CSDN新首页上线啦,邀请你来立即体验!(http://blog.csdn.net/)

立即体

ふ



博客 (//blog. **(#kulnywet/9def:ntxt/9lled+)**toolba**学**院 (//edu.csdn.net?ref=toolbar)

下载 (//download.csdn.net?ref=toolbar)

GitChat (//gitbook.cn/?ref=csdn)

更多心

1

Q





登录 (https://passport.csdr///\\#titte://passport.csdr//\#titte://passport.csdn.net/account/mobileregister?ref=toolbar&action=mobileRegister) ref=toollbar) source=csdnblog1)

TensorFlow学习(十一):保存TFRecord文件

原创

2017年06月01日 16:11:39

标签:tensorflow(http://so.csdn.net/so/search/s.do?q=tensorflow&t=blog)

3565

ಹ್ಳ

做过kaggle竞赛的应该很熟悉.csv文件了,.csv文件非常方便,但是通常读取的时候,是一次性读取到内存里面的. 要是内存小的话,就要想其他的办法了,那就变得很麻烦了.

或者有时候,从硬盘上面直接读取图片啊什么的,因为图片的文件格式,存放位置各种各样等等一些因素,要是 想在训练阶段直接这么使用的话.就更加麻烦了.所以.对于数据进行统一的管理是很有必要的.TFRecord就是 对于输入数据做统一管理的格式.加上一些多线程的处理方式,使得在训练期间对于数据管理把控的效率和舒 适度都好于暴力的方法.

小的任务什么方法差别不大,但是对于大的任务,使用统一格式管理的好处就非常显著了.因此,TFRecord的使 用方法很有必要熟悉.

这节并不准备将TFRcord文件的读取,只讲怎么保存为TFRecord文件,读取还涉及到其他的操作,所以之后会和 其他的操作一起讲.

加入CSDN,享受更精准的内容推荐,与5000万程序员共同成长!



谢小小XH (http://blog.csd...

+ 关注

(http://blog.csdn.net/xierhacker)

码云

未开诵 原创 粉丝 (https://gite 87 382 utm sourc

▮他的最新文章

更多文章 (http://blog.csdn.net/xierhacker)

机器学习笔记十九:正则化思想 (http:// blog.csdn.net/xierhacker/article/details/ 76689572)

机器学习笔记十八:概率分类思想 (htt p://blog.csdn.net/xierhacker/article/det ails/76161440)

深度学习笔记十:生成对抗网络再思考f -GAN (http://blog.csdn.net/xierhanker/a rticle/details/75759538)

一.几个例子

这里一开始就直接写两个非常常见的例子(有多常见你看一下就知道了),来体会一下TFRecord可能的用法,在这里可以并不用知道程序运行每一行的具体含义,但是要大概知道是怎么回事.更加详细的对这两个例子的讲解会在后面的内容.这里的目的就是先感受一下.

在这个例子中使用的.csv文件就是kaggle竞赛里面MNIST手写体识别的train.csv文件,可以在官网上面下载实验一下.

I.csv文件转化为TFRecord格式

···

ಹ್

(https://passport.csdn.net/a 深度学习笔记九:生成对抗网络GAN (基本理论) (http://blog.csdn.net/xierhac ker/article/details/75216652)

C++学习:对象和类详细总结 (http://blo g.csdn.net/xierhacker/article/details/74 939880)

■相关推荐

TFRecords 文件的生成和读取 (http://blo g.csdn.net/u012222949/article/details/72 875281)

TensorFlow学习记录-- 7.TensorFlow高效读取数据之tfrecord详细解读 (http://blog.csdn.net/qq_16949707/article/details/53483493)

Tensorflow之构建自己的图片数据集TFrec ords (http://blog.csdn.net/csuzhaoqinghu i/article/details/51377941)

TensorFlow高效读取数据的方法 (http://bl og.csdn.net/u012759136/article/details/5 2232266)

加入CSDN,享受更精准的内容推荐,与5000万程序员共同成长!

