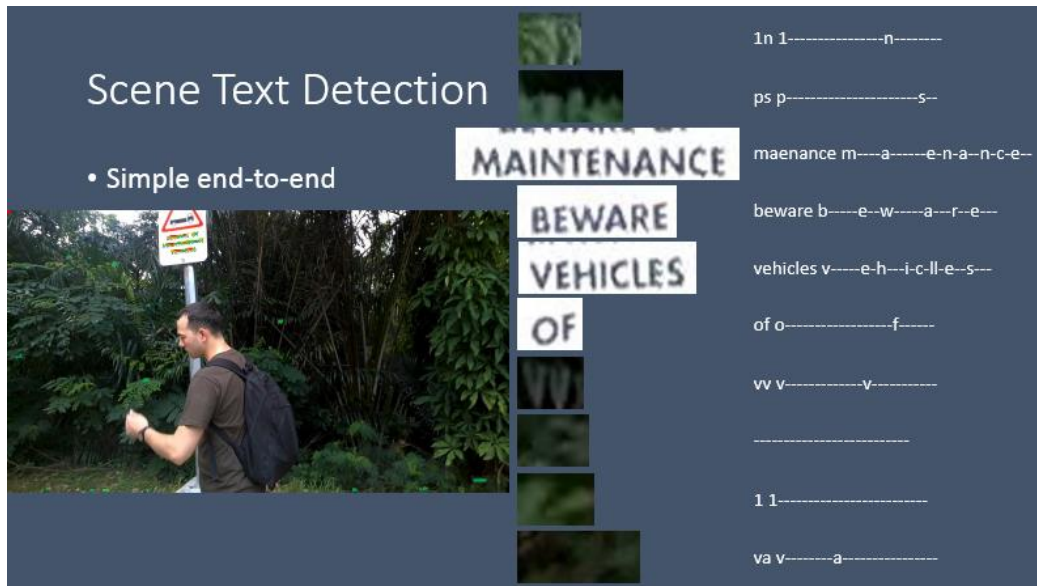


1. 场景文本检测

- a) 上次报告中提到的多行文本会有粘连的问题已经能够处理的不错了。处理的方法就是在训练中 Drop 掉置信度很高的负样本。



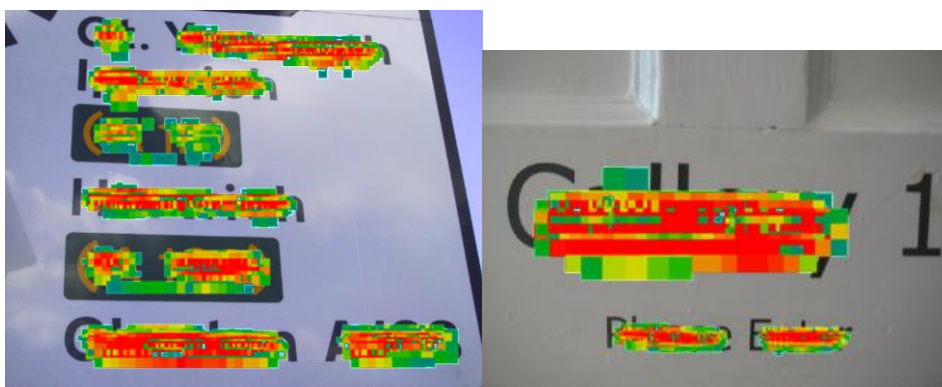
- b) 由于 Challenge4 中有垂直文本，所以我把训练样本做了旋转的扰动，结果发现会出现粘连的现象，分析原因是因为分类器会认为这些多行水平文本在竖直方向上是一个文本行，目前不知道该怎么解决这个问题。
- c) 针对目前检测结果中存在 False Positive 的问题，我想用 RNN 识别的结果来对这些样本做剔除。做法是在 Torch7 上，利用 Bai Xiang 提供的识别框架对检测提取的 Proposal 进行识别。结果发现对噪声并不能特别好的剔除，不过从中我也想到了一个方法，就是把串识别的 RNN 部分作为一个 Task 进行训练，识别做的好的话可以做到 End to End Training and Testing，识别做的不好也可以拿来校正检测的结果。下图从左到右依次是 带有 Heatmap 的原图，利用 Heatmap 截取出来的 Proposal，RNN 识别的结果，RNN 中 CTC 对齐前的结果



- d) 利用 Heatmap 对倾斜以及弯曲的文本行做校正。跟冯伟师弟一起尝试用 Heatmap 来对文本行做校正，把倾斜以及弯曲的文本行校正为水平的，效果如下图。这个方法的特点是完全基于 Heatmap，不利用任何原图的信息，而且没有训练参数，纯靠规则。个人觉得可以为之后的识别框架提供支撑。



- e) 在 Challenge2 上简单地测试了一下，对于整词比较长的 Case 能够检测很好，但是假如是孤立的一个字母就很难检测，主要是跟训练样本有关，因为实验是基于 Challenge4 做的，Challenge4 里面对于特别短的词是不要求必须检测到的。



2. 公式识别

- 将符号集合扩展到了 128 类，除去根号需要做一下预处理与后处理，其他新加的符号跟之前比并没有增加太多处理难度。
- 复现了之前公式识别的框架，跟之前相比，收敛速度快了很多，应该是因为在 Multi-task 阶段对结构做了些调整的缘故。

3. 之后的工作

- 在 Challenge4 跟 Challenge2 上给一个初步的结果
- 将 STN 加入到公式识别的框架：假设公式全图是做了一个统一的仿射变换，然后利用 STN 对整图做逆变换
- 准备组会报告，报告内容是有关 YOLO (You only look once) 的两篇文章