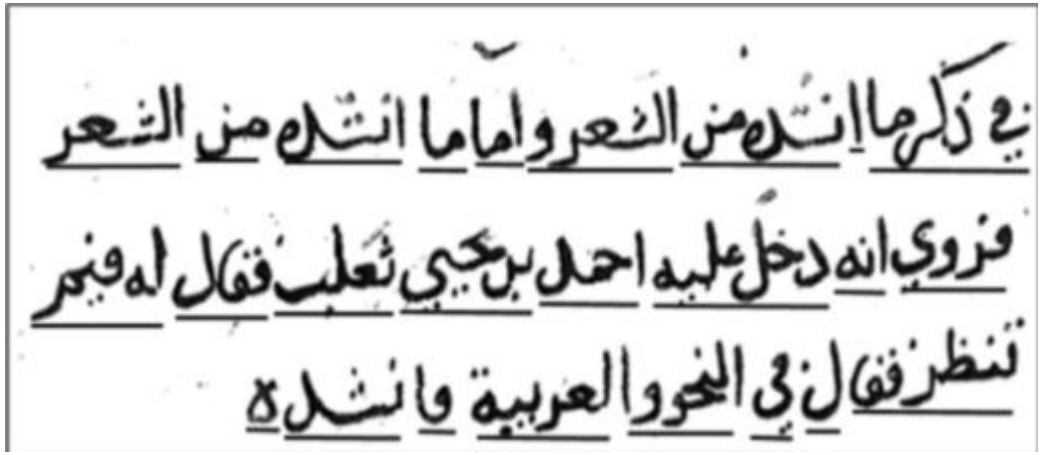
文档布局分析

胡子凌

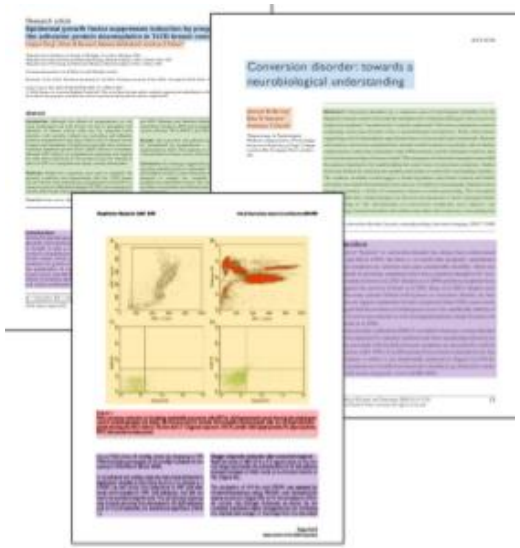
Document layout analysis

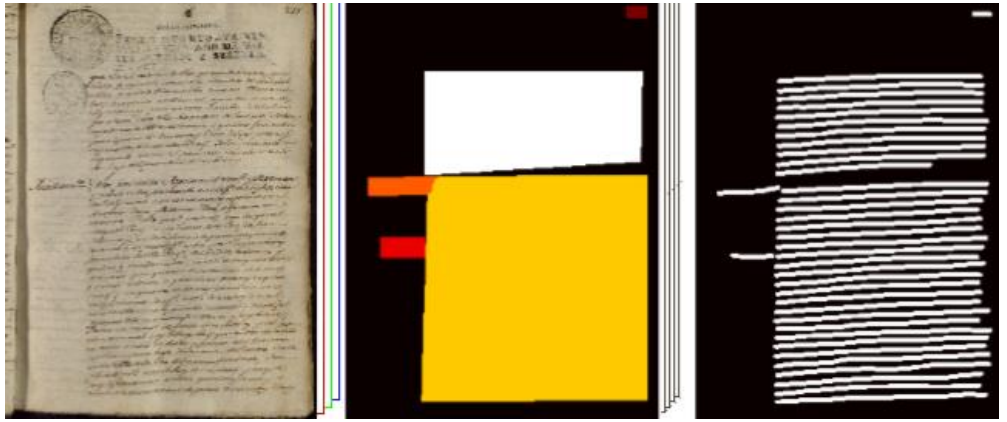
- 任务性质：文档布局分析是文档理解系统的中间步骤
- 任务应用：应用于文档检索、内容分类、文本识别等领域
- 任务场景：打印文档、手写类文档
- 任务方法：基于目标检测以及基于图像分割



DLA的主要问题

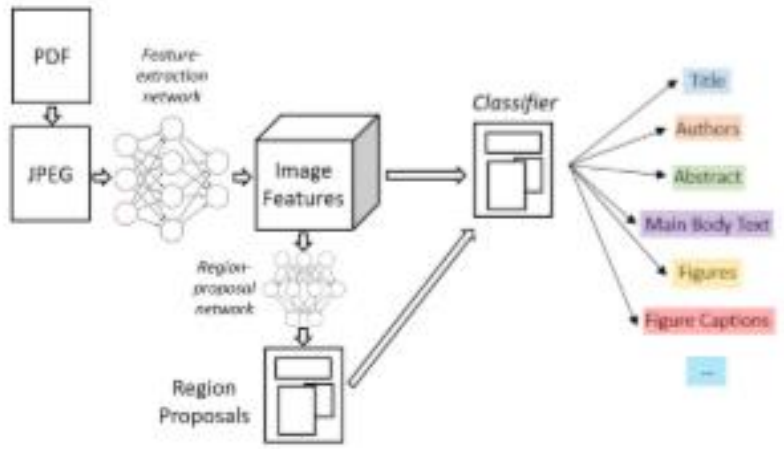
- 任务场景不统一，少有统一的benchmark dataset
- 任务偏工程化，不同方法间使用的应用场景差异大
- 可用数据少,且标注困难
- 手写文档图片质量普遍较差



