**一、全切片数据、标注数据及转格式数据**

所有切片原始数据保存在160服务器的H:\TCTDATA。备份数据在两个硬盘中，其中9.09T的盘中保存了3D，our，szsq\_sfy1, szsq\_sfy3,szsq\_sfy5。另一个盘中保存了szsq\_tj6,szsq\_tj7,szsq\_tj8,szsq\_xyw1,szsq\_xyw2。其他未备份的数据需要10T空间保存。

标注文件保存在上述原始数据下对应的LabelFiles文件中。其中model3的标注文件保存在160服务器的下面路径下：H:\transSrp\SZSQ\_originaldata\Model3Labelfiles。整理后的标注文件保存在L:\业务文档\TotalLabelFiles路径下，其中originalBackup中为原始标注文件的备份，xmlLabelFiles中为将所有标注和阳性标注的XML文件，其中将所有标注名称进行统一，其对应名称如下所示：

posdic={

'AGC':['AGC','2213','2222'],

'ASCUS':['ASC-US','ASCUS'],

'ASCH':['ASC-H','ASCH'],

'HSIL':['HISL','HSIL'],

'LSIL':['LSIL','LISL'],

'SCC':['SCC'],

'AIS':['2230'],

'Tumor':['3000'],

'Pos':['pos']}

negdic={

'HCG':['HCG','Normal-0400',],

'NGEC':['nGEC',],

'AMC':['2212'],

'AEC':['2211'],

'IUD':['1123'],

'Mould':['mould'],

'Trichomond':['1210','trichomonad'],

'Conified':['1112'],

'Actinomycetes':['1240'],

'Metaplasia':['1111','1113'],

'Inflammation':['1121'],

'Herpes':['1250'],

'Flora':['1230'],

'Bubble':['bubble'],

'Nplus':['nplus','several','singel','tens'],

'Neg':['Normal','null','Normal-0100','Normal-0300','Normal-0500','Normal-0600','0600']}

转格式数据分别保存在下面的路径中：G:\transSrp路径中保存的为3D、Tianchi、szsq\_tj4、szsq\_tj5的数据，其余批次的转格式数据保存在H:\transSrp路径下。

**二、样本标注格式及路径**

用于推荐系统的AdaptData，origin文件夹中为原始分辨率下尺寸为768\*0.586/mpp的样本数据，gamma\_simple中为origin数据经过gamma变换，并resize到1536下的样本数据。初始的AdaptData保存在H:\AdaptDATA路径下，后续model3标注中的阳性标注保存在K:\AdaptDATA路径下，BD的阳性标注保存在K:\AdaptDATA\BD路径下，阴性标注，其中tj10,tj11的所有阴性与阳性标注在相同的文件夹，而xiehe的n保存在J:\xiehe，n\_w保存在N:\xiehe，肿瘤医院的n保存在M:\srp\_zhongliuyiyuan1，n\_w保存在N:\zhongliuyiyuan1中。

用于model3的数据，为原始分辨率下尺寸为512\*0.293/mpp再resize到512下的样本数据，其保存路径为H:\model3\data。

测试切片的名称保存在该路径下H:\fql\fqlcode\utils\totalTestSlide.json，其中BD切片的测试切片名称保存在该路径下H:\fql\fqlcode\utils\BDTestSlide.json。

困难切片名称保存在该路径下H:\fql\fqlcode\utils\ hardslidesname.json。各个批次困难切片的标注文件在Labelfiles下对应原始标注的小文件夹里，如szsq\_xyw2的困难切片的全标注标注文件在H:\TCTDATA\SZSQ\_originaldata\Labelfiles\xml\_xyw2\xml\_xyw2路径下。困难样本的标注详情在表szsqlHardSlides.xlsx中，集中的困难样本切片及标注在本机的E:\hardSlide\a\_new路径下。

**三、各种批次中间结果的路径**

模型为model1\_700,model2\_658,location\_36,modle3模型为7th\_v2\_360，使用韩威提供的工程化工具（H:\fql\Model3CheckTools），计算的所有中间结果保存在路径H:\fql\rnnResult\rnn1000下。

使用最新模型（偏向于同济数据的）model1\_236,model2\_200,location\_128, modle3模型为7th\_v2\_360，计算刘思博RNN未使用的其余约4000张切片的中间结果保存在L:\ModelResult\new下。4000张切片的具体情况可以参考数据详情.xlsx中的lsb4000一表，其中绿色部分为已经完成计算的。

**四、交叉检查和亚类数据**

交叉检查数据为model3中人工挑选的细胞与标注不同的标注，用于标注组进行再次检查。数据保存路径为：L:\crossCheckData\problem。

亚类数据为从model3的阴性中挑选边缘细胞、成团腺细胞、基底细胞萎缩细胞和角化细胞、化生细胞，该类细胞保存在L:\sub\_nplus路径下。

**五、相关代码**

韩威提供的计算中间结果的工程化工具保存路径为H:\fql\Model3CheckTools，会输出model2推荐的前十个区域，model3推荐的前十个区域，以及model3Input的50个区域，以及rnn的结果等。使用时需要在model以及rnn文件夹中替换相应的模型，并在config中修改模型的路径。

韩威提供的制作阳性样本的工具保存F:\MrHan\AILabTools\sample\_generate路径下，但是由于当时还未对所有标注名称进行整理，可能工具中设置的阳性名称不完整，还需要进行测试。

转格式代码保存在H:\fql\fqlcode\utils\all2srp.py中，只需更换原始数据的路径以及保存srp格式的路径即可。

之前整理的代码放在H:\fql\fqlcode路径下，其中generateAdaptData为制作样本的代码，model1、binaryModel2、locationModel为训练model1、model2以及定位模型的代码以及patch测试的代码，H:\fql\fqlcode\recommend\ predictOptimizeCombine.py为目前从全切片读图加model1，model2，model3整个流程对应的Python代码。H:\fql\fqlcode\weights\validWeight中保存了现有几批重要权重。该代码备份在2019.12.31号购入的4T硬盘中。