文字檢測

傳統打印/掃描樣本:

排列整齊，字體規範清楚，圖片背景單一，容易分割。

場景文本:

多變文本字體，高度複雜訊息，外界干擾嚴重(強光線.陰影.遮擋)。

場景文本辨識包含了車牌識別，智能閱讀，身分證辨識等，若要進行文字檢測首先就要提到目標檢測方法。

目標檢測透過強大的深度學習能力，用來檢測圖像目標位置以及目標對象的類型，而目標檢測可分為基於實例的目標檢測(Instance-base )和基於類別的目標檢測(Category-based)。

前者是找尋某一特定物體的實例，後者為找尋圖像中所有同一類別的技術。

目前可參考論文有，ResNet(2016，He et al.)，DenseNet(2016，Huang et al.)，Mask RCNN(2016，He et al. )，Fast RCNN(2015，Ren et al.)

目標檢測目前的技術主要分成三類:

1)傳統非深度學習方法，基於傳統特徵(Hog，Sift)的方法又被稱作Region Proposal Based方法，此方法又可分為三步驟，即區域選擇，特徵提取，分類回歸。

2)Two-stage方法，該方法包含一個預處理階段，負責生成一個標框區域，並提取CNN特徵，然後再進入第二階段進行分類並修正位置。

3)One-stage方法，該方法進一步簡化Two-stage方法，略過了生成標框區域，又被稱作Region Proposal Free方法。

而目標檢測的相關術語又分作:

1. IOU(交並比)(Intersection Over Union):

此值用來衡量生成的預測檢測框和標註框之間的重疊程度，IOU= DectionResultGroundTruth/DectionResultGroundTruth

1. NMS(非極大值抑制)(Non-Maximum Suppression):

由於標框算法會標出一些錯誤或不準確的樣本，所以需要一個有效的算法來消除多餘的檢測框架，此算法的本質是搜索局部最大值，抑制非極大值元素，利用置信度分數(最高值以及大於閥值的值)來刪除檢測框。

p.s.置信度分數在不同的論文中定義會有所不同，一般會用到的是分類置信度，分類置信度confidence是介於0和1(或100%)之間的數字，描述模型預測邊界框包含某類別目標的機率，一般多分類最後一層輸出會接softmax函數，會輸出n個前景和1個背景的概率分數，但YOLOv1的定義為Pr(classi)\*IOU(truth pred)，還有一些論文預測了預測框與GT的IOU，此IOU又被稱作是定位置信度，此置信度通常是用在CNN上使用。

1. 目標檢測的評價指標主要是Average Precision(平均精度)，主要基於準確率(Precision)和召回率(Recall)計算出來的，Precision=TP/TP+FP，

Recall=TP/TP+FN。

FI度量=Precision\*Recall\*2 / Precision + Recall

MAP(加權平均精度均值)

兩者都是AP的改良版，可以更好地描述試驗方法的解釋力。

TP=True Positive TN=True Negative

FP=False Positive FN=False Negative

1. 選擇性搜索(Selective Search)是一種流行的生成候選框方法，其步驟為將圖像分割成小區域(通常為1000個到2000個)，來獲取原始分割區域，並計算每個相鄰區域的相似度，按照合併規則合併區域重複上述步驟，而考量的因素在論文中主要有四個參數分別為:

1.顏色相似度

2.紋理相似度(高斯分布做梯度統計)

3.尺寸相似度

4.空間交疊相似度

1. 之後會利用特徵提取(Haar，LBP，HOG)加上分類回歸的方式(HOG + SVM)(Deformable Part Model)(Bounding-Box Regression)來增加框架準確性。

而其他經典的區域選擇算法有Objectness，Bing，Edge Boxes，MCG，Rigor。

相較於傳統學習方法，Two-Stage方法則是(Region Proposal + CNN)，其算法流程為輸入測試圖像 -> 並利用Selective Search提取2000個左右的Region Proposals -> 然後將每個Region Proposals縮放成227\*227的像素並輸入CNN將CNN的fc7的輸出做為CNN的特徵 -> 並將CNN特徵輸入SVM進行分類 -> 再進行細部微調並修正候選框的位置。

但因其缺點是訓練耗時且速度慢(每一張圖如果有2000個RP那就得進行2000次運算)並且所有圖片需同尺寸，而SPPNet很好地解決了尺寸必須完全相同的問題，且也可先提取特徵避免圖片重複計算量大的問題。

p.s.SVM的算法改進是可以在這邊去進行微調與改善的

<http://www.360doc.com/content/10/0519/22/1472642_28469588.shtml>

後來Fast R-CNN提出了ROI Pooling的操作以進行R-CNN的實時檢測，其瓶頸主要來自於Selective Search的計算量問題上。

改良版的Faster R-CNN的主要組件主要為

1) CNN-Layers(運用Conv + ReLU + Pooling)提取特徵

2) RPN (Region\_Proposal\_Network)此層生成Region Proposal並且利用Bounding Box Regression修正Anchor獲取較好的Proposal

3)ROI Pooling再將目標特徵細緻化

4)Classification利用Proposal Feature Map計算Proposal類別並再一次利用Bounding Box Regression檢測框的最終位置。

RFCN論文中提到希望在分類時網路神經層足夠深則具有更好的平移不變性，但在檢測時希望網路不要太深，否則就喪失了平移可變性。

Mask R-CNN 則可同時做目標檢測(Object Betection)和實例檢測

(Instance Segmentation)，在提升精度的同時並無增加檢測時間。