登录 注册



```
import tensorflow as tf
               import numpy as np
               import pandas as pd
           4
               #load .csv file
               train frame=pd.read csv(filepath or buffer="train.csv")
               #print(train_frame.head())
               train labels frame=train frame.pop(item="label")
      1
               print(train_labels_frame.shape)
               train values=train frame.values
               train labels values=train labels frame.values
          12
               train_size=train_values.shape[0]
          14
     16
                  -----#
               writer=tf.python_io.TFRecordWriter(path="train.tfrecords")
          18
          19
               for i in range(train_size):
          20
                 image_raw=train_values[i].tostring()
          21
          22
                 example=tf.train.Example(
          23
                   features=tf.train.Features(
                     feature={
          24
          25
                        "image_raw":tf.train.Feature(bytes_list=tf.train.BytesList(value=[image_raw])),
                        "label":tf.train.Feature(int64 list=tf.train.Int64List(value=[train labels values[i]]))
          26
                     }
          27
          28
          29
                 writer.write(record=example.SerializeToString())
          30
加入CSDN , 享受更精准的内容推荐 , 与5000万程序员共同成长 ! szt. writer.close()
```





高倍显微镜

零基础学习手绘

四十平小户装修 OA办公系统 loft公寓 日语常用语 生日蛋糕网上.. 普吉岛游多少钱 70平米室内装修 小美式装修 130平米家装 二手玛莎...







手持式三维扫描仪

别墅装潢





it培训机构排名

四十平小户装修

他的热门文章

TensorFlow学习(四):优化器Optimizer (http://blog.csdn.net/xierhacker/article/det ails/53174558)

35349

TensorFlow学习(二):变量常量类型(h ttp://blog.csdn.net/xierhacker/article/detail s/53103979)登录

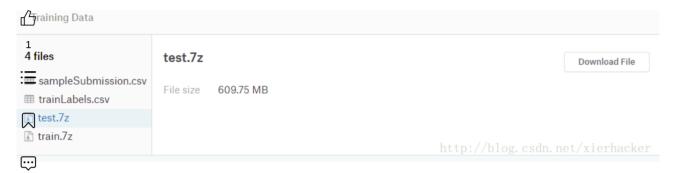


II.硬盘上图片转成TFRecord格式

这里使用的例子还是来自于kaggle,是在kaggle中的CIFAR-10识别比赛中的数据集.

CIFAR-10 - Object Recognition in Images (https://www.kaggle.com/c/cifar-10/data)

数据集主要是两个压缩包、train数据集和test数据集、解压之后就是两个文件夹、里面放了一大推图片.



要是不用TFRecord的话,直接处理是比较低效的.那么来看看用TFRecord怎么处理.

ಹ

24997

(https://passport.csdn.net/a

Windows+VS2013爆详细Caffe编译安装 教程 (http://blog.csdn.net/xierhacker/article/details/51834563)

24646

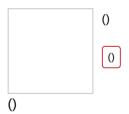
TensorFlow学习(三): Graph和Session (http://blog.csdn.net/xierhacker/article/details/53860379)

17279

爆详细Ubuntu16.04,CUDA8.0,OpenCV3. 1,Tensorflow完全配置指南 (http://blog.cs dn.net/xierhacker/article/details/5303598 9)

12922

微博



加入CSDN,享受更精准的内容推荐,与5000万程序员共同成长!

登录

注册



```
1 import matplotlib.pyplot as plt
                import matplotlib.image as mpimg
                import numpy as np
            3
                import tensorflow as tf
                import pandas as pd
            6
                import os
      6
       1
                #get the amount of files in folder
                def sizeOfFolder(folder_path):
      :≡ 10
                  fileNameList = os.listdir(path=folder_path)
          12
                  size = 0
      \square 13
                  for fileName in fileNameList:
                     if (os.path.isfile(path=os.path.join(folder_path, fileName))):
           14
      ₩ 15
                       size += 1
                  return size
           16
      ॡ 17
          18
                #path(path of folder)
           19
           20
                #if isTrain=True,labels can't be None
                def pics_to_TFRecord(folder_path,labels=None,isTrain=False):
           21
                  size=sizeOfFolder(folder_path=folder_path)
           22
           23
                  #train set
           24
           25
                  if isTrain:
                     if labels is None:
           26
           27
                       print("labels can't be None!!!")
                       return None
           28
                     if labels.shape[0]!=size:
           29
                       print("something wrong with shape!!!")
加入CSDN , 享受更精准的内容推荐 , 与5000万程序员共同成长!
32 writer=tf.python_io.TFRecordWriter("../data/TFRecords/train.tfrecords")
```

