

Dear all ,

我是洛加大（苏州）先进技术研究院的大数据应用工程师，目前，[我们公司大数据项目组正在智能生物医疗领域做探索](#)。我们想以“天池医疗大赛-肺部结节智能诊断”为起点，开始人工智能在生物医学领域的研发。

Gihub 上的代码是我自己开发，并且已经在公司服务器上用户天池数据集来训练模型和预测测试图像肺部结节坐标和疑似结节为真结节的概率。我在 Github 上对“肺部结节智能诊断”提交了 2 个项目：pulmonary\_nodules\_AI\_diagnosis 和 pulmonary-nodules-segmentation。这些项目是我在开发过程中递增式地提交的，所以有部分代码我已经不使用了（expired，这些程序是我开发中的调试程序）。这个“肺部结节智能诊断”项目代码的运行顺序是：

1、天池 Train Set, Val Set 中 CT（.mhd）图像的预处理，[使用项目 pulmonary-nodules-segmentation 中的代码](#)如下：

```
*_dataset_preprocessing_2DUnet.py  
或者，  
*_dataset_mask_extraction.py , *_dataset_segment_lung_ROI.py
```

2、使用 U-net 深度卷积神经网络训练肺结节分割器，代码如下：

```
deep_learning_2DUnet.py  
或者，  
deep_learning_unet.py
```

3、对测试图像数据集的预测肺部结节 mask 图像（JPEG），使用 openCV 进行图像的亮度调高，代码如下：

```
adjust_preds_image_brightness.py
```

然后，使用 openCV 提取测试图像数据集肺部结节 mask 图像中疑似结节的坐标，代码如下：

```
imgs_mask_test_coordinate.py
```

4、对天池测试数据集中的肺部 CT 图像按照一个个立方体方式进行切割 进行切割，包括统一肺结节的坐标，并且把分离出来的肺部，生成肺部图像对应切割立方体的世界坐标和.npy 二维图像，代码如下：

```
test_dataset_extract3d_lung.py  
convert_voxel_to_world.py  
test_dataset_extract_nodules_coord.py
```

5、找到疑似结节后，使用常见的图像分类算法(如 CNN 等)对疑似结节进行分类，得出疑似肺结节是否为真正肺结节的概率，使用项目 `pulmonary_nodules_AI_diagnosis` 中的代码，Caffe 深度卷积神经网络训练中配置文件的生成代码如下：

```
caffe_CNN_training/generate_caffe_train_file.py
caffe_CNN_training/generate_caffe_val_file.py
caffe_CNN_training/generate_caffe_det_synset_words_file.py
caffe_CNN_training/generate_caffe_synset_words_file.py
caffe_CNN_training/generate_caffe_synsets_file.py
```

然后，训练 Caffe 深度卷积神经网络模型。待 Caffe 深度学习模型训练好后，统计肺结节是否为真正肺结节的概率，使用项目 `pulmonary-nodules-segmentation` 中的代码如下：

```
caffe-recognition/nodule_classification.py
```

6、提交 `submission.csv` 文件，第一行标记每一列的名称，分别为图像 ID 号,坐标和概率。从第二行之后的每一行都标记一个检测到的结节，坐标为检测到的结节的中心坐标  $x, y, z$  的数值。其代码如下：

```
tianchi-submission/generate_submission.py
```

上述就是“天池肺部结节智能诊断”的整个代码运行流程。

*Best Regards,*

*Yingyi Wu*