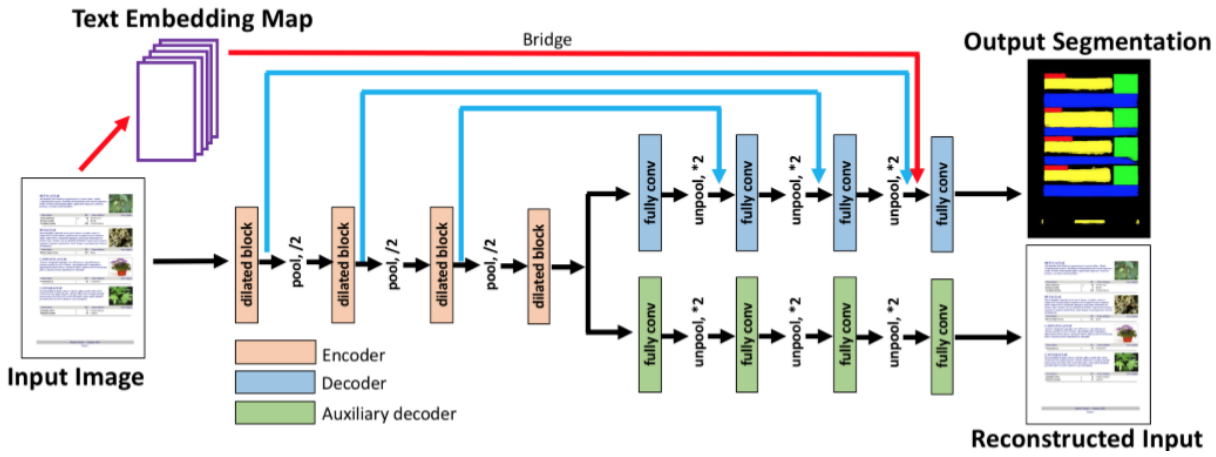


DLA的主要方法

- 基于目标检测的方法

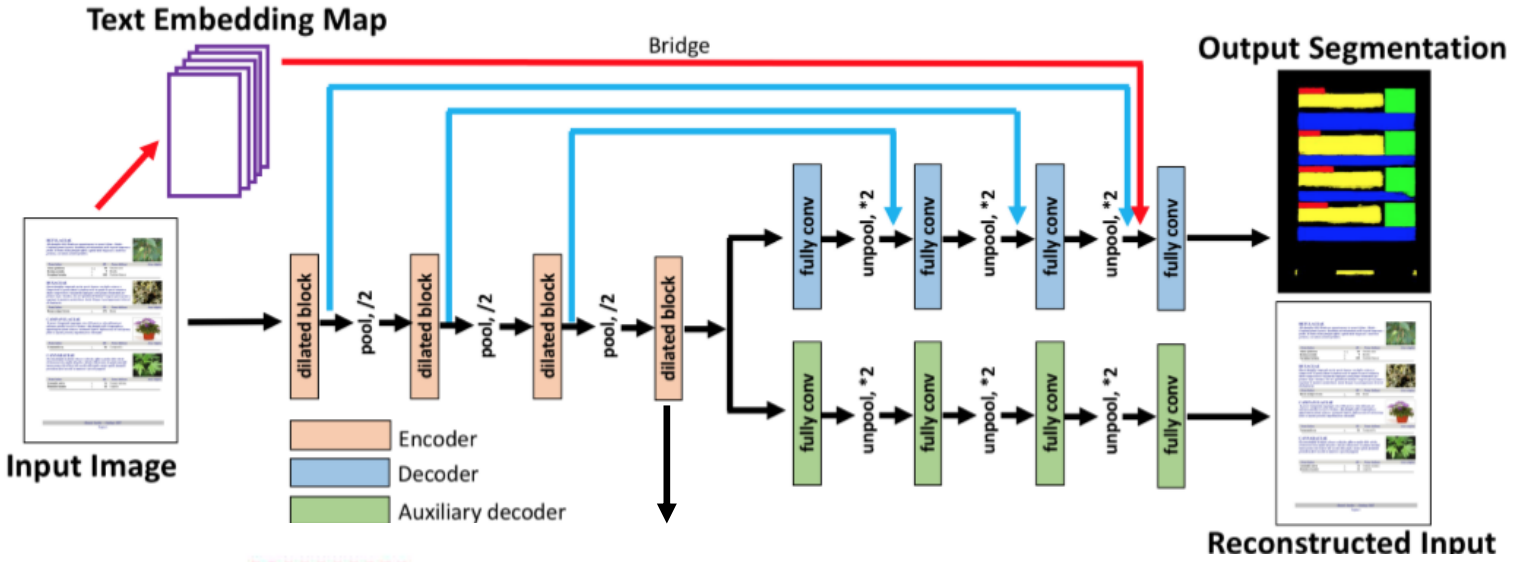


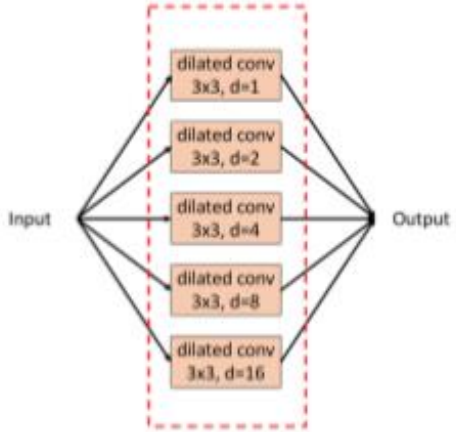
- 基于图像分割的方法

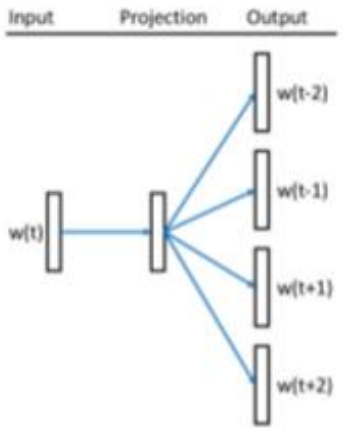


图像分割方法

MFCNN—CVPR17







- 通过加入Embedding,实现了fine grain的图像分割
- 提出使用无监督任务进行辅助
- 提出了新的数据集与数据生成方法

Learning to Extract Semantic Structure from Documents Using Multimodal Fully Convolutional Neural Networks

