

[CSDN首页 \(http://www.csdn.net?ref=toolbar\)](http://www.csdn.net?ref=toolbar)[学院 \(http://edu.csdn.net?ref=toolbar\)](http://edu.csdn.net?ref=toolbar)[下载 \(http://download.csdn.net?ref=toolbar\)](http://download.csdn.net?ref=toolbar)[更多 ▼](#)[下载 CSDN APP \(http://www.csdn.net/app?ref=toolbar\)](http://www.csdn.net/app?ref=toolbar)[写博客 \(http://write.blog.csdn.net/postedit?ref=toolbar\)](http://write.blog.csdn.net/postedit?ref=toolbar)[登录 \(https://passport.csdn.net/account/login?ref=toolbar\)](http://blog.csdn.net/) | [注册 \(http://passport.csdn.net/account/mobile/register?ref=toolbar&action=mobileRegister\)](http://passport.csdn.net/account/mobile/register?ref=toolbar&action=mobileRegister)[首页 \(http://blog.csdn.net/\)](http://blog.csdn.net/) [移动开发 \(http://blog.csdn.net/mobile/newarticle.html\)](http://blog.csdn.net/mobile/newarticle.html) [全部 ▮](#)**CSDN** (<http://www.csdn.net>)

## 目录 Tensorflow 训练自己的数据集（一）（数据直接导入到内存）



原创 2017年04月12日 11:32:13

2377

1

2

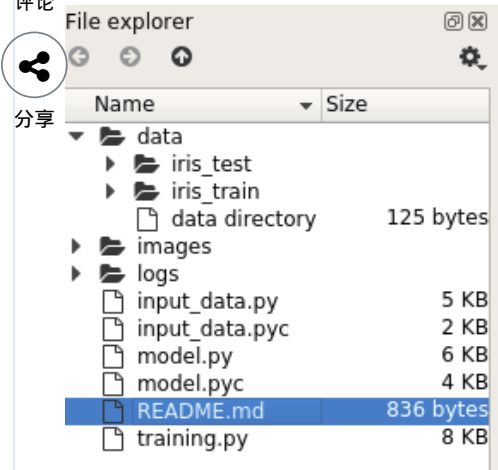
喜欢



## 收藏 制作自己的训练集



评论 下图是我们数据的存放格式，在data目录下有验证集与测试集分别对应iris\_test， iris\_train



分享

为了向伟大的MNIST致敬，我们采用的数据名称格式和MNIST类似

ruiweicas ([http://blog.csdn.net/Best\\_Coder](http://blog.csdn.net/Best_Coder))[+ 关注](#)[http://blog.csdn.net/Best\\_Coder](http://blog.csdn.net/Best_Coder)

码云

未开通

原创

粉丝

喜欢

<https://github.com/ruiweicas>

38

32

0

### 他的最新文章

[更多文章 \(http://blog.csdn.net/Best\\_Coder\)](http://blog.csdn.net/Best_Coder)

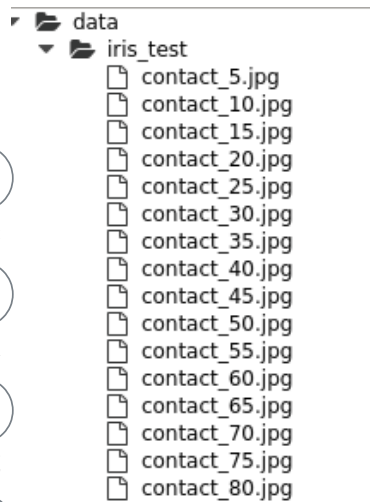
Tensorflow框架下Faster-RCNN实践  
（二）——用自己制作的数据训练  
Faster-RCNN网络（附代码）  
(/best\_coder/article/details/76615000)

Tensorflow框架下Faster-RCNN实践  
（一）——Faster-RCNN所需数据集制  
作（附代码）  
(/best\_coder/article/details/76577544)

window10 编译安装opencv3.2+contrib  
(/best\_coder/article/details/71515722)

