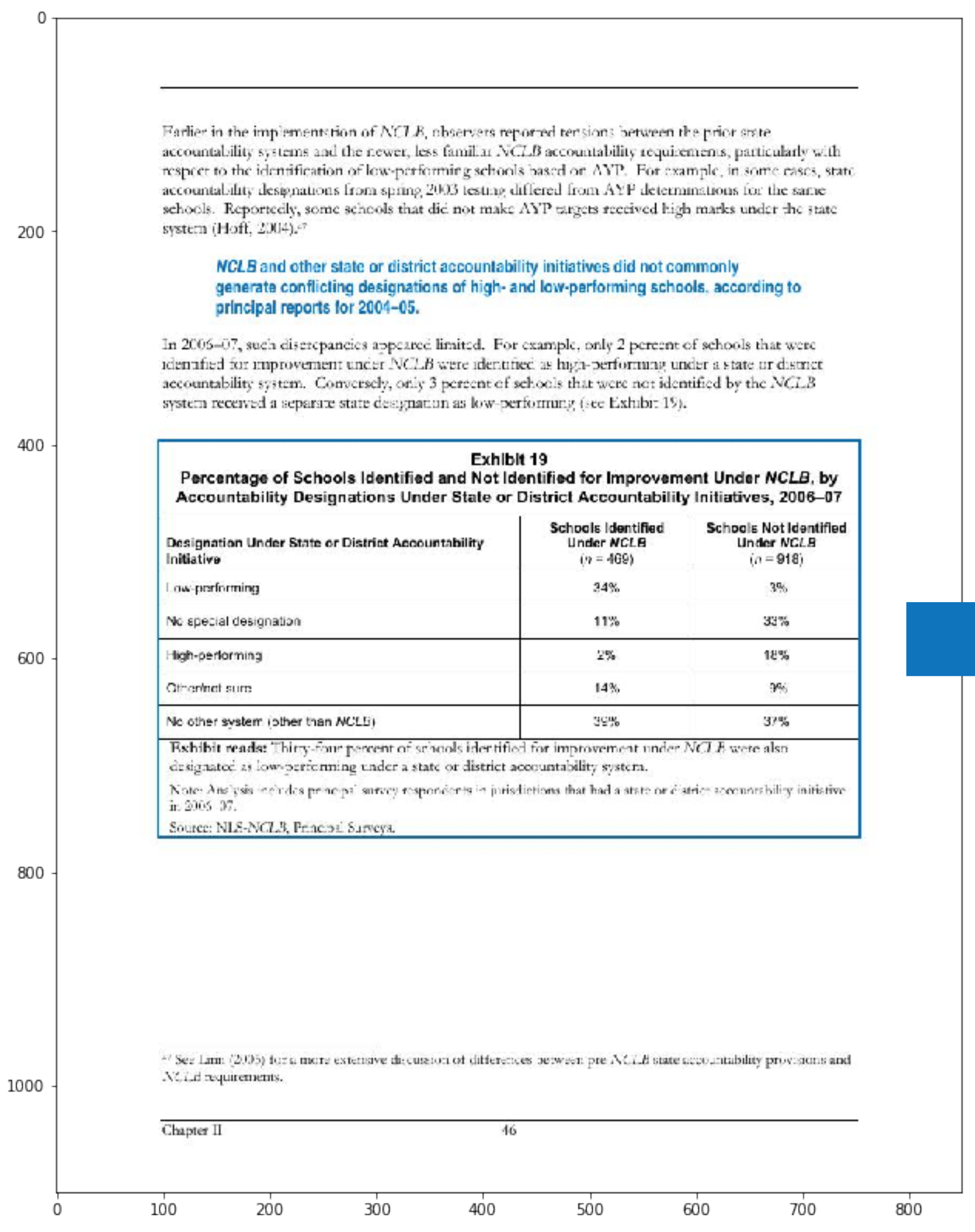
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

School of Computer Science and Engineering
College of Engineering

A Novel Pipeline for Table Extraction using Deep Learning

Student: Lee Seng Cheong **Supervisor: Dr Loke Yuan Ren**

Tables and tabular formats are structured manners of presenting information. Much of the information within tables cannot be recovered without preserving the structure of the table, and manual extraction of tables is tedious. However, automated systems for table extracting must deal with tables from different domains, font, styles, structures and mediums with such systems mainly oriented towards digital documents only. We propose an image-based deep learning pipeline for extracting tables that can accommodate the variety in formats and styles from both images and digital documents.