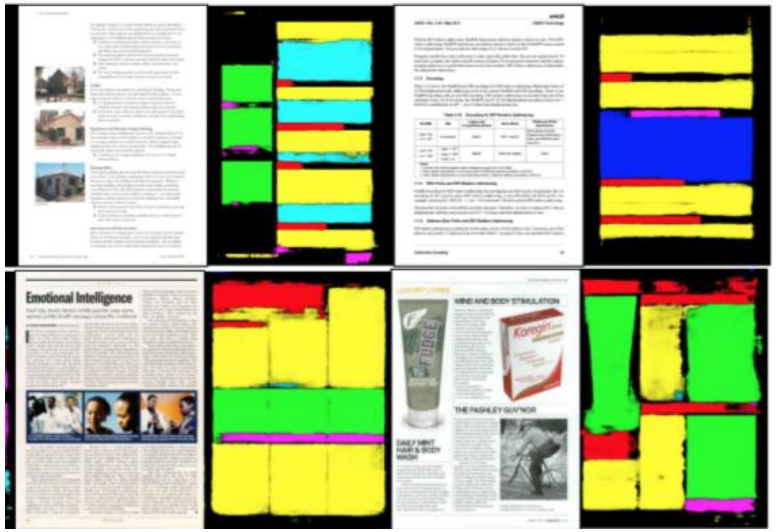
图像分割方法

MFCNN—CVPR17

base	dataset	text	bkg	figure	table	section	caption	list	para.	mean
Model5	D	none	**84.6**	83.3	**79.4**	58.3	61.0	66.7	77.1	73.0
Model5	D	extract	83.9	**83.7**	79.7	**59.4**	**61.1**	**68.4**	**79.3**	**73.3**
Model5	S	none	87.7	83.1	84.3	70.8	70.9	82.3	83.1	79.6
Model5	S	extract	88.8	85.4	86.6	73.1	71.2	83.6	87.2	82.2
Model5	S	real	**91.2**	**90.3**	**89.0**	**78.4**	**75.3**	**87.5**	**89.6**	**86.0**

Methods	non-text	text
Leptonica [9]	84.7	86.8
Bukhari et al. [11]	90.6	90.3
Ours (binary)	**94.5**	**91.0**
Methods	**figure**	**text**
Fernandez et al. [18]	70.1	85.8
Ours (binary)	**77.1**	**91.0**

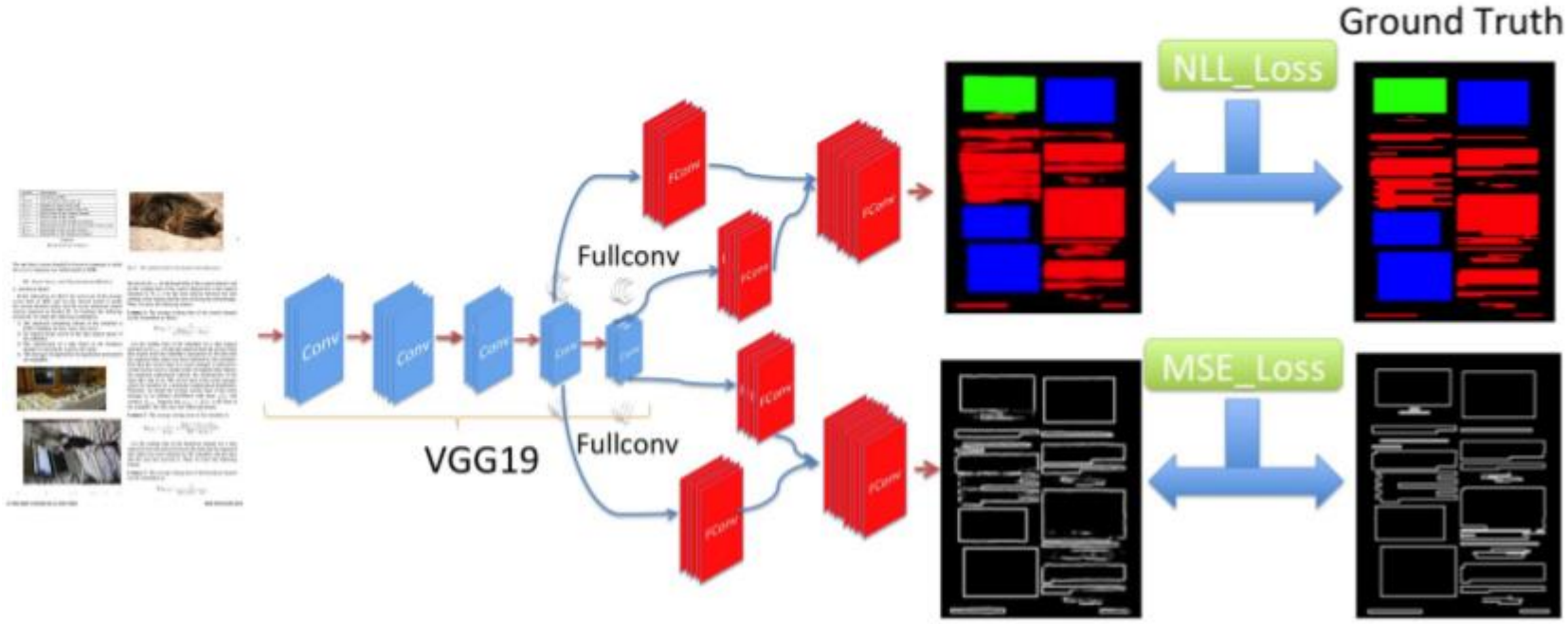
Table 4: IoU scores (%) for page segmentation on the ICDAR2015 dataset. For comparison purpose, only IoU scores for non-text, text and figure are shown. However our model can make fine-grained predictions as well.



