

# 面向深度学习目标类型识别任务的样本标注软件

## 操作手册

### 目录

简介.....	2
1. 生成训练列表.....	2
1.1. 输入.....	2
1.2. 操作流程.....	2
1.3. 输出.....	3
2. 获取文件列表.....	4
2.1. 输入.....	4
2.2. 操作流程.....	4
2.3. 输出.....	5
3. 收缩验证列表.....	5
3.1. 输入.....	5
3.2. 操作流程.....	5
3.3. 输出.....	6
4. 挑选样本 ImageNet.....	6
4.1. 输入.....	6
4.2. 操作流程.....	7
4.3. 输出.....	8
5. 子目录合并.....	8
5.1. 输入.....	8
5.2. 操作流程.....	8
5.3. 输出.....	8
6. 文件批量重命名等功能.....	9
6.1. 输入.....	9
6.2. 操作流程.....	9
6.3. 输出.....	10
总结.....	10

## 简介

本软件是一套面向深度学习目标类型识别任务的样本标注软件，全部功能基于 C++ 开发，适用于为目标识别网络标注数据，包括从现有数据集精选数组，或制作自己的数据集。其中提供的批量自动处理功能，能够替代大部分人工处理样本的工作，能够节省大量时间。本软件具有生成训练列表、获取文件列表、收缩验证列表、挑选样本 ImageNet、子目录合并、文件批量重命名等功能。功能概述如下：

生成训练列表——针对分类任务，搜集多个类别的图片，按类存放样本，只需辅助一个样本索引文件，结合本助手，就能生成训练列表文件，免去手动标注的工作。

获取文件列表——遍历某个目录，获取所有文件的列表，往往在测试时需要用到该列表。该功能还支持 2 个延伸功能：嵌套遍历子目录、列表乱序。

收缩验证列表——精选某些类别的验证集子集，也可用于训练集列表的收缩精选。

挑选样本 ImageNet——按照文件列表挑选样本图像子集，输出到一个新目录。

子目录合并——将多个子目录里面的文件重命名拷贝到一个目录。

文件批量重命名——把一个文件夹里面的文件名按照一定的前缀统一重命名。

## 1. 生成训练列表

针对分类任务，搜集多个类别的图片，按类存放样本，只需辅助一个样本索引文件，结合本助手，就能生成训练列表文件，免去手动标注的工作。

### 1.1. 输入

(1) 每个目录放置一个类别的图像样本，目录以类别命名，名称要求是英文；

(2) 一个类别索引文本文件 classIndex.txt，内容一行一个类别名称；

### 1.2. 操作流程

(1) 点击“工具”菜单，点击子菜单“生成训练列表”，如下图：

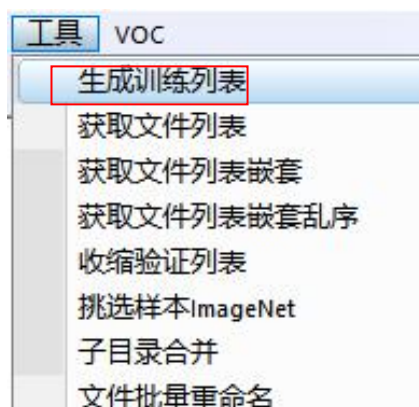


图 1-1 菜单项“生成训练列表”

