数据集分两部分

自然数据集

英语和坎那达语 还有手写数字和其他的字符数据集 都是来自自然场景下

英语：作者按大小写和字符分为62个类。坎那达语提供了49个基本类，不区分大小写。

我们拍摄了一组1922照片，大部分是布告板，广告牌和广告，但我们也包括一些超市里的产品图片商店。这里使用了多边形框和矩形框作为标记，发现多边形框几乎没有优势。英语数据集有12503个字符，其中4798幅图像效果较差由于过度咬合，低分辨率或噪音。我们使用了剩下的7705字符图像。类似地，对于Kannada，总共有4194个字符被提取出来的只有3345个使用。

手写和打印数据集

使用平板电脑捕获手印数据集（Hnd），笔的厚度设置为与手绘信息中的平均厚度相匹配。对于英语，55名志愿者共生成了3410个字符。 对于卡纳达语，25名志愿者总共创作了16425个字符。一些示例图像如图7所示。字体数据集仅用于英文字符。 我们尝试了4种风格的254种不同的字体（普通，粗体，斜体和粗体+斜体），共生成62992个字符。 这个数据集将被称为Fnt数据集。

Png文件

64类（0-9，A-Z，a-z）

从自然图像获得7705个字符

3410手绘字符使用平板电脑

62992来自计算机字体的合成字符

标注：在我们的实验中用于训练和测试的文件列表（以MatLab数据文件格式“.MAT”）。 请使用这些拆分，以便与上面的论文中公布的结果进行公平比较。

每个文件都有一个包含这些元素的数据结构“列表”：

ALLlabels：每个样品的类别标签

ALLnames：每个样本的图像的子目录和名称

类标签：该数据集中的标签（类）集合，用数字编码，例如 10 = A，11 = B，...，64 = z

classnames：存储每个类的样本的目录的边界

NUMclasses：此数据集中类的总数

TRNind：训练样本的索引。 如果使用20个分裂，则这是N\_train\_samples X 20的矩阵

TSTind：测试样本的索引。 如果使用20个分裂，则这是N\_test\_samples X 20的矩阵

VALind：验证样本的索引。 如果使用20个分裂，则这是N\_validation\_samples X 20的矩阵

TXNind：文本样本的索引，即用于用视觉词袋方法构建词汇表的样本。 如果使用20个分割，则这是N\_texton\_samples X 20的矩阵

作者是测试和训练样本数目保持相等，然后进行抽样每类15个样本进行训练和测试

自己理解：这个数据集中图片是来自各个自然场景，图像的长，宽分布较广，应该是同一块场景文本中的一部分，是否可以以在这个数据集让后面图像检测网络学到自然场景下文本的底层特征，然后在加一些层识别整块的样本，也就是在目标区域候选层后面再加一些层进行一个整块文本的检测，在训练时加可视化观察前面层是否学到了单个字符的特征，后面再说，还有这类样本过多会不会影响候选区域层提出感兴趣区域。