Learning to Extract Semantic Structure from Documents Using Multimodal Fully Convolutional Neural Networks

图像分割方法

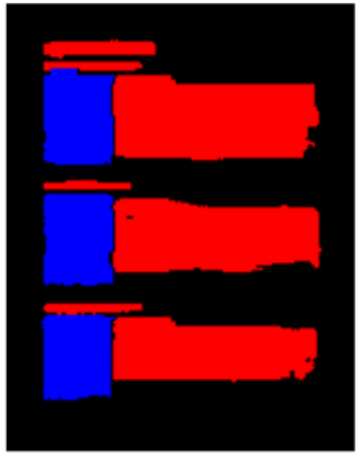
Multi-scale Multi-task FCN for Semantic Page Segmentation and Table Detection —ICDAR17

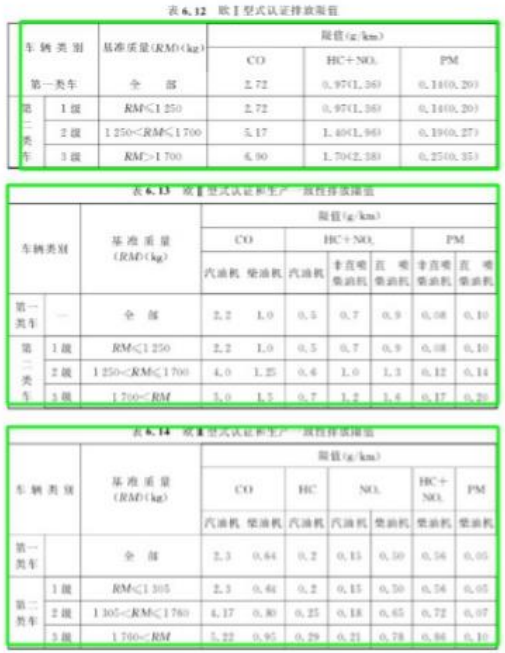


- 使用了不同层多尺度的视觉特征
- 加入了轮廓检测任务作为辅助
- 能同时实现页面分割和表格检测

图像分割方法

Multi-scale Multi-task FCN for Semantic Page Segmentation and Table Detection —ICDAR17