0
200
400
600
800
1000
100
200
300
400
500
600
700
800
Exhibit 19
Percentage of Schools Identified and Not Identified for Improvement Under NCLB, by Accountability Designations Under State or District Accountability Initiatives, 2006–07

Input document as image

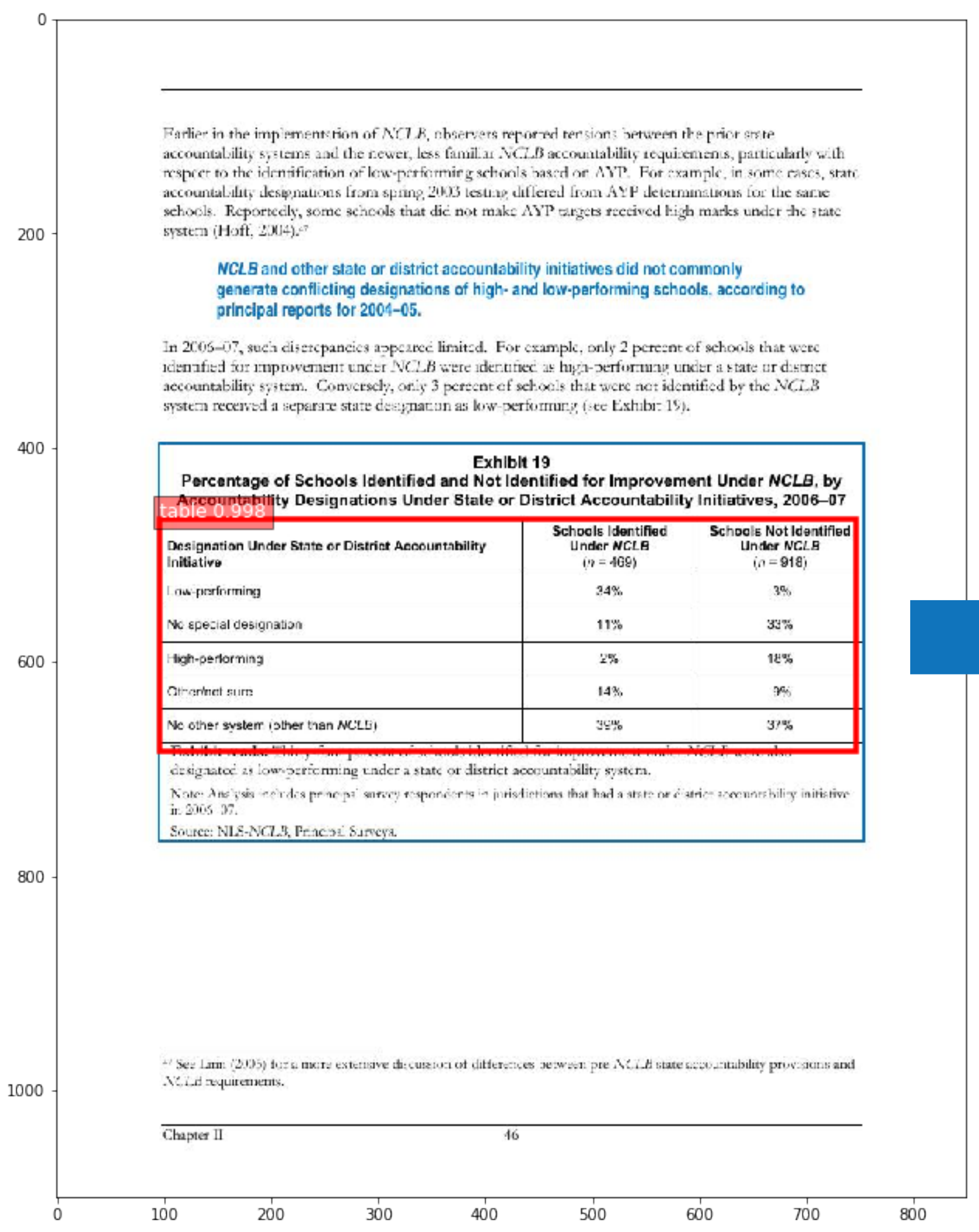
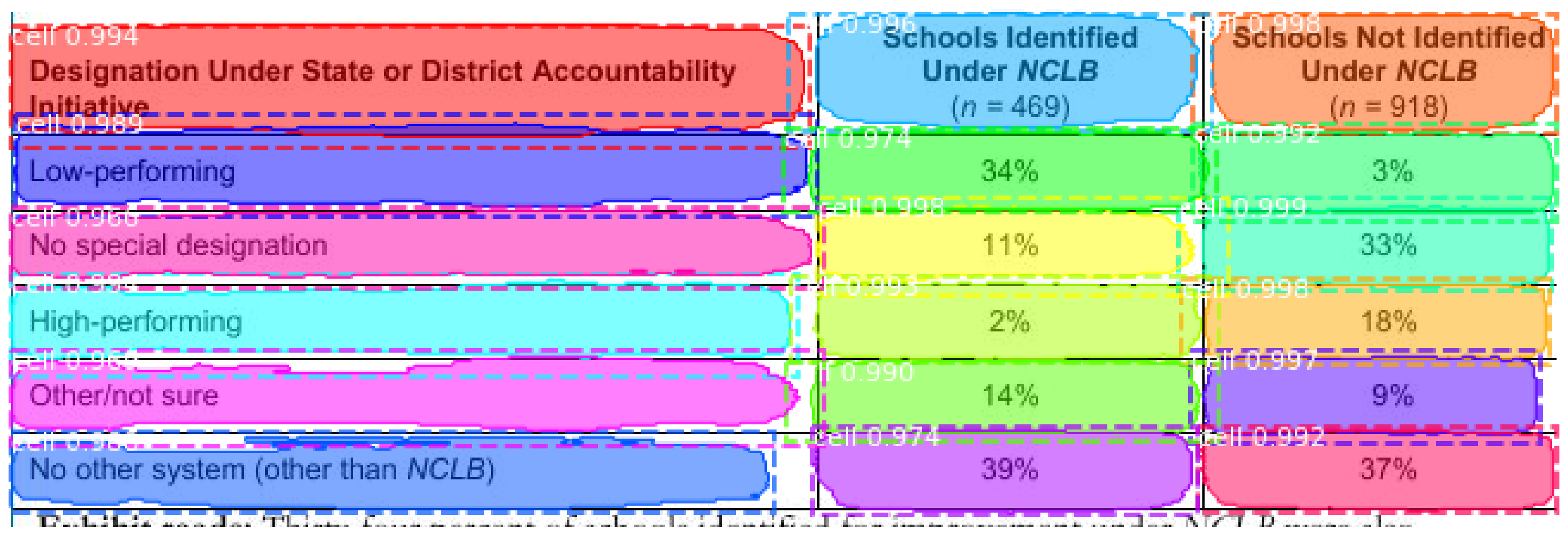
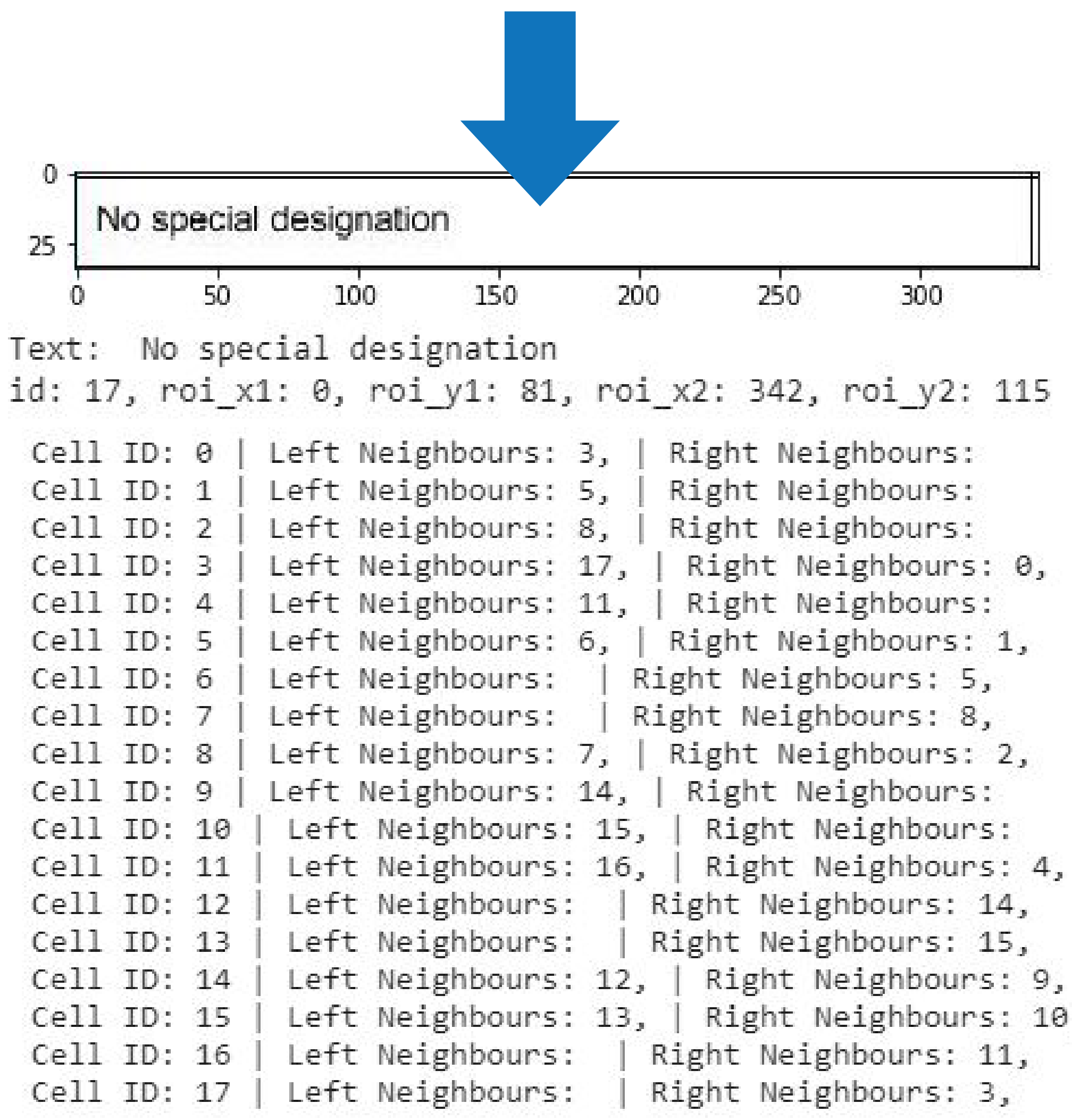

