = io.imread(filename)
34.             return True
35.     except KeyboardInterrupt:
36.         raise Exception() # multiprocessing doesn't catch keyboard exceptions
37.     except:
38.         return False
39.
40. def mydownload_image(args_tuple):
41.     "For use with multiprocessing map. Returns filename on fail."
42.     try:
43.         url, filename = args_tuple
44.         if not os.path.exists(filename):
45.             urllib.urlretrieve(url, filename)
46.             with open(filename) as f:
47.                 assert hashlib.sha1(f.read()).hexdigest() != MISSING_IMAGE_SHA1
48.             test_read_image = io.imread(filename)
49.             return True
50.     except KeyboardInterrupt:
51.         raise Exception() # multiprocessing doesn't catch keyboard exceptions
52.     except:
```

[关闭](#)

caffe 教程 Fine-tuning CaffeNet for Style Recognition on "Flickr Style" Data (1938)

机器学习中的内核方法 (1873)

DBN的训练过程 (1790)

评论排行

连续特征离散化达到更好 (4)

machine learning week6 (3)

error LNK2019: 无法解析 (2)

深入分析C++引用 (1)

人为什么会浮躁 (1)

Python 列表的清空 (1)

XGBoost Stopping to Avoid Overfitting (1)

Andrew Ng的 Machine Learning (1)

配置caffe的python接口及caffe2 (1)

caffe 教程 Fine-tuning CaffeNet for Style Recognition on "Flickr Style" Data (1)

推荐文章

* Android安全防护之旅---带你把Apk混淆成中文语言代码

* TensorFlow文本摘要生成 - 基于注意力的序列到序列模型

* 创建后台任务的两种代码模式

* 一个屌丝程序员的人生 (六十)

* WKWebView与js交互之完美解决方案

* 年轻人,“砖砖瓦瓦”不应该成为你的梦想!

```

53.         return False
54.
55.
56.
57. if __name__ == '__main__':
58.     parser = argparse.ArgumentParser(
59.         description='Download a subset of Flickr Style to a directory')
60.     parser.add_argument(
61.         '-s', '--seed', type=int, default=0,
62.         help="random seed")
63.     parser.add_argument(
64.         '-i', '--images', type=int, default=-1,
65.         help="number of images to use (-1 for all [default])",
66.     )
67.     parser.add_argument(
68.         '-w', '--workers', type=int, default=-1,
69.         help="num workers used to download images. -
x uses (all - x) cores [-1 default].")
70.     )
71.     parser.add_argument(
72.         '-l', '--labels', type=int, default=0,
73.         help="if set to a positive value, only sample images from the first numbe
74.     )
75.
76.     args = parser.parse_args()
77.     np.random.seed(args.seed)
78.     # Read data, shuffle order, and subsample.
79.     csv_filename = os.path.join(example_dirname, 'flickr_style.csv')
80.     df = pd.read_csv(csv_filename, index_col=0, compression='gzip')
81.     df = df.iloc[np.random.permutation(df.shape[0])]
82.     if args.labels > 0:
83.         df = df.loc[df['label'] < args.labels]
84.     if args.images > 0 and args.images < df.shape[0]:
85.         df = df.iloc[:args.images]
86.
87.     # Make directory for images and get local filenames.
88.     if training_dirname is None:
89.         training_dirname = os.path.join(caffe_dirname, 'data')
90.     images_dirname = os.path.join(training_dirname, 'images')

```

关闭



最新评论

Andrew Ng的 Machine Learning
fupf1303: 写的挺好的, 可惜只有2和4, 有其他的课程笔记吗?

tensorflow MNIST数据集上简单
倾城一少: 博主, MLP网络的全称是什么?

tensorflow CNN for mnist
xjbada: 我运行这个代码为什么会出现这个错误呢*** TypeError: __init__() got an...

tensorflow中关于队列使用的实验
yuehanliushuang: very good

tesnsorflow 使用LSTM进行分类
qq_27590277: 为什么说我出错
TypeError: __init__() got an unexpected ...

error LNK2019: 无法解析的外部:
m0_37640107: 多谢博主! 我的opencv用vs生成的时候也一直报这个错误, 上网其他方案都没用, 直到看了你的才发现...

machine learning week6 诊断机
Starry5cm:
rand_seq=round(rand(1,i)*(m-1))+1;%生成i个随机序列 0~m这里改...

machine learning week6 诊断机
Starry5cm:
rand_seq=round(rand(1,i)*(m-1))+1;%生成i个随机序列 0~m这里改...

人为什么会浮躁
annipiao: 相当有见地的分析, 受教了

tensorflow中dropout的用法,防止
Wxilong: 博主你好, 你在文中说“train的时候才是dropout起作用的时候,train和test的时候不应...