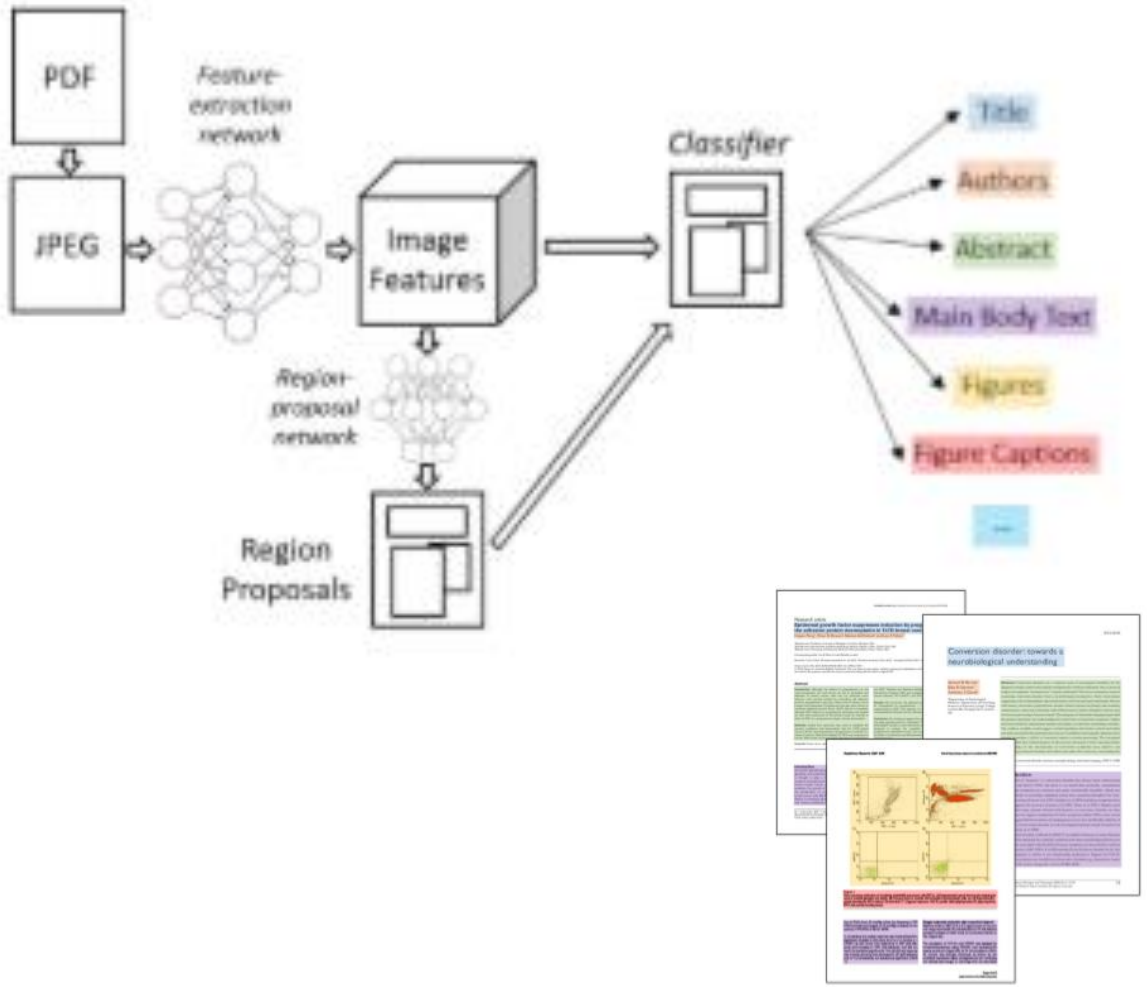




Chinese Documents			English Documents		
method	precision	recall	method	precision	recall
Liu [25]	0.44	0.28	Liu [25]	0.48	0.35
Yildiz [29]	0.35	0.34	Yildiz [29]	0.35	0.38
Fang [30]	0.89	0.80	Fang [30]	0.58	0.49
Our(0.9 IOU)	0.493	0.491	Our(0.9 IOU)	0.47	0.45
Our(0.8 IOU)	0.77	0.761	Our(0.8 IOU)	0.753	0.70

目标方法

Visual Detection with Context for Document Layout Analysis—IJCNlP19

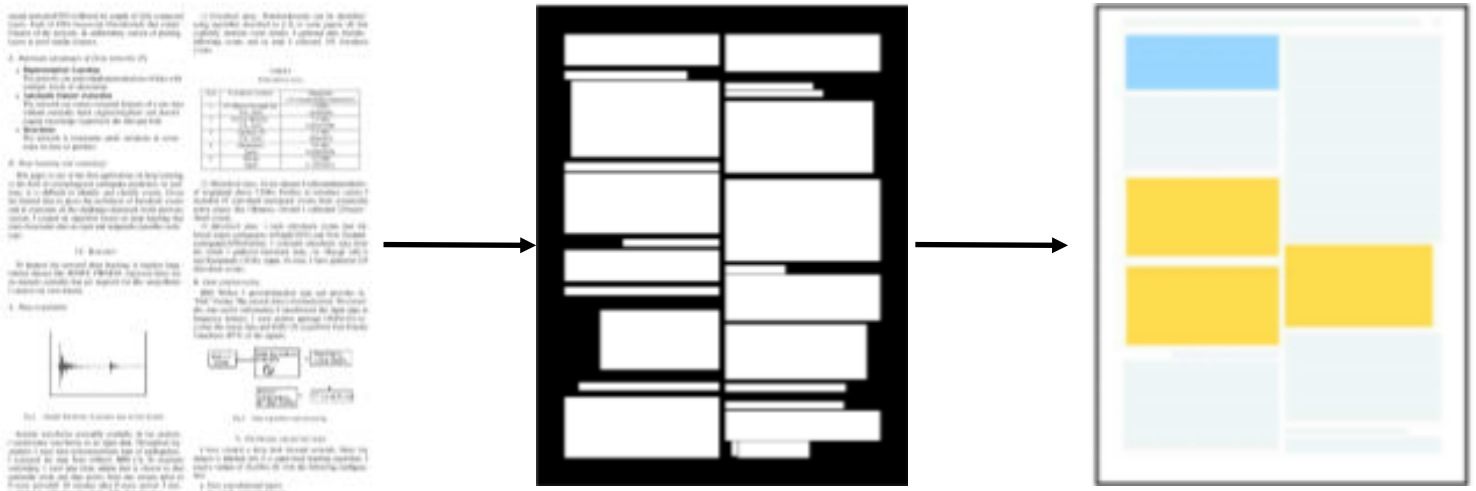


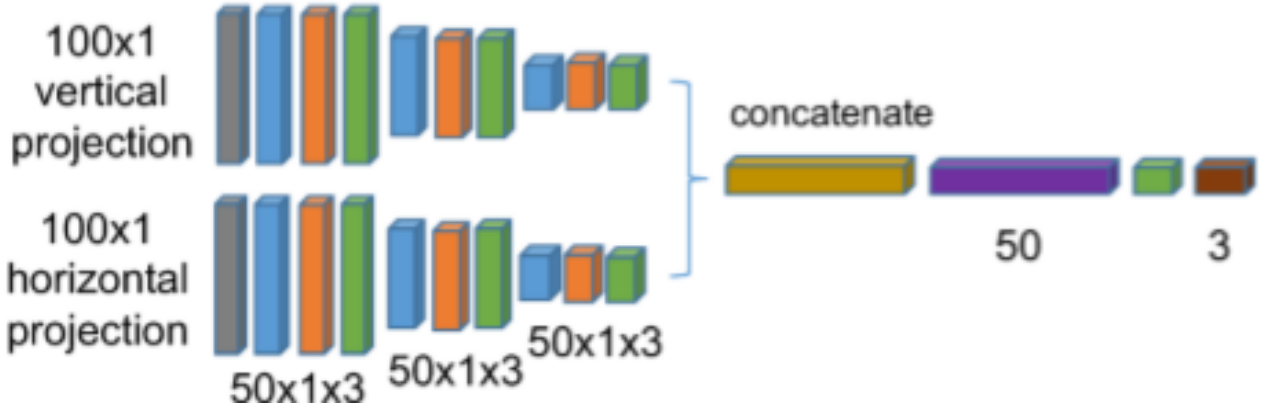
- 加入了Context Feature Embedding
- 实现了fine grain的检测分类
- 提出了新的数据集

