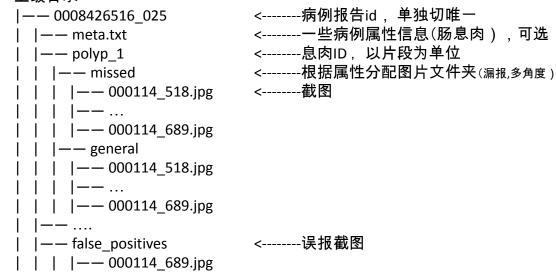
标注文件目录结构和文件命名规范

- 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



1.2 胶囊内镜分类数据集

上级目录 <----1级目录,以病例号命名 |--0008426516<-----2级目录,命名为videos I——videos <-----3级目录,以视频号命名 编号从0开始,按顺序递增 每段视频要有单独文件夹 <-----诊断报告,命名为report |—— report.pdf 可以是pdf、word、txt等格式 <-----胶囊内镜视频,命名为video |−− video.mp4 可以是mp4、avi等格式 <-----4级目录,命名为frames |—— frames 存放视频中的所有帧 | |—— 0000001.jpg <-----视频的每一帧图片, 图片名为视频中的帧号 | |-- 0010000.jpg

—— disease	<2级目录,命名为disease
1 1	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/ 0 /frames来
	则此目录也命名为 0
	<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第3部分
— 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
— 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

- 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
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_foilicie_hyperplasia
7	虫体	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