登录 注册

```
33
                   for i in range(1,size+1):
          34
                     print("-----processing the ",i,"\'th image-----")
                     filename=folder_path+str(i)+".png"
          35
                     img=mpimg.imread(fname=filename)
          36
                     width=img.shape[0]
          37
          38
                     print(width)
          39
                      #trans to string
     40 کا
                     img_raw=img.tostring()
      1 41
          42
                     example=tf.train.Example(
      :≡ 43
                        features=tf.train.Features(
                          feature={
          44
          45
                            "img_raw":tf.train.Feature(bytes_list=tf.train.BytesList(value=[img_raw])),
          46
                            "label":tf.train.Feature(int64 list=tf.train.Int64List(value=[labels[i-1]])),
     ... 47
                            "width":tf.train.Feature(int64_list=tf.train.Int64List(value=[width]))
          48
          50
          51
          52
                     writer.write(record=example.SerializeToString())
          53
                   writer.close()
          54
          55
                 #test set
                 else:
          56
          57
                   writer = tf.python_io.TFRecordWriter("../data/TFRecords/test.tfrecords")
                   for i in range(1, size + 1):
          58
                     print("-----processing the ", i, "\'th image-----")
          59
          60
                     filename = folder_path + str(i) + ".png"
                     img = mpimg.imread(fname=filename)
          61
                     width = img.shape[0]
          62
                     print(width)
加入CSQAI,享受專精准的協論推荐,与5000万程序员共同成长!
                                                                                                                                                         登录
```

登录 注册

```
65
                       img_raw = img.tostring()
           66
                       example = tf.train.Example(
           67
           68
                         features=tf.train.Features(
           69
                            feature={
                              "img_raw": tf.train.Feature(bytes_list=tf.train.BytesList(value=[img_raw])),
           70
                              "width": tf.train.Feature(int64_list=tf.train.Int64List(value=[width]))
           71
      72 كأ
      <sub>1</sub> 73
           74
      := 75
           76
                       writer.write(record=example.SerializeToString())
      □ 77
                     writer.close()
          78
      ... 79
           80
           82
           83
                train_labels_frame=pd.read_csv("../data/trainLabels.csv")
                train_labels_frame_dummy=pd.get_dummies(data=train_labels_frame)
                #print(train_labels_frame_dummy)
                train_labels_frame_dummy.pop(item="id")
                #print(train_labels_frame_dummy)
           87
                train_labels_values_dummy=train_labels_frame_dummy.values
           88
                #print(train_labels_values_dummy)
                train_labels_values=np.argmax(train_labels_values_dummy,axis=1)
                #print(train_labels_values)
           92
           93
                #write train record
95 pics_to_TFRecord(folder_path="../data/train/",labels=train_labels_values,isTrain=True)
加入CSDN , 享受更精准的内容推荐,与5000万程序员共同成长!
```

登录 注册

97 #write test record

98 pics_to_TFRecord(folder_path="../data/test/")

(https://passport.csdn.net/a

上面的代码运行之后,就会在设置的文件夹下面得到两个.tfrecords文件. 其中训练集的有500多M,测试集达到了惊人的3G多.但是为了后面训练的方便.这是值得的.

血.存为多个TFRecord文件

.

通过前面两个方法,我们知道可以把你想要的文件或者记录通过或多或少的方法转为TFRecord格式.那么数据量很大的时候,你会发现,单个TFRecord文件是非常非常大的,这对于硬盘是不小的负担,所以,可以通证存储多个TFRecord文件来解决问题.

 $\overline{\odot}$

ಹೆ

加入CSDN,享受更精准的内容推荐,与5000万程序员共同成长!