Context Feature Embedding：
将候选区域框的位置，大小，文档页数编码作为context feature嵌入到视觉特征中

目标方法

Fast CNN-based document layout analysis—CVPR17 workshop



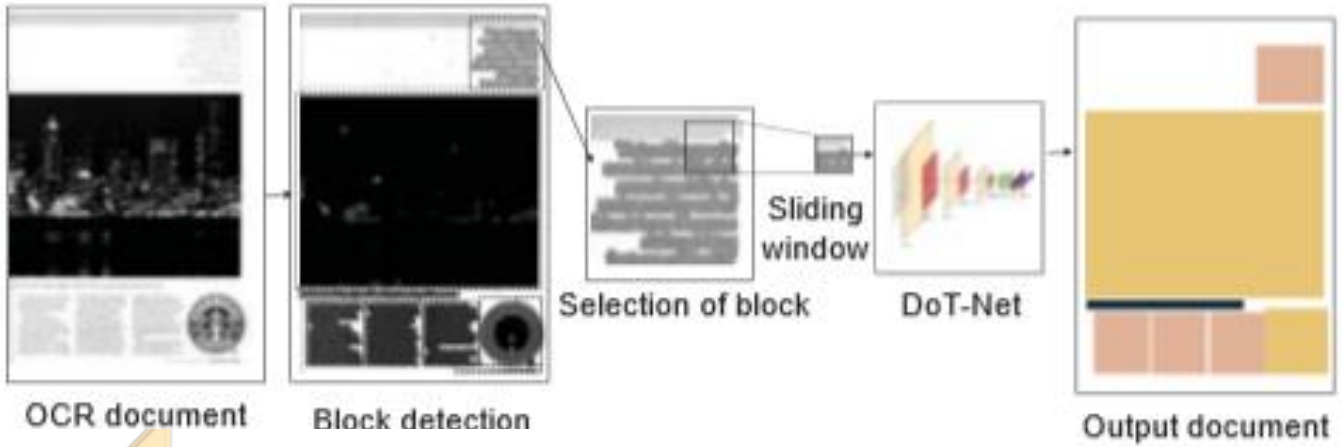


- 提出了1D利用卷积快速进行分类
- 提出了新的数据集

Fast CNN-based document layout analysis

目标方法

DoT-Net—ICDAR19



- 实现了多分类
- 用dilated卷积替代了所有的普通卷积

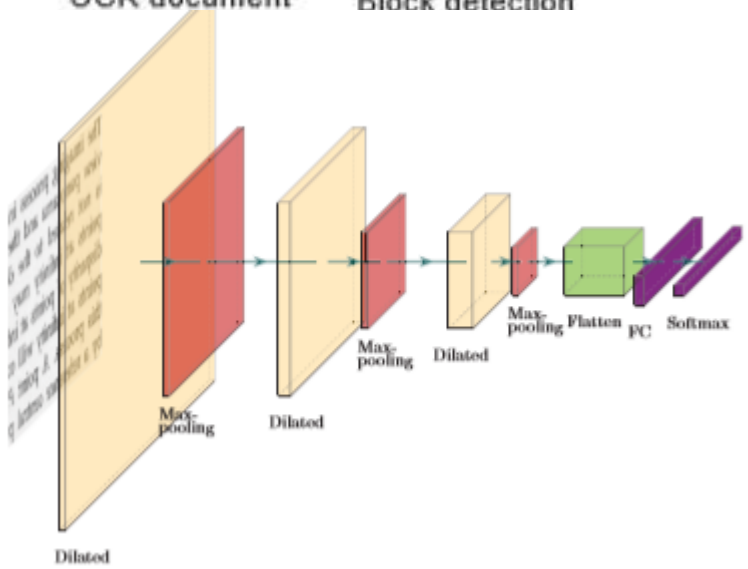


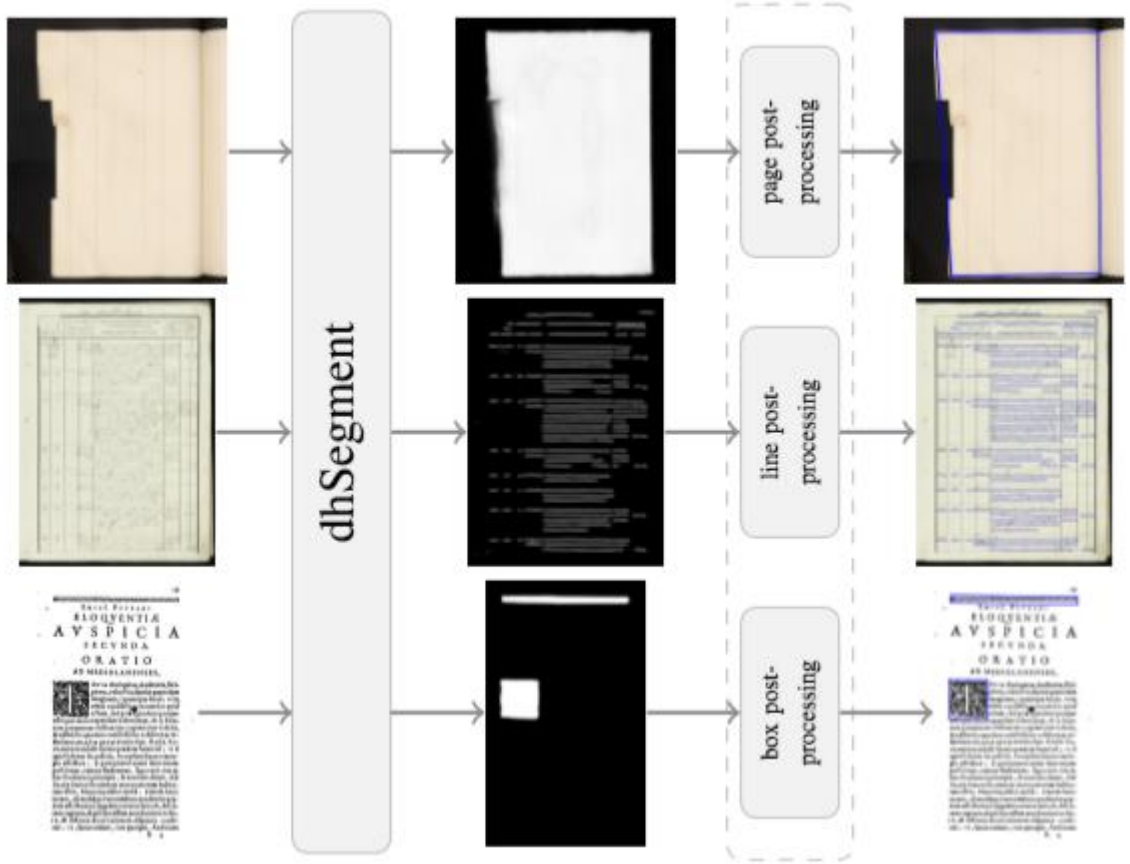
TABLE V: Performance with block images in multiclass classification