```

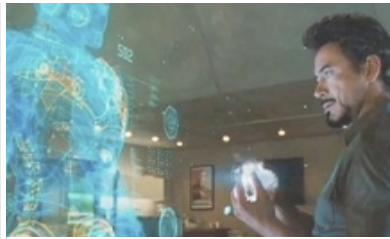
91.         if not os.path.exists(images_dirname):
92.             os.makedirs(images_dirname)
93.         df['image_filename'] = [
94.             os.path.join(images_dirname, _.split('/')[1]) for _ in df['image_url']]
95.         ]
96.
97.         # Download images.
98.         num_workers = args.workers
99.         if num_workers <= 0:
100.             num_workers = multiprocessing.cpu_count() + num_workers
101.         print('Downloading {} images with {} workers...'.format(
102.             df.shape[0], num_workers))
103.         #pool = multiprocessing.Pool(processes=num_workers)
104.         map_args = zip(df['image_url'], df['image_filename'])
105.         #results = pool.map(download_image, map_args)
106.         socket.setdefaulttimeout(6)
107.         results = []
108.         for item in map_args:
109.             value = mydownload_image(item)
110.             results.append(value)
111.             if value == False:
112.                 print 'Flase'
113.             else:
114.                 print '1'
115.         # Only keep rows with valid images, and write out training set
116.         print len(results)
117.         df = df[results]
118.         for split in ['train', 'test']:
119.             split_df = df[df['_split'] == split]
120.             filename = os.path.join(training_dirname, '{}.txt'.format(split))
121.             split_df[['image_filename', 'label']].to_csv(
122.                 filename, sep=' ', header=None, index=None)
123.         print('Writing train/val for {} successfully downloaded'.format(
124.             df.shape[0]))

```

修改主要有以下几点：

关闭





全息投影沙盘



1、#!/usr/bin/env python3 使用python3

2、

```
[python]
01. #pool = multiprocessing.Pool(processes=num_workers)
02.     map_args = zip(df['image_url'], df['image_filename'])
03.     #results = pool.map(download_image, map_args)
04.     socket.setdefaulttimeout(6)
05.     results = []
06.     for item in map_args:
07.         value = mydownload_image(item)
08.         results.append(value)
09.         if value == False:
10.             print 'Flase'
11.         else:
12.             print '1'
13.     # Only keep rows with valid images, and write out training file lists.
14.     print len(results)
```

只使用单线程下载，不使用多线程，多进程下载。并且，设定连接的超时时间为6s,socket.setdefaulttime

经过上述改进，就可以把数据下载下来。

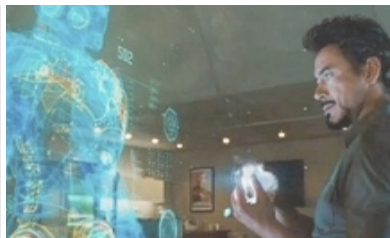
2、

在运行命令：