登录

注册



```
import tensorflow as tf
      2
          num files=3
      3
          num_instance=100
          for i in range(num_files):
            print("write ",i," file")
            fileName=("test.tfrecords-%.5d-of-%.5d" % (i,num_files))
 1
            writer=tf.python_io.TFRecordWriter(path=fileName)
            for j in range(num_instance):
              print("write ",j," record")
    12
              example=tf.train.Example(
                 features=tf.train.Features(
    14
₩ 15
                      feature={
                        "i":tf.train.Feature(int64_list=tf.train.Int64List(value=[i])),
    16
& 17
                        "j":tf.train.Feature(int64_list=tf.train.Int64List(value=[j]))
    18
    19
    20
    21
    22
              writer.write(record=example.SerializeToString())
    23
            writer.close()
```

二.常用的类

I tf.python_io.TFRecordWriter

加入CSDN 享受更精准的内容推荐,与5000万程序员共同成长! 把记录写入到TFRecords文件的类。 (https://passport.csdn.net/a

登录 注册

函数

(https://passport.csdn.net/a

__init__ (path,options=None)

作用:创建一个TFRecordWriter对象,这个对象就负责写记录到指定的文件中去了.

心 参数:

1 path: TFRecords 文件路径

options: (可选) TFRecordOptions对象

dose()

₩

作用:关闭对象.

æ

write(record)

作用:把字符串形式的记录写到文件中去.

参数:

record: 字符串,待写入的记录

I.tf.train.Example

这个类是非常重要的,TFRecord文件中的数据都是通过tf.train.Example Protocol Buffer的格式存储的.

在这里,不会非常详细的讲这个类,但是会给出常见的使用方法和一些重要函数的解释.其他的细节可以参考文加入CSDN,享受更精准的内容推荐,与5000万程序员共同成长!

登录 注册

class tf.train.Example

(https://passport.csdn.net/a

属性:

features Magic attribute generated for "features" proto field.

1

遠数:

__init__ (**kwargs)

○ 这个函数是初始化函数,会生成一个Example对象,一般我们使用的时候,是传入一个tf.train.Features对象进去.

SerializeToString()

作用:把example序列化为一个字符串,因为在写入到TFRcorde的时候,write方法的参数是字符串的.

Ⅲ.tf.train.Features

class tf.train.Features

定性 加入**CSDN**,享受更精准的内容推荐,与5000万程序员共同成长!

登录 注册

×

feature

函数:

__init__ (**kwargs)

作用:初始化Features对象,一般我们是**传入一个字典**,字典的键是一个字符串,表示名字,字典的值是一个ff.train.Feature对象.

≔

W.tf.train.Feature

qlass tf.train.Feature

属性:

bytes_list

float_list

int64_list

函数:

__init__ (**kwargs)

作用:构造一个Feature对象,一般使用的时候,传入 tf.train.Int64List ,tf.train.BytesList ,tf.train.FloatList 对象

对象. 加入CSDN,享受更精准的内容推荐,与5000万程序员共同成长!

登录 注册

×

V.tf.train.Int64List, tf.train.BytesList, tf.train.FloatList

使用的时候,一般传入一个具体的值,比如学习任务中的标签就可以传进value=tf.train.Int64List,而图片就可以 先转为字符串的格式之后,传入value=tf.train.BytesList中.



说明:



以上的函数的API都可以对照着例子的代码来熟悉,看在例子中是怎么使用的这些对象.

 $\overline{\odot}$

版取声明:本文为博主原创文章,未经博主允许不得转载。

Д



Maan_liaa (/Maan_liaa) 前天 21:21

1楼

(/Maa有没有读取的博客啊

回复

加入CSDN,享受更精准的内容推荐,与5000万程序员共同成长了 相关又量推荐

登录

注册

×

TFRecords 文件的生成和读取 (http://blog.csdn.net/u012222949/article/details/72875281)

TensorFlow提供了TFRecords的格式来统一存储数据,理论上,TFRecords可以存储任何形式的数据。 TFRecords文件中的数据都是通过tf.train.Example ...



u012222949 (http://blog.csdn.net/u012222949) 2017年06月06日 10:09 2018

മ

TensorFlow学习记录-- 7.TensorFlow高效读取数据之tfrecord详细解读 (http://blog.csdn.ne...



一 why tfrecord?对于数据量较小而言,可能一般选择直接将数据加载进内存,然后再分batch输入网络进行训练(tip:使用这种方法时,结合yield 使用更为简洁,大家自己尝试一下吧,我就不...



