

Andrew Ng机器学习课程笔记11——图像文字识别

yalewoo 最后修改于 2017-03-10 发表于 2016-07-10 1,159 1
Ng笔记 Ng笔记, 机器学习

图像文字识别的任务是从一张给定的图片中识别文字。不同于文档的文字识别，图像文字识别首先要发现图像中哪个地方有文字。然后在进行文字识别。



完成这个任务的流水线如下：



1. 文字侦测 (Text Detection) —— 确定图片中的文字在哪里



[目录开关]

- 1 滑动窗口 (Sliding windows)
 - 2 获得大量人工数据
 - 3 上限分析 (Geiling Analysis)
- Character Segmentation) —— 把文字分割成单一的字符

ANTIQUE MALL

3. 字符分类 (Character Classification) —— 确定每个字符是什么



滑动窗口 (Sliding windows)

滑动窗口用来从图像中抽取对象。

假设要在图片中寻找行人。下图中的绿框就是滑动窗口，它首先从左上角开始。



然后调用模型，判断滑动窗口内的内容是不是行人。然后，按照一定的规则移动滑动窗口，一直检测完整个图片



还可以改变窗口大小，然后重新监测整张图片。最终结果是这样的

[目录开关]

- 1 滑动窗口 (Sliding windows)
- 2 获得大量人工数据
- 3 上限分析 (Geiling Analysis)



滑动窗口的思路也被用于文字识别中的文字探测。这时模型可以区分字符和非字符。找到图片中的字符后，把相邻的字符结合起来作为文字区域即可。单个的字符可能被舍弃。



获得大量人工数据

当我们的模型高偏差时，获得更多训练模型会提高性能。有时可以把已有的数据进行一些处理，获得更多的数据。

例如：对已有的训练集中的文字进行扭曲和旋转，获得新的数据。

上限分析（Geiling Analysis）

文字识别中的流水线，如何判断提高流水线的哪一段对性能影响最大呢？



流水线中，上一部分的输出就是下一步的输入。我们可以手动设置100%正确的输出结果，然后

[目录开关]

- 1 滑动窗口 (Sliding windows)
- 2 获得大量人工数据
- 3 上限分析 (Geiling Analysis)

72%，如果令文字侦测部分的结果100%正确，总体正确率提升到89%。这投入时间精力来提高我们的文字侦测部分。

我们再让字符切分输出的结果100%正确，系统的总体效果只提升了1%，这意味着，字符切分部分可能已经足够好了。

Component	Accuracy
Overall system	72% ← ↓ 17%
→ Text detection	89% ← ↓ 1%
Character segmentation	<u>90%</u> ← ↓ 10%
Character recognition	100% ←

点赞 1 赏 分享

如果文章对你有帮助，欢迎点赞或打赏（金额不限）。你的打赏将全部用于支付网站服务器费用和提高网站文章质量，谢谢支持。

版权声明:

本文由 yalewoo 原创，商业转载请联系作者获得授权。
非商业转载请注明作者 yalewoo 或 雅乐网，并附带本文链接：
http://www.yalewoo.com/andrew_ng_machine_learning_notes_11_photo_ocr.html

上一篇：
[Andrew Ng机器学习课程笔记10——大规模机器学习](#)

下一篇：
[Social Network Analysis笔记1：介绍](#)

与 Ng笔记, 机器学习 相关的文章

- [Andrew Ng机器学习课程笔记10——大规模机器学习](#)
- [Andrew Ng机器学习课程笔记9——异常检测和推荐系统](#)
- [Andrew Ng机器学习课程笔记目录](#)
- [Andrew Ng机器学习课程笔记8——聚类和降维](#)
- [Andrew Ng机器学习课程笔记7——支持向量机（SVM）](#)
- [Andrew Ng机器学习课程笔记6——机器学习应用建议和系统设计](#)
- [Andrew Ng机器学习课程笔记5——神经网络：学习](#)
- [Andrew Ng机器学习课程笔记4——神经网络：表达](#)

文章《Andrew Ng机器学习课程笔记11——图像文字识别》共有1条评论：

我要评论

[目录开关]