[编辑推荐](#)[最热专栏](#)[返回顶部](#)

classification\_index.jpg



目录



喜欢



收藏



评论



分享

图像的index都是5的整数倍是因为我们选择测试集的原则是每5个样本，选择一个样本作为测试集，其余的作为训练集和验证集

生成这样数据的过程相对简单，如果有需要python代码的，可以给我发邮件，或者在我的github下载

至此，我们的训练集，测试集，验证集就生成成功了，之所以我们的文件夹只有训练集和测试集是因为我们在后续的训练过程中，会在训练集中分出一部分作为验证集，所以两者暂时合称为训练集

## 将数据集写入到Tensorflow中

### 1. 直接写入到队列中

SqueezeNet (/u014568921/article/deta...

TensorFlow——训练自己的数据（四）...

### 在线课程



【免费】深入理解Docker

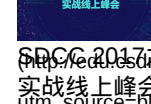
内部原理及网络配置

(http://edu.csdn.net/course/detail/563?utm\_source=blog9)

讲师: 左巍

Course/detail/563?

blog9)



SDCC 2017之区块链技术应用

(http://edu.csdn.net/course/series\_detail/66?utm\_source=blog9)

实战线上峰会

(http://edu.csdn.net/course/series\_detail/66?utm\_source=blog9)

Course/series\_detail/66?

utm\_source=blog9)



返回顶部



目录



喜欢



收藏



评论



分享

```
1 import tensorflow as tf
2 import numpy as np
3 import os
4
5 train_dir = '/home/ruyiwei/data/iris_train/' # your data directory
6 def get_files(file_dir):
7     """
8     Args:
9         file_dir: file directory
10    Returns:
11        list of images and labels
12    """
13    iris = []
14    label_iris = []
15    contact = []
16    label_contact = []
17    for file in os.listdir(file_dir):
18        name = file.split('_')
19        if name[0] == "iris":
20            iris.append(file_dir + file)
21            label_iris.append(0)
22        else:
23            contact.append(file_dir + file)
24            label_contact.append(1)
25    print('There are %d iris\nThere are %d contact' % (len(iris), len(contact)))
26
27    image_list = np.hstack((iris, contact))
28    label_list = np.hstack((label_iris, label_contact))
29
30    temp = np.array([image_list, label_list])
31    temp = temp.transpose()
32    np.random.shuffle(temp)
33
34    image_list = list(temp[:, 0])
35    label_list = list(temp[:, 1])
36    label_list = [int(i) for i in label_list]
37
```



返回顶部

38

```
return image_list, label_list
```

为了大家更方便的理解和修改代码，我们对代码进行讲解如下

1-3行：导入需要的模块

5行：定义训练集合的位置，这个需要根据自己的机器进行修改

7行：定义函数 `get_files`

18行：`os.listdir(file_dir)` 获取指定目录`file_dir`下的所有文件名词，也就是我们的训练图片名称

18行：`for file in os.listdir(file_dir):` 遍历所有的图片

19行：`name`为一个数组，由于我们根据MINIST来定制的图片名词，所以`file.split('_')`会将图像名称分为两部分，第一部分为classification，通过`name[0]`来获得分类信息。

21行、24行：`iris.append(file_dir + file)/contact.append(file_dir + file)`将图像的绝对路径放入到`iris/contact`中

22行、25行：`label_iris.append(0)/label_contact.append(1)`给对应的图片贴标签

28-29行：将二分类的图像与标签压入到list中

31-33行：合并二分类图像，然后打乱

38行：返回打乱后对应的图像与标签



目录



喜欢



收藏



评论



分享

在spyder下执行如上代码后会返回如下信息

```
In [4]: image_list, label_list = get_files(train_dir)
There are 7975 iris
There are 4288 contact
```

