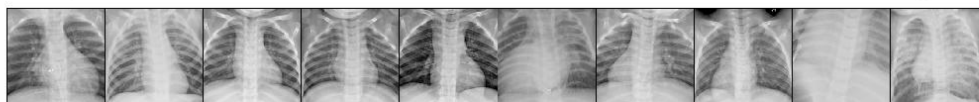


# 机器学习课程 期末报告

---使用 CNN 来检测病人是否患有肺癌

- 1.数据源 kaggle, 链接: <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>
- 2.因为数据量较大, 因此本文件夹里没有包含数据集, 若需要试运行, 只需要更改 detect\_pneumonia.py 里的 data\_dir 即可
- 3.思路原理: 数据集的图片分辨率不一, 首先将这些图片缩放到 256x256 后统一裁剪到 254x254 的大小, 文件夹里有 5216 张训练图片和 624 张测试图片, 这些图片有具有标签 pneumonia、normal 两种类型, 然后建立基本的 CNN 网络模型, 卷积层-->激励函数(RELU)-->池化层-->卷积层-->激励函数(RELU)-->池化层-->全连接层 x 3, 优化函数采用随机梯度下降算法, 损失函数采用交叉熵。采用 RELU 是为了避免梯度消失、梯度爆炸等原因,
- 4.训练中的随机 10 张图片以及标签为

['PNEUMONIA', 'NORMAL', 'NORMAL', 'NORMAL', 'NORMAL', 'PNEUMONIA', 'NORMAL', 'NORMAL', 'PNEUMONIA', 'PNEUMONIA']



- 5.完成 9 轮 Epoch 后, 可以看到测试集上的准确率最高达到了 0.77

```
Epoch: 5 | train loss: 0.2951 | test accuracy: 0.75
Epoch: 6 | train loss: 0.1060 | test accuracy: 0.76
Epoch: 6 | train loss: 0.0081 | test accuracy: 0.74
Epoch: 6 | train loss: 0.0054 | test accuracy: 0.77
Epoch: 6 | train loss: 0.2681 | test accuracy: 0.70
Epoch: 6 | train loss: 0.1351 | test accuracy: 0.76
Epoch: 6 | train loss: 0.0259 | test accuracy: 0.73
Epoch: 7 | train loss: 0.1554 | test accuracy: 0.75
Epoch: 7 | train loss: 0.1032 | test accuracy: 0.75
Epoch: 7 | train loss: 0.5196 | test accuracy: 0.77
Epoch: 7 | train loss: 0.0195 | test accuracy: 0.74
Epoch: 7 | train loss: 0.1183 | test accuracy: 0.74
Epoch: 7 | train loss: 0.2165 | test accuracy: 0.75
Epoch: 8 | train loss: 0.0553 | test accuracy: 0.75
Epoch: 8 | train loss: 0.0194 | test accuracy: 0.73
Epoch: 8 | train loss: 0.2030 | test accuracy: 0.73
Epoch: 8 | train loss: 0.0937 | test accuracy: 0.72
Epoch: 8 | train loss: 0.0056 | test accuracy: 0.73
Epoch: 8 | train loss: 0.0480 | test accuracy: 0.74
```

6.总结：通过 CNN 的训练初步使分类器具有了基本识别肺癌与普通肺癌的能力，但是最终分类器在测试集上的正确率只达到了 77%的水平，通过后续的训练分数不升反降，明显发生了过拟合，若要后续优化分类器的识别准确率，需要优化 CNN 的网络结构。

7.文件夹中的 `detect_pneumonia.py` 可直接运行。