	Accuracy	F1 score	AUC
F1DCNN	0.842 (0.019)	0.831 (0.013)	0.874 (0.023)
FFN	0.532 (0.035)	0.451 (0.027)	0.593 (0.041)
CNN	0.681 (0.024)	0.643 (0.013)	0.716 (0.029)
GSVM	0.449 (0.004)	0.394 (0.007)	0.512 (0.009
HOGMLP	0.491 (0.019)	0.371 (0.005)	0.532 (0.019)
DoT-Net	**0.941 (0.021)**	**0.929 (0.011)**	**0.952 (0.017)**

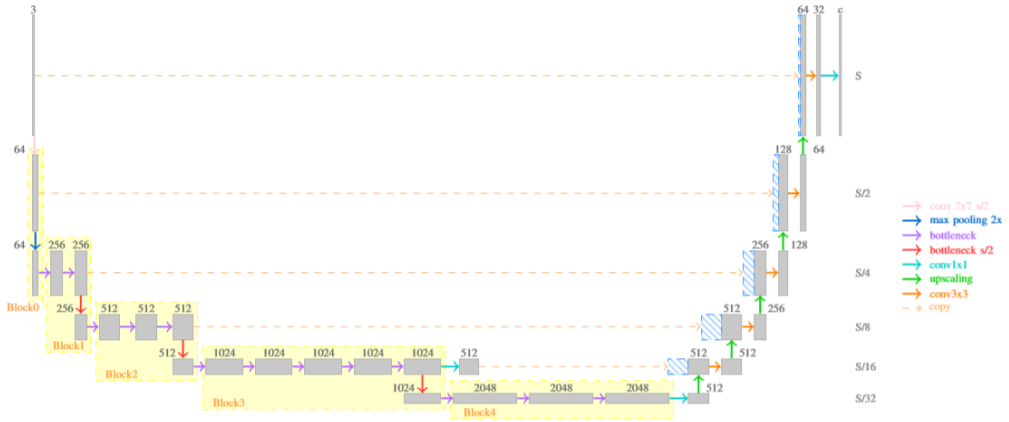
DoT-Net: Document Layout Classification Using Texture-based CNN