这样图像和标签信息就被load到了内存中，我们接下来就可以利用现有的模型对图像进行分类训练，模型的选择和训练的过程，我们会在后面进行讲解。



返回顶部

版权声明：本文为博主原创文章，如果特别强烈的想转载，那就转载吧，有问题联系ruyiwei.cas@gmail.com

 举报

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

训练模型 (<http://so.csdn.net/so/search/s.do?q=训练模型&t=blog>) /

自制数据集 (<http://so.csdn.net/so/search/s.do?q=自制数据集&t=blog>) /

深度学习 (<http://so.csdn.net/so/search/s.do?q=深度学习&t=blog>) /

机器学习 (<http://so.csdn.net/so/search/s.do?q=机器学习&t=blog>) /



目录



喜欢



收藏



评论



分享



**u011325670 (/u011325670)** 2017-08-25 21:45

1楼

(/u011325670)

回复

## 相关文章推荐

### Ubuntu启动问题以及Grub Rescue修复方法 (/xixinyan/article/details/6969469)

Ubuntu启动问题以及Grub Rescue修复方法 问题：之前系统是Windows7 64bit（C盘）+ D，E盘（都是NTFS）+ Ubuntu。今天，在...



**xixinyan** (<http://blog.csdn.net/xixinyan>) 2011-11-14 17:22

📖 0

  
返回顶部

## SqueezeNet (/u014568921/article/details/72565976)

深度学习方法（七）：最新SqueezeNet 模型详解，CNN模型参数降低50倍，压缩461倍！



u014568921 (<http://blog.csdn.net/u014568921>) 2017-05-19 21:00 239



## 精选：深入理解 Docker 内部原理及网络配置 ([http://edu.csdn.net/huiyiCourse/detail/563?utm\\_source=blog10](http://edu.csdn.net/huiyiCourse/detail/563?utm_source=blog10))

网络绝对是任何系统的核心，对于容器而言也是如此。Docker 作为目前最火的轻量级容器技术，有很多令人称道的功能，如 Docker 的镜像管理。然而，Docker的网络一直以来都比较薄弱，所以我们有必要深入了解Docker的网络知识，以满足更高的网络需求。



收藏

## TensorFlow——训练自己的数据（四）模型测试 (/xinyu3307/article/details/75008458)



评论

测试一张图片获取一张图片函数：def get\_one\_image(train): 输入参数：train,训练图片的路径 返回参数：image，从训练图片中随机抽取一张图片 n = len(train)...



分享

xinyu3307 (<http://blog.csdn.net/xinyu3307>) 2017-07-12 12:25 539

## TensorFlow——训练自己的数据（三）模型训练 (/xinyu3307/article/details/74979842)

参考：Tensorflow教程-猫狗大战数据集 文件training.py 导入文件 import os import numpy as np import tensorflow as tf im...





xinyu3307 (<http://blog.csdn.net/xinyu3307>) 2017-07-11 16:45 1113

## TensorFlow——训练自己的数据——CIFAR10（一）数据准备 (/xinyu3307/article/details/77072789)

返回顶部

参考教程:Tensorflow教程：深度学习 图像分类 CIFAR10数据集Reading Data 所用函数def read\_cifar10(data\_dir, is\_train, batch\_size):

 xinyu3307 (<http://blog.csdn.net/xinyu3307>) 2017-08-10 20:06  207

## 白平衡 (/ken\_yjj/article/details/5593779)



白平衡是摄像领域一个非常重要的概念，通过它可以解决色彩还原和色调处理的一系列问题。白平衡是随着电子影像再现色彩真实而产生的，在专业摄像领域白平衡应用的较早，现在家用电子产品（家用摄像机、数码照相机）中...



 ken\_yjj ([http://blog.csdn.net/ken\\_yjj](http://blog.csdn.net/ken_yjj)) 2010-05-14 23:15  933

喜欢





## Tensorflow框架下Faster-RCNN实践（二）——用自己制作的数据训练Faster-RCNN网络（附代码） (/best\_coder/article/details/76615000)



Tensorflow框架下Faster RCNN实践





 Best\_Coder ([http://blog.csdn.net/Best\\_Coder](http://blog.csdn.net/Best_Coder)) 2017-08-03 11:12  662



分享

## 利用TensorFlow Object Detection API 训练自己的数据集 (/shz\_0519/article/details/73471128)

利用谷歌TensorFlow Object Detection API 训练自己的数据集今天终于通过Tensorflow Object Detection API中的faster\_rcnn\_inception\_v2\_train.py训练成功...

 shz\_0519 ([http://blog.csdn.net/shz\\_0519](http://blog.csdn.net/shz_0519)) 2017-06-19 18:37  3340

## Tensorflow 训练自己的数据集（一）（数据直接导入到内存） (/best\_coder/article/details/70141075)

制作自己的训练集下图是我们数据的存放格式，在data目录下有验证集与测试集分别对应iris\_test，iris\_train 为了向伟大的MNIST致敬，我们采用的数据名称格式和MNIST类似 ...