🧑 qq 16949707 (http://blog.csdn.net/qq 16949707) 2016年12月06日 10:06 🕮 7210





惊呆了! 微博和阿里背后的数据库有多厉害?

想不到!数据库作为最关键的基础设施,渗透技术领域的方方面面,我阿里和微博的师哥们是这么分享的...

(http://www.baidu.com/cb.php?c=IgF_pyfqnHmknjTzrjb0IZ0qnfK9ujYzP1nsrjD10Aw-5Hc3rHnYnHb0TAq15HfLPWRznjb0T1d-PynznWT4uhNhrymvP1bk0AwY5HDdnHn3nHnLn1n0IgF_5y9YIZ0IQzq-uZR8mLPbUB48ugfEpZNGXy-jULNzTvRETvNzpyN1gvw-IA7GUatLPjqdIAdxTvqdThP-5yF_UvTkn0KzujYk0AFV5H00TZcqn0KdpyfqnHRLPjnvnfKEpyfqnHc4rj6kP0KWpyfqP1cvrHnz0AqLUWYs0ZK45HcsP6KWThnqnW0sP1b)

Tensorflow之构建自己的图片数据集TFrecords (http://blog.csdn.net/csuzhaoqinghui/articl...

学习谷歌的深度学习终于有点眉目了,给大家分享我的Tensorflow学习历程。 tensorflow的官方中文文档比较生涩,数据集一直采用的MNIST二进制数据集。并没有过多讲述怎么构建自己的图...

加美SBN,享受更精准的内容推荐,与5000万程序员共同成长!

登录

注册

TensorFlow高效读取数据的方法 (http://blog.csdn.net/u012759136/article/details/52232266)

概述关于Tensorflow读取数据,官网给出了三种方法: 供给数据(Feeding): 在TensorFlow程序运行的每一步, 让Python代 码来供给数据。 从文件读取数据: 在TensorFl...

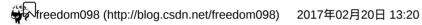


u012759136 (http://blog.csdn.net/u012759136) 2016年08月17日 19:20 €35598



Tensorflow建立与读取TFrecorder文件 (http://blog.csdn.net/freedom098/article/details/560...

Tei____orflow建立与读取TFrecorder文件除了直接读取数据文件,比如csv和bin文件,tensorflow还可以建立一种自有格式的数 据文件,称之为tfrecorder,这种文件储存类似于...



□14820



1.80/只 批发50*50防水盒 防水 接线盒 IP65电缆接线



6.00/只 户外防水接线盒分线盒 塑料端子盒丁控密封盒



1.80/只 开孔-电缆接头盒防水

TFRecord 的使用 (http://blog.csdn.net/xueyingxue001/article/details/68943650)

PS: 这段内容摘自 http://wiki.jikexueyuan.com/project/tensorflow-zh/how tos/readi... 什么是 TFRecord



xueyingxue001 (http://blog.csdn.net/xueyingxue001) 2017年04月01日 13:53

Tensorflow构建自己的图片数据集TFrecords (http://blog.csdn.net/tengxing007/article/deta...

:C/C++ 图像二进制存储与读取中,已经讲解了如何利用C/C++的方法存储与读取二进制图像文件,本文继续讲述如何根据C IFAR-10的格式制作自己的数据集。 所述博文与代码均已同步至GitHub ...

□5085

登录 注册

TensorFlow 学习(二)制作自己的TFRecord数据集,读取,显示及代码详解(http://blog.cs...

前言在跑通了官网的mnist和cifar10数据之后,笔者尝试着制作自己的数据集,并保存,读入,显示。 TensorFlow可以支持cif ar10的数据格式, 也提供了标准的TFRecord 格式, 而...



miaomiaoyuan (http://blog.csdn.net/miaomiaoyuan) 2017年02月24日 20:33 □5510

tensorflow中的TFRecord格式文件的写入和读取 (http://blog.csdn.net/UESTC_C2_403/artic...

在t氧sorflow中,TFRecord格式的文件是可以将样本和标签放在一起,是在模型之前的一个预处理步骤,这种方式可以很大 的提高效率和节约运行的内存,这种格式对数据是不进行压缩的。 1.写入 im...



MUESTC C2 403 (http://blog.csdn.net/UESTC C2 403) 2017年07月05日 15:11 🖫 1287



TensorFlow TFRecord数据集的生成与显示 (http://blog.csdn.net/chaipp0607/article/details...

