

# 标注文件目录结构和文件命名规范

## 1. 目录结构及图片命名规范

### 1.1 肠镜息肉数据集

目录结构应该以病人为单位

以cw肠息肉0008426516\_025-息肉1.mp4\_000114.518.jpg 为例子

上级目录

-- 0008426516_025	<-----病例报告id，单独切唯一
-- meta.txt	<-----一些病例属性信息(肠息肉)，可选
-- polyp_1	<-----息肉ID，以片段为单位
-- missed	<-----根据属性分配图片文件夹(漏报,多角度)
-- 000114_518.jpg	<-----截图
-- ...	
-- 000114_689.jpg	
-- general	
-- 000114_518.jpg	
-- ...	
-- 000114_689.jpg	
-- ....	
-- false_positives	<-----误报截图
-- 000114_689.jpg	

## 1.2 胶囊内镜分类数据集

### 上级目录

-- 0008426516	<-----1级目录，以病例号命名
-- videos	<-----2级目录，命名为videos
-- 0	<-----3级目录，以视频号命名 从0开始，按顺序递增 每个视频要有单独文件夹
-- report.pdf	<-----诊断报告，命名为report 可以是pdf、word、txt等格式
-- video.avi	<-----胶囊内镜视频，命名为video 可以是mp4、avi等格式
-- frames	<-----4级目录，命名为frames 存放视频中的所有帧 可使用VLC，详见下面的链接
-- 0000001.jpg	<-----视频的每一帧图片， 图片名为视频中的帧号
-- ...	
-- 0010000.jpg	
-- 1	
-- ...	
-- disease	<-----2级目录，命名为disease 此目录中放分类的图片
-- 0	<-----3级目录，以视频号命名 编号必须对应原视频的编号 若图从0008426516/videos/ <b>0</b> /frames来， 则此目录也命名为 <b>0</b>
-- ulcer	<-----4级目录，根据病变命名 是溃疡，则ulcer；是糜烂，则erosion 其他病变的英文名详见本文件第3部分
-- ...	
-- erosion	
-- 0000100.jpg	<-----从frames中选取的有病变的图片 不要改动图片原有的名称 如果一张图上有多个病变， 则这几个病变的目录下都要包含此图
-- ...	
-- 0000110.jpg	
-- 0000600.jpg	
-- ...	
-- 0000610.jpg	
-- ...	

胶囊内镜数据集整理详细步骤参看：

[https://github.com/wangxizhe2017/XY\\_Cap\\_Endo\\_Dataset\\_Guide.git](https://github.com/wangxizhe2017/XY_Cap_Endo_Dataset_Guide.git)

## 2. 不应出现的符号

- 1) 中文
- 2) 空格
- 3) 各种括号
- 4) 点(除了.jpg, .png, .mp4等文件格式的点)
- 5) - (中杠)
- 6) \
- 7) /
- 8) 其它标点符号

这些符号应该使用下划线 \_ 来代替, 并且为英文半角。

## 3. 胶囊内镜分类数据集病变名称中英文对照

	病变名称中文	病变名称英文 ( 单词之间用 <u>下划线</u> )
1	溃疡	ulcer
2	糜烂	erosion
3	粘膜充血水肿	mucous_membrane_hyperemia_edema
4	血管改变	vascular_changes
5	隆起	protuberant
6	淋巴滤泡增生	lymph_follicle_hyperplasia
7	息肉	polypide
8	异物	foreign_body
9	疤痕	scar
10	出血	bleeding
11	白色颗粒样变	white_granular_degeneration
12	绒毛改变	villous_change
13	憩室	diverticulum
14	狭窄	stenosis
15	虫体	parasite