(2) 选择文件夹，如下图：

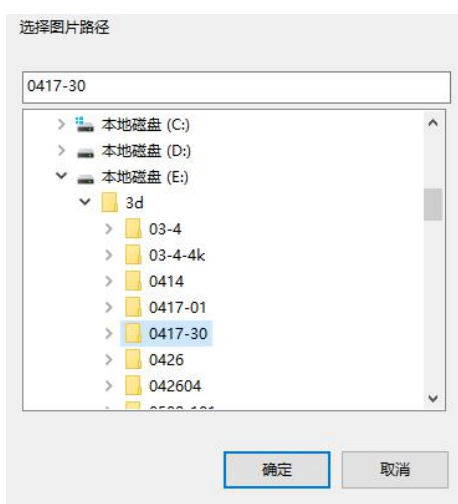


图 1-2 选择图片路径

### 1.3. 输出

输出一个训练列表文件 `filelist.txt`，格式如下：

相对路径\样本图像名称 类别索引

如下图所示：

```

new 1 x new 2 x filelist.txt x
91 boat/virtual_boat67.jpg 0
92 boat/virtual_boat7.jpg 0
93 boat/virtual_boat8.jpg 0
94 boat/virtual_boat9.jpg 0
95 buoy/110.jpg 1
96 buoy/111.jpg 1
97 buoy/114.jpg 1
98 buoy/115.jpg 1
  
```

图 1-3 训练列表示例

## 2. 获取文件列表

遍历某个目录，获取所有文件的列表，往往在测试时需要用到该列表。

该功能还支持 2 个延伸功能：嵌套遍历子目录、列表乱序。

### 2.1. 输入

(1) 一个目录放置了一个类别的图像样本；

### 2.2. 操作流程

(1) 点击“工具”菜单，点击子菜单“获取训练列表”，如下图：

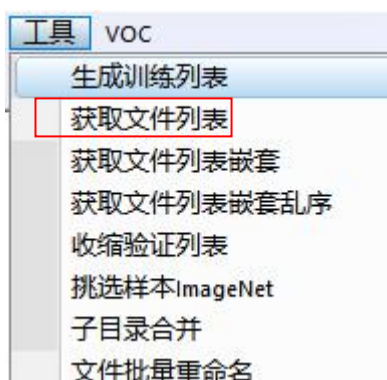


图 2-1 菜单项“获取训练列表”

如果选择“~嵌套”，除了遍历所选目录，还会遍历其中的子目录；

如果选择“~嵌套乱序”，文件列表还将打乱顺序，便于训练时不同类别交替进行，避免过拟合；

(2) 选择文件夹，如下图：



图 2-2 选择图片路径

## 2.3. 输出

输出一个列表文件 `filelist.txt`，格式如下：

样本图像 1 名称

样本图像 2 名称

如下图所示：

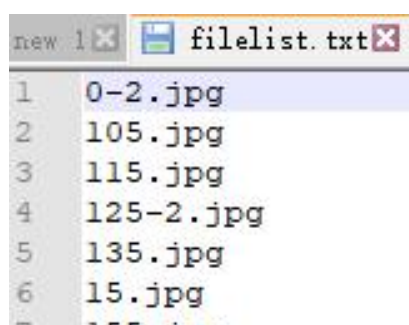


图 2-3 文件列表示例

## 3. 收缩验证列表

精选某些类别的验证集子集，也可用于训练集列表的收缩精选。

### 3.1. 输入

(1) 初始验证列表 `val-full.txt`，文件每行格式：验证样本名称 类别索引，如下图：

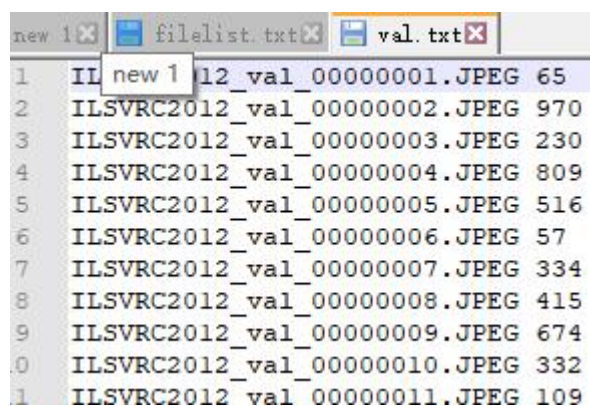


图 3-1 菜单项“获取训练列表”