Best\_Coder ([http://blog.csdn.net/Best\\_Coder](http://blog.csdn.net/Best_Coder)) 2017-04-12 11:32 2378

### Fast RCNN训练自己的数据集 (/jxl\_1207/article/details/72900315)



转载请注明出处，楼赙(yi)航的blog，<http://www.cnblogs.com/louyihang-loves-baiyan/> <https://github.com/YihangLou/>...

目录



JXL\_1207 ([http://blog.csdn.net/JXL\\_1207](http://blog.csdn.net/JXL_1207)) 2017-06-07 15:32 205



喜欢

### Faster-RCNN训练自己的数据集 (/gvfdbdf/article/details/52214008)



收藏

就目前我在了解到的资料来看，训练RCNN系列，一般有两种思路：修改CNN的工程代码，适应自己的数据集格式 将自己的数据格式做成VOC2007形式的数据集 从工作量来看，无疑后者更容易一些，本文的思路...



评论



GVFDDBDF (<http://blog.csdn.net/GVFDDBDF>) 2016-08-15 19:34 5313



### darknet yolov2 训练自己的数据集 (/qq\_27840681/article/details/63682694)

分享

ubuntu14.04+opencv2.4.10+cuda7.5+cudnn5.0 训练集：20张图片 00001.jpg-00020.jpg 测试集：5张图片 00021.jpg...



qq\_27840681 ([http://blog.csdn.net/qq\\_27840681](http://blog.csdn.net/qq_27840681)) 2017-03-19 10:53 629

### TensorFlow入门之训练mnist数据集 (/rongrongyaofeiqi/article/details/65445830)

```
import sys,os
import numpy as np
import tensorflow as tf
from tensorflow.examples.tutorials.mnist import im...
```




rongrongyaofeiqi (<http://blog.csdn.net/rongrongyaofeiqi>) 2017-03-23 17:06 246



## caffe训练自己的数据集 (/fahail/article/details/74935756)



利用python生成txt文件后，就可以利用caffe自带工具将图片转换为lmdb格式，其代码如下#!/usr/bin/env sh # Create the imagenet lmdb inputs...

 fahail (<http://blog.csdn.net/fahail>) 2017-07-10 19:31 171

## 用YOLOv2训练自己的数据集 (/zhy8623080/article/details/73188542)


目录

1. 系统初始环境系统:Ubuntu16.04: ubuntu-16.04-desktop-amd64.isocuda安装文件: cuda-repo-ubuntu1604-8-0-local\_8.0...

 喜欢  zhy8623080 (<http://blog.csdn.net/zhy8623080>) 2017-06-13 17:08 602

## Faster RCNN 训练自己的数据集 (/apple\_\_fly/article/details/51712402)

收藏

 Faster RCNN 训练自己的数据集

 评论  Apple\_\_fly ([http://blog.csdn.net/Apple\\_\\_fly](http://blog.csdn.net/Apple__fly)) 2016-06-19 14:55 1952

## 分享用SSD训练自己的数据集 (/dongfang1984/article/details/74640219)

1构建 数据集 先来看一下我们构建数据集应该是什么样的,假设总数据为1000张。为了方便，我们将数据放在 / home / bin golwang/data 文件夹下。 /home/bingolwa...

 dongfang1984 (<http://blog.csdn.net/dongfang1984>) 2017-07-07 11:11 251

## Detection : pvanet训练自己数据 (/byplane/article/details/72472805)

pvanet训练自己数据

 byplane (<http://blog.csdn.net/byplane>) 2017-05-18 16:17 561

  
返回顶部