Table 2: Quantitative results on ExcelChart400K with proposed evaluation metrics in Section 6 (the higher the better)

Methods	Bar	Pie	Line
ChartOCR (Ours)	**0.919**	**0.918**	**0.962**
ChartOCR (Ours) + GT Point	0.989	0.996	0.991
ResNet+Faster-RCNN [17, 6]	0.802	-	-
Revision [23]	0.582	0.838	-
ResNet+RotationRNN [17]	-	0.797	-
ResNet+RNN	0.000	0.411	0.644

through the graphical interface for each example, hence we only show qualitative comparison for it by visualizing input images and predicted results.

7.2. Quantitative Analysis

We report the quantitative results on ExcelChart400K with the proposed evaluation metrics in Section 6.

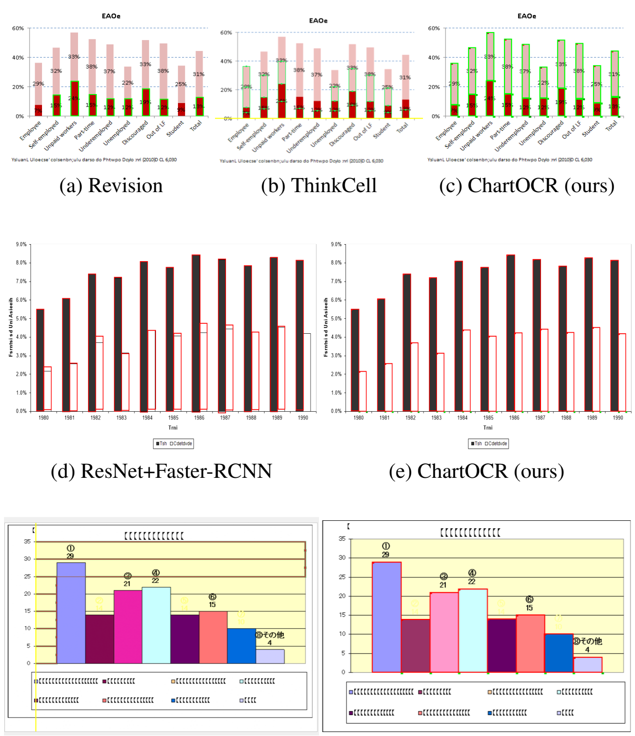
In Table 2, our method has the highest scores in all three types of charts. For **bar chart**, compared with ResNet+Faster-RCNN, our method achieves nearly 14.5% improvement. The improvement is majorly due to the accurate detection of key point positions. As shown in Figure 5, the ChartOCR is more accurate in defining the bars via key points compared with traditional object detection approaches. For **pie chart**, our model outperforms Revision by over 9.5% which is due to the generalization ability of deep key point method. As shown in Figure 7, in the case of detached sectors, key point method can still work well while rule-based method cannot. For ablation study, we also perform an experiment where key point detection results are replaced by the ground-truth key point locations. (See ChartOCR (Ours) + GT Point entry in Table 2). This result can be seen as the upper-bound of our ChartOCR model with the key point based approach. We can see that the performance of bar and pie charts reading has been improved to a nearly perfect score, which means that our rule-based module is exceptionally reliable. For **line chart**, our method improves around 50% compared with the ResNet+RNN baseline.

We also follow the experiment setting[3] of Vis[6] and report the Mean Error Rate in Table 3. For ablation study, we use the ground-truth label information to replace the OCR result, reported as **GT OCR**. Since WebData does not have ground-truth OCR results available, we only compared the methods with and without GT for the FQA dataset. Compared with Vis, ChartOCR shows significant improvement on line and pie type charts. It is because that in Vis, the detection of pie and line components is still based on pure rule-based approach. Ground-truth OCR also greatly reduces the error for FQA data set for bar charts, which indicates that the major error for bar data extraction is caused by poor OCR performance on this dataset.

[3]The comparison is based on reported results in the paper

Table 3: Comparison on the public datasets: FQA and WebData. (* numbers are taken from the original paper)

FQA Mean Error ↓	Bar	Pie	Line
ChartOCR(Ours) + GT OCR	0.093	0.038	0.496
ChartOCR(Ours)	**0.185**	**0.038**	**0.484**
Vis [6]	0.330*	1.010*	2.580*
Revision[23]	0.500	0.120	-
WebData Mean Error ↓	Bar	Pie	Line
ChartOCR(Ours)	**0.285**	**0.439**	**0.740**
Vis [6]	0.450*	0.810*	2.070*
Revision [23]	2.230	0.570	-



(a) Revision (b) ThinkCell (c) ChartOCR (ours)

(d) ResNet+Faster-RCNN (e) ChartOCR (ours)

(f) ThinkCell (g) ChartOCR (ours)

Figure 5: Comparison of the methods on bar charts. First row: stacked bar charts; Second row: clustered bar charts; Third row: tight clustered bar charts.

7.3. Qualitative Analysis

In this section, we compare our method with state-of-the-art methods and show the performance for different type of chart images by visualizing the predicted results.

Bar Chart Cases: In Figure 5, compared with rule-based methods, ChartOCR is more stable on different types of bar