Dataset：用来处理单个样本，如何从磁盘中提取特征、标签，还会进行相应的预处理，并映射成相应的X Y

自定义dataset必须实现三个函数，\_\_init\_\_ \_\_ len \_\_ \_\_getitem\_\_

Init方法中，首先要传入文件名和其所在的目录，annotation\_file（文件名）、img\_dir（文件目录）

len函数，是返回样本长度

getitem是通过idx返回一个样本，img\_path是照片相应的路径，image提取出照片，label标签

源码中：

Init函数：

dataset是实体化的对象

batch\_size默认值是1但通常要调大一些

shuffle在每个训练周期以后是否需要对数据进行打乱

sampler是决定了我们怎样对数据进行采样

batch\_sampler同sampler

num\_workers默认值为0意为默认使用主进程加在数据，可以改变数据但要根据设备的cpu个数

pin\_memory可以选择是否将tensors保存在gpu中

drop\_last是否丢掉多与数据

collare\_fn对采样的小批次进行再处理，输入是batch输出也是batch