both deep and rule-based methods. As shown in Figure 2, our method first run common information extraction to obtain key points and chart type. Then, we apply type-specific rules to construct the data components (e.g. bar components, sector components) and data range. Finally, we transform these components into structured data format (e.g. tables). It not only exploits the generalization ability of deep methods, but also generates semantic rich intermediate results as in rule-based methods. When dealing with chart images with new styles, we only need to enrich the training data for key point detection network without changing other parts of the framework. The experiments on three datasets FQA, WebData and ExcelChart400K show that our method has good performance on three major chart types, including bar, pie and line charts.

The contribution of this work can be summarized as follows: (1) We propose CharOCR, a deep hybrid framework that combines the advantages of deep-learning and rule-based methods. ChartOCR achieves state-of-the-art performance on chart data extraction task for all three major chart types. (2) We also design new evaluation metrics for these chart types. (3) We collect a fully annotated chart data set with 400K Excel chart images to enable the training of deep learning models.

2. Related Work

2.1. Rule-based Methods

The feature-based methods [4, 8, 20, 23] have been the mainstream for solving chart element extraction problem. They use color continuous searching and edge extraction to find the raw components. Afterwards, predefined rules are applied to eliminate the wrong candidates. However, those methods highly rely on hand-crafted rules and pre-defined features. They are efficient for the data with certain styles, but they can not generalize well on various types