ಹ್ಳ TensorFlow提供了TFRecord的格式来统一存储数据,TFRecord格式是一种将图像数据和标签放在一起的二进制文件,能更 好的利用内存,在tensorflow中快速的复制,移动,读取,存储...



tutorial cifar10 tfrecord.py解读 (http://blog.csdn.net/NeutronT/article/details/53541250)

把Github上TensorLayer的示例——tutorial cifar10 tfrecord.py做了一些中文注解。针对TF r0.12的情况,将源代码做了响应 的改动。 对涉及的函数做了介绍。...



NeutronT (http://blog.csdn.net/NeutronT) 2016年12月09日 16:45

制作自己的tfrecords数据集 (http://blog.csdn.net/GISLN/article/details/69831347)

加入CSDN,享受更精准的内容推荐,与5000万程序员共同成长!

登录 注册

最近想要制作自己的图像数据集,看了很多网上的教程,但是坑太多。。。。现在为了让大家少踩坑,附上全部代码,其中包 括tfrecords格式数据集的制作与读取。 import os os.environ[...



学习笔记TF016:CNN实现、数据集、TFRecord、加载图像、模型、训练、调试 (http://blog.cs...

AlexNet(Alex Krizhevsky,ILSVRC2012冠军)适合做图像分类。层自左向右、自上向下读取,关联层分为一组,高度、宽度减 小 ¹深度增加。深度增加减少网络计算量。训练模型数据集 S...

♣=weixin 38776853 (http://blog.csdn.net/weixin 38776853) 2017年05月30日 20:00 □1484

Tensorflow 中Tfrecords的使用心得 (http://blog.csdn.net/u014802590/article/details/68495...

运篇博客主要讲了如何用Tensorflow中的标准数据读取方式的简单的实现对自己数据的读取操作。...

__ou014802590 (http://blog.csdn.net/u014802590) 2017年03月30日 20:10 🕮 1480

[TFRecord格式数据]利用TFRecords存储与读取带标签的图片 (http://blog.csdn.net/u013555...

利用TFRecords存储与读取带标签的图片觉得有用的话,欢迎一起讨论相互学习~Follow MeTFRecords其实是一种二进制文 件,虽然它不如其他格式好理解,但是它能更好的利用内存,更方便复制和...

TensorFlow学习笔记(二十四)自制TFRecord数据集 读取、显示及代码详解 (http://blog.csd...

在跑通了官网的mnist和cifar10数据之后,笔者尝试着制作自己的数据集,并保存,读入,显示。 TensorFlow可以支持cifar10 的数据格式, 也提供了标准的TFRecord 格式, 而关于...

登录 注册

X

(https://passport.csdn.net/a

[TFRecord格式数据]基本介绍 (http://blog.csdn.net/u013555719/article/details/77894827)

标准TensorFlow格式 TFRecords觉得有用的话.欢迎一起讨论相互学习~Follow MeTFRecords可以允许你讲任意的数据转换为 TensorFlow所支持的格式,这种方法可以使T...



u013555719 (http://blog.csdn.net/u013555719) 2017年09月08日 14:56

Tensorflow中使用tfrecord方式读取数据 (http://blog.csdn.net/u010358677/article/details/70...

前言本博客默认读者对神经网络与Tensorflow有一定了解,对其中的一些术语不再做具体解释。并且本博客主要以图片数据为 例选行介绍,如有错误,敬请斧正。使用Tensorflow训练神经网络时,我们可以...



Tensorflow 训练自己的数据集(二)(TFRecord)(http://blog.csdn.net/Best_Coder/articl...

1. TFRecord输入数据格式2. TFRecord格式介绍3. 将自己的数据转化为TFRecord格式准备数据转换数据import os import ten sorflow as tf fr...



深入浅出的TensorFlow数据格式化存储工具TFRecord用法教程 (http://blog.csdn.net/gg_188...

本文详尽地介绍了使用TFRecord的好处、必要性,然后介绍相关核心概念,最后把它整合到我们的模型训练代码当中。...



加入CSDN,享受更精准的内容推荐,与5000万程序员共同成长!

登录 注册