



大数据，成就未来



期末大作业

背景介绍

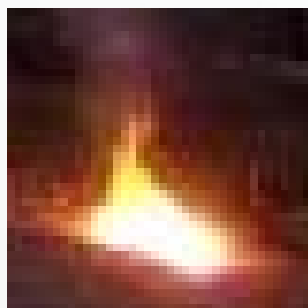
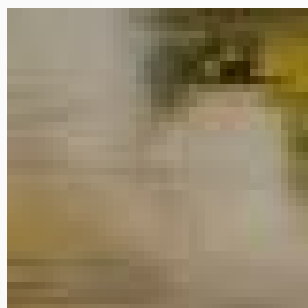
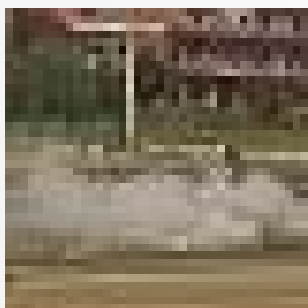
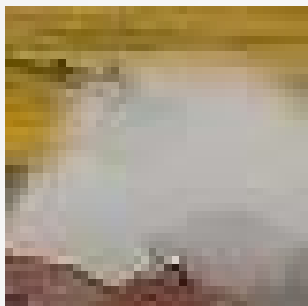
2018年中，武汉纺织大学数计学院与中国移动湖北分公司合作，共同开发一款基于户外监控摄像头的山火/非法焚烧秸秆的预防系统。希望能够在最短的时间内基于监控画面确定是否有烟火发生，然后人工快速介入，确定是否是山火/非法焚烧秸秆的事件。最终交由当地的联防/公安/森林等部门进行快速响应。

我们因此采集到了海量的户外图像，其中大致分为两类：没有任何烟火的图像，有明显的烟/火出现的图像。图像的获得是基于经过培训的人工判断然后直接从监控画面上截图。

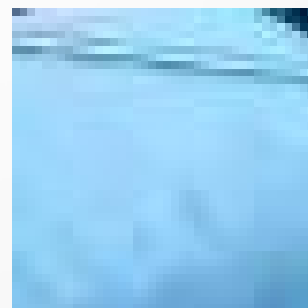
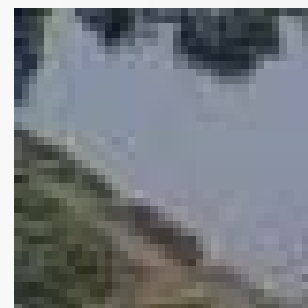
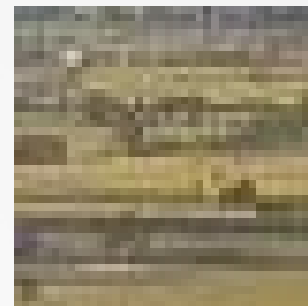
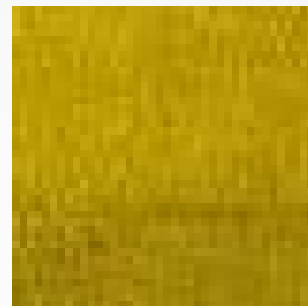
本次大作业的目标是基于提供的图像，获得一个识别烟火的图像分类器。
要求

- (1) 分类准确性 >0.85 ， $AUC>0.9$
- (2) 提供误报率和漏报率的结果，你认为那个指标相对更加重要？如何提升这个更加重要的指标呢？

有明显的烟/火出现的图像



没有烟/火出现的图像



大作业提交要求

- (1) 遵循完整的数据挖掘流程，清晰的体现从数据探索、数据预处理、数据建模到模型评估的整个过程
- (2) 至少采用两种传统机器学习算法（SVM/神经网络/随机森林等）进行建模分析
- (3) (bonus+20分) 采用CNN建模，是否能获得比传统分类算法更好的性能？并对比两种方法的程序执行效率。
- (4) 采用AdaBoost构建分类器，对比模型性能。

参考文献：

- (1) 暗通道去雾算法原理及实现
<https://www.jianshu.com/p/df9c963a392a>
- (2) Adaboost 算法
<https://www.jianshu.com/p/389d28f853c0>
- (3) 西瓜书

大作业提交要求

- (1) 必须用Jupyter Notebook完成，保存为ipynb格式提交
- (2) 提交文件名为 **“Python数据分析大作业+班级+姓名.ipynb”**
- (3) 提交到刘顿同学的邮箱 **810792334@qq.com**
- (4) 提交的截止时间为**2020年5月29日下午17:00**



大数据，成就未来



Thank you!