(2) 所需类别索引文件 `classNo.txt`，文件每行包含一个类别索引；

### 3.2. 操作流程

(1) 点击“工具”菜单，点击子菜单“收缩验证列表”，如下图：

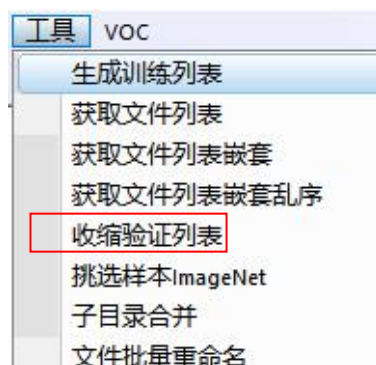


图 3-2 菜单项“收缩验证列表”

(2) 选择文件夹，如下图：

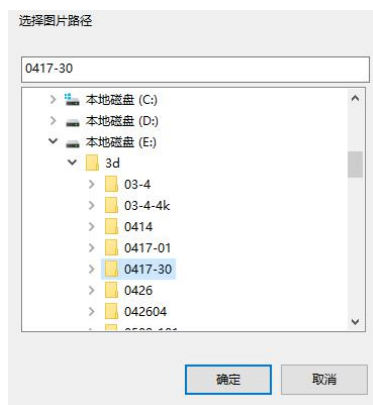


图 3-3 选择图片路径

### 3.3. 输出

输出一个列表文件 val.txt，只包含 classNo.txt 所列类别的验证集列表。

## 4. 挑选样本 ImageNet

按照文件列表挑选样本图像子集，输出到一个新目录。

### 4.1. 输入

(1) 即将挑选的文件列表 val.txt，文件每行格式：样本名称 类别索引，如下图：

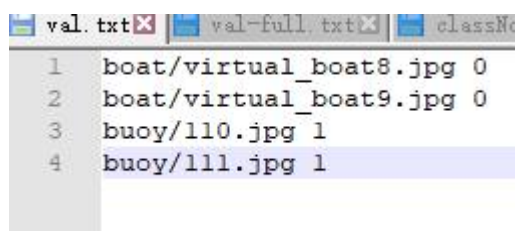


图 4-1 挑选文件列表

(2) 包含 val.txt 所列样本文件的文件夹，跟列表文件在同一目录，如下图：

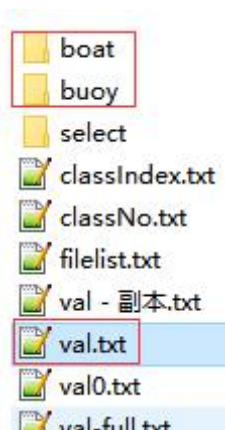


图 4-2 待挑选文件夹

## 4.2. 操作流程

(1) 点击“工具”菜单，点击子菜单“挑选样本 ImageNet”，如下图：

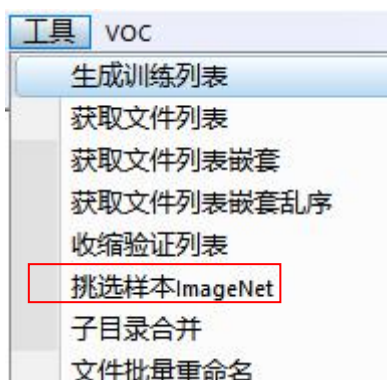


图 4-3 菜单项“挑选样本 ImageNet”