0
200
400
600
800
1000
100
200
300
400
500
600
700
800
table 0.998
Exhibit 19
Percentage of Schools Identified and Not Identified for Improvement Under NCLB, by Accountability Designations Under State or District Accountability Initiatives, 2006–07

Table detection using object detection approaches



Designation Under State or District Accountability Initiative	Schools Identified Under *NCLB* (*n* = 469)	Schools Not Identified Under *NCLB* (*n* = 918)
Low-performing	34%	3%
No special designation	11%	33%
High-performing	2%	18%
Other/not sure	14%	9%
No other system (other than *NCLB*)	39%	37%

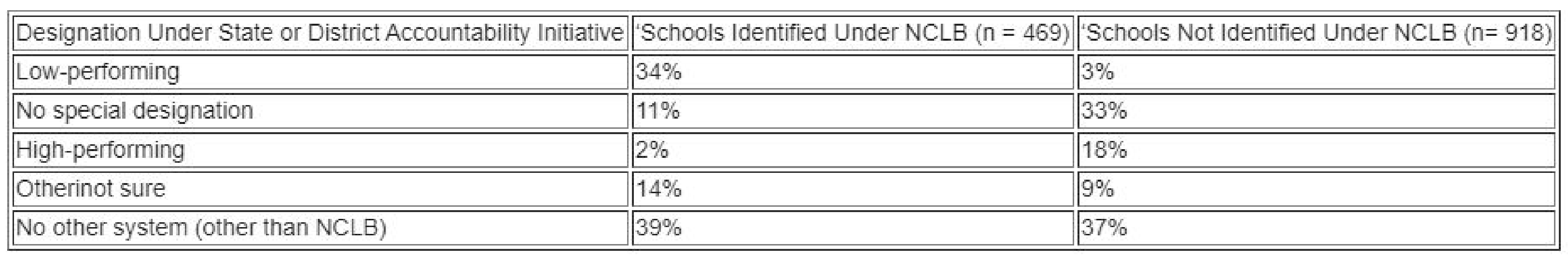
Table cell segmentation using instance segmentation approaches



```
Text:  No special designation
id: 17, roi_x1: 0, roi_y1: 81, roi_x2: 342, roi_y2: 115

 Cell ID: 0 | Left Neighbours: 3, | Right Neighbours:
 Cell ID: 1 | Left Neighbours: 5, | Right Neighbours:
 Cell ID: 2 | Left Neighbours: 8, | Right Neighbours:
 Cell ID: 3 | Left Neighbours: 17, | Right Neighbours: 0,
 Cell ID: 4 | Left Neighbours: 11, | Right Neighbours:
 Cell ID: 5 | Left Neighbours: 6, | Right Neighbours: 1,
 Cell ID: 6 | Left Neighbours:  | Right Neighbours: 5,
 Cell ID: 7 | Left Neighbours:  | Right Neighbours: 8,
 Cell ID: 8 | Left Neighbours: 7, | Right Neighbours: 2,
 Cell ID: 9 | Left Neighbours: 14, | Right Neighbours:
 Cell ID: 10 | Left Neighbours: 15, | Right Neighbours:
 Cell ID: 11 | Left Neighbours: 16, | Right Neighbours: 4,
 Cell ID: 12 | Left Neighbours:  | Right Neighbours: 14,
 Cell ID: 13 | Left Neighbours:  | Right Neighbours: 15,
 Cell ID: 14 | Left Neighbours: 12, | Right Neighbours: 9,
 Cell ID: 15 | Left Neighbours: 13, | Right Neighbours: 10,
 Cell ID: 16 | Left Neighbours:  | Right Neighbours: 11,
 Cell ID: 17 | Left Neighbours:  | Right Neighbours: 3,
```

Extraction of table cell contents and reconstruction using Tesseract OCR



Designation Under State or District Accountability Initiative	'Schools Identified Under NCLB (n = 469)	'Schools Not Identified Under NCLB (n= 918)
Low-performing	34%	3%
No special designation	11%	33%
High-performing	2%	18%
Otherinot sure	14%	9%
No other system (other than NCLB)	39%	37%

Output as desired format (e.g. HTML,JSON)

1. To perform table detection, a modified Single Shot Detector (SSD) model is utilized with deformable convolutions for enhanced table detection performance
2. To perform table cell segmentation, Mask RCNN is used to identify and locate table cell instances
3. Using Tesseract OCR, the text content in each table cell is recovered
4. Using the position coordinates of each table cell instance, each table cell's row and column position is determined to reconstruct the table

F1-measure @ IoU Threshold	Table Detection Performance on ICDAR2013 table dataset
0.6	79.9%
0.7	78.5%
0.8	72.2%
Table Cell Segmentation	**Pascal-VOC mAP@ IoU:0.5 on UNLV table dataset**
Mask RCNN + Resnet-101	63.5%

The pipeline thus presents an automated end-to-end solution for table extraction, allowing automated extraction of tables from both digital documents and images. In addition, a modified version of the SSD model is proposed, incorporating deformable convolutions that reports improved table detection performance over the original version. It is portable as it can accept tables from different domains and mediums via conversion of documents to an image-based format. Thus, tables from sources such as scanned documents and photos can be automatically converted to a digital format. In addition, being a computer vision approach, it is robust to the various table styles and fonts present in documents of different domains.