图像分割方法

dhSegment— ICFHR18



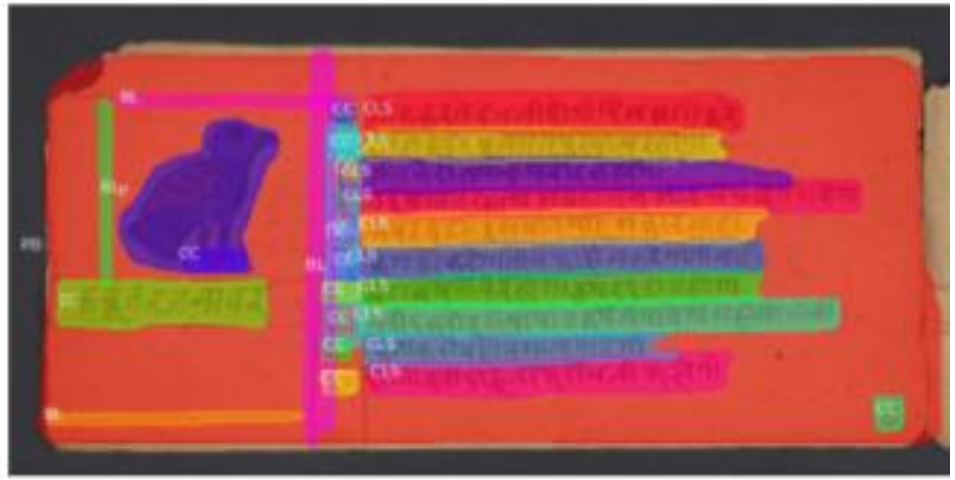
- 使用一个框架完成多个layout任务



Multi-scale Multi-task FCN for Semantic Page Segmentation and Table Detection

图像分割方法

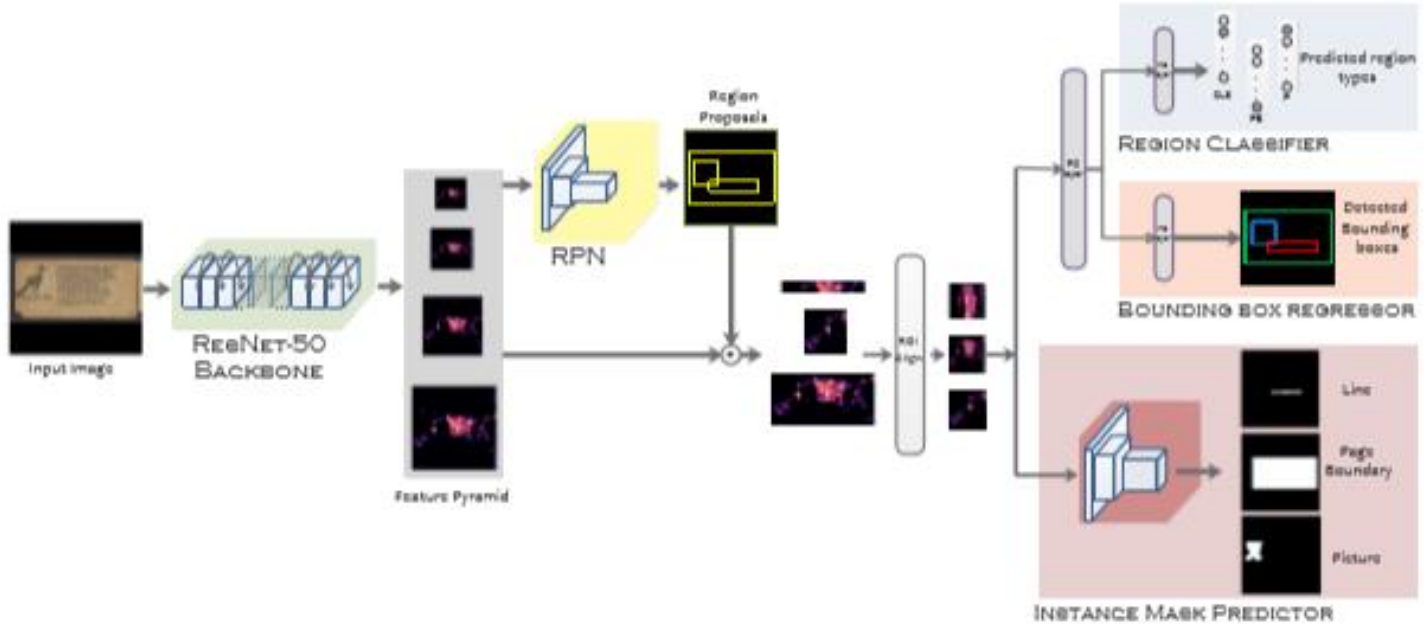
Indiscapes— ICDAR19



Indiscapes: Instance Segmentation Networks for Layout Parsing of Historical Indic Manuscripts

图像分割方法

Indiscapes— ICDAR19

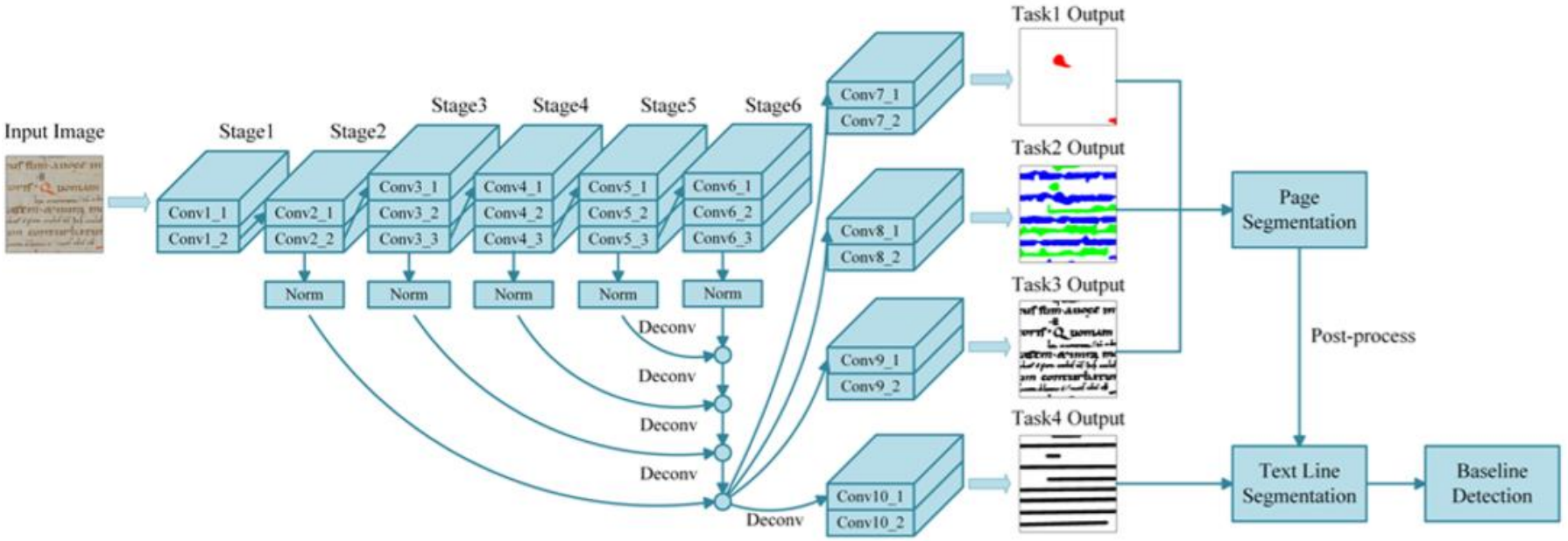


- 完成了一个可视化标注工具，并提出了了一个数据集
- 使用了一个基于深度学习的实例分割方法

Indiscapes: Instance Segmentation Networks for Layout Parsing of Historical Indic Manuscripts

图像分割方法

Multi-task Layout Analysis for Historical Handwritten Documents Using Fully Convolutional Networks— IJCAI18



图像分割方法

Multi-task Layout Analysis for Historical Handwritten Documents Using Fully Convolutional Networks— IJCAI18

(a) (b) (c) (d)