$$loss_{push} = \frac{1}{\mathbf{C}_K^2} \sum_i \sum_{j>i} \max((1 - |e_m^i - e_m^j|), 0) \quad (5)$$

$$loss_{embedding} = loss_{pull} + loss_{push} \quad (6)$$

The total loss of key point detection network for line chart is defined as $loss_{point'} = loss_{point} + \lambda \cdot loss_{embedding}$ where $loss_{point}$ is the summation of the probability map loss and smooth L1 loss described in Section 3.1. We use λ =0.1 in experiments.

To form lines given key points, we adopt the hierarchical clustering strategy to group the embedding of the key points with the classical union-find algorithm. (The details of this algorithm can be found in the supplemental material.) In this way, each cluster contains points that belong to the same line. However, some points may belong to two (or more) lines and they are usually treated as outliers in the clustering algorithm. We call these points *intersection points* and propose the QUERY network to predict which lines they should be assigned to. For each pair, let (x_s, y_s) denote the location of intersection point s, and $e = (x_e, y_e)$ denote the closest point from s that has been assigned to a cluster. We sample K points equidistantly on the line $s - e$. The location of sample points are calculated using following equations:

$$p_k = (x_s + (k-1)d_x,\ y_s + (k-1)d_y) \quad (7)$$

$$d_x = \frac{x_e - x_s}{k},\ d_y = \frac{y_e - y_s}{k} \quad (8)$$

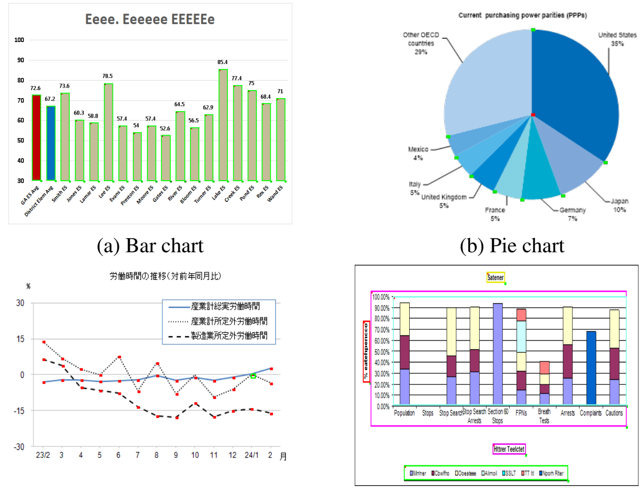
where k means the kth sample point. Since the sample point locations are float numbers instead of integers, we use linear interpolation to obtain the feature of the sample point. Then we can use the QUERY network to take the K sample points as input and classify if point s and e should belong to the same line.

4. Data Set

FQA [6] This dataset contains 100 synthetic images for bar chart, pie chart and line chart. However the variation on chart style not large.

WebData [6] This dataset has the same size as FQA. The images are crawled from the web, and the variation in chart style is much larger than FQA.

ExcelChart400K Deep neural networks easily overfit on small datasets like FQA and WebData, thus we collect a large-scale dataset that contains 386,966 chart images by crawling public Excel sheets from the web. We first capture the chart image with Excel APIs, then extract the underlying data values of the chart. (To protect privacy, we have conducted data anonymization by overwriting texts in the charts with random characters.) The collected dataset not only



(a) Bar chart (b) Pie chart

(c) Line chart (d) Common chart components

Figure 4: Example chart images from ExcelChart400K with annotated ground-truth positions of chart components. a) the bounding box of each bar for bar chart; b) the key point positions of each sector for pie chart; c) the data points of each line for line chart; d) the bounding boxes of chart components (only the position of plot area is used in this paper).

Table 1: ExcelChart400K dataset statistics

Type	train	val	test
Bar	173,249	6,935	6,970
Line	116,745	3,073	3,072
Pie	73,075	1,924	1,923

provides the bounding boxes locations for the chart components but also the numerical readings of the charts. Figure 4 shows some samples and the annotations from this dataset. Table 1 summaries the statistics of our dataset. Compared with previous chart data sets used in[6, 23, 12, 17], this dataset has a wider range of variations in type and style. Moreover, they are authentic images used in real-world scenarios instead of synthesized from data-generation.

5. Training Details

In the design of the keypoint detection network, for all three types of chart images, we use the same backbone network- HourGlass Net with 104 layers. During training, we use Adam optimizer with learning rate 2.5e-4 and decrease the learning rate to 2.5e-5 for the last 5,000 batches. Batch size is set to be 27. $\alpha = 2, \beta = 4$. Soft-NMS is applied to merge key points from the heat map. All experiments are conducted in the same environment with 4 Tesla P100 GPUs. For training details of different type of charts please refer to the supplemental material. A vali-