```
[plain]
01. ./build/tools/caffe train -solver models/finetune_flickr_s
weights models/bvlc_reference_caffenet/bvlc_reference_caff
```

关闭





全息投影沙盘



时遇到错误：

Failed to parse NetParameter file: models/bvlc_reference_caffenet/bvlc_reference_caffenet.caffemodel

出错的原因是我们传入的数据bvlc_reference_caffenet.caffemodel 并不是二进制的。

原因：因为我是在win7下，把bvlc_reference_caffenet.caffemodel下载下来，再使用winSCP传输到服务器上，直接在服务器上使用wget下载，速度太慢了，但是在传输的过程中winSCP就把bvlc_reference_caffenet.caffemodel的格式给篡改了,导致bvlc_reference_caffenet.caffemodel不是二进制的。

解决方案，把winSCP的传输格式设置成二进制，那么就可以解决这个问题。

详情见博客:<http://blog.chinaunix.NET/uid-20332519-id-5585964.html>

顶 踩
1 0

关闭

上一篇 vowpal_wabbit是在单机上性能极高的机器学习库,比较online le

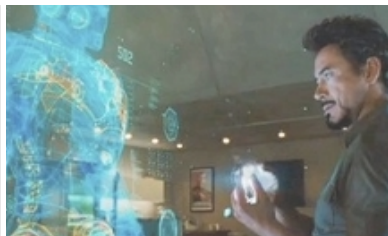
下一篇 配置caffe的python接口及其易错点

我的同类文章

caffe教程及遇到问题的解决方案（1）

- 配置caffe的python接口及其... 2016-01-19 阅读 2193





全息投影沙盘



新东方
XDF.CN
北京总部

要留学先游学
走访世界名校 提升留学背景

广告

咨询详情

参考知识库



Python知识库

22340 关注 | 1443 收录



.NET知识库

3647 关注 | 833 收录

猜你在找

全网服务器数据备份解决方案案例实践

windows server 2008 r2服务器操作系统

iis+php网站服务器环境配置视频教程

2017最新Linux集群全网服务器数据备份方案超细实战!

windows批处理教程

解决方案The method getTextContent is undefined for the

Caffe fine-tuning 微调网络

转Caffe fine-tuning 微调网络

Caffe fine-tuning 微调网络

Caffe fine-tuning 微调网络

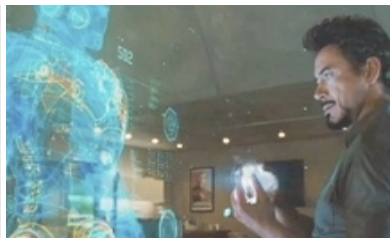
关闭

查看评论

1楼 [ainijingdan](#) 2016-09-28 11:46发表

博主您好，这个文件怎么使用下载指定网站上的图片和标签呢？恳请博主指导下

您还没有登录,请[登录](#)或[注册](#)



全息投影沙盘



* 以上用户言论只代表其个人观点，不代表CSDN网站的观点或立场

核心技术类目

全部主题 Hadoop AWS 移动游戏 Java Android iOS Swift 智能硬件 Docker OpenStack
VPN Spark ERP IE10 Eclipse CRM JavaScript 数据库 Ubuntu NFC WAP jQuery
BI HTML5 Spring Apache .NET API HTML SDK IIS Fedora XML LBS Unity
Splashtop UML components Windows Mobile Rails QEMU KDE Cassandra CloudStack FTC
coremail OPhone CouchBase 云计算 iOS6 Rackspace Web App SpringSide Maemo
Compuware 大数据 aptech Perl Tornado Ruby Hibernate ThinkPHP HBase Pure Solr
Angular Cloud Foundry Redis Scala Django Bootstrap

[公司简介](#) | [招贤纳士](#) | [广告服务](#) | [联系方式](#) | [版权声明](#) | [法律顾问](#) | [问题报告](#) | [合作伙伴](#) | [论坛反馈](#)

网站客服

杂志客服

微博客服

webmaster@csdn.net

400-600-2320

| 北京创新乐知信息技术有限公司 版权所有 | 江苏知之为计算机有限公司 | 江苏乐知网络技术有限公司

京 ICP 证 09002463 号 | Copyright © 1999-2016, CSDN.NET, All Rights Reserved



2

关闭