(2) 选择文件夹，如下图：

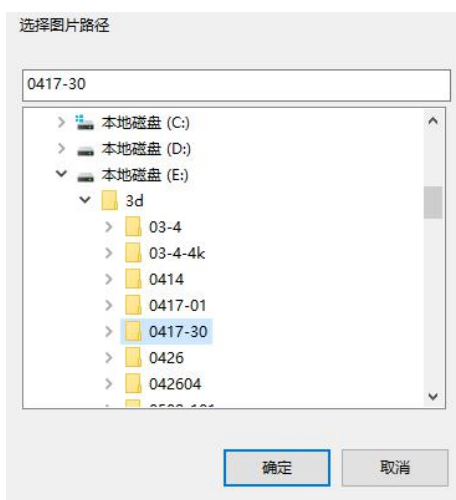


图 4-4 选择图片路径

### 4.3. 输出

新建 select 文件，里面包含挑选出来的样本图像文件。

## 5. 子目录合并

将多个子目录里面的文件重命名拷贝到一个目录。

### 5.1. 输入

同 4 - 挑选样本 ImageNet

### 5.2. 操作流程

(1) 点击“工具”菜单，点击子菜单“子目录合并”，如下图：

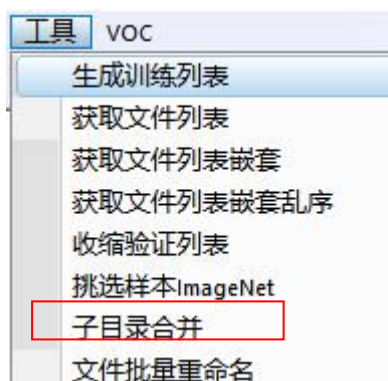


图 5-1 菜单项“子目录合并”

(2) 选择文件夹，如下图：

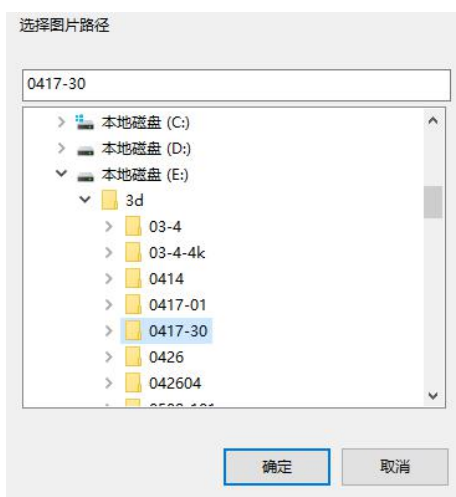


图 5-2 选择图片路径

### 5.3. 输出

新建 select 文件，里面包含挑选出来的样本图像文件，所有文件在同一目录



下，子目录名称作为文件名的前缀，如下图：





	-boat-virtual_boat8.jpg	2020/7/14 16:54
	-boat-virtual_boat9.jpg	2020/7/14 16:54
	-buoy-110.jpg	2019/1/11 6:24
	-buoy-111.jpg	2019/1/11 6:24

图 5-3 合并目录结果

## 6. 文件批量重命名等功能

把一个文件夹里面的文件名按照一定的前缀统一重命名。

### 6.1. 输入

包含一系列图片样本文件的目录，支持嵌套，将拷贝到新的子目录 select，并按照一致的前缀统一重命名。

### 6.2. 操作流程

(1) 点击“工具”菜单，点击子菜单“文件批量重命名”，如下图：

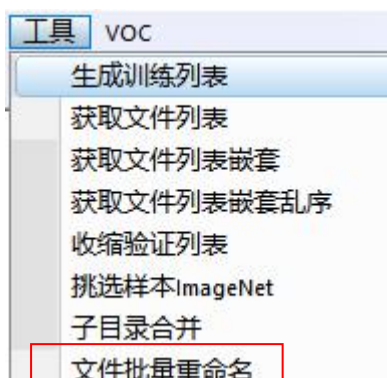


图 6-1 菜单项“文件批量重命名”

(2) 选择文件夹，如下图：



图 6-2 选择图片路径

### 6.3. 输出

新建子目录 `select`，输入目录下的所有文件（含嵌套子目录）拷贝到新目录，并按照一致的前缀统一重命名。

### 总结

以上文档内容详细介绍了软件的使用方法——面向深度学习目标类型识别任务的样本标注软件，提供了包括输入数据要求、功能调用步骤以及预期输出等信息。本软件全部功能基于 C++ 开发，适用于为目标识别网络标注数据，包括从现有数据集精选数组，或制作自己的数据集。具有生成训练列表、获取文件列表、收缩验证列表、挑选样本 ImageNet、子目录合并、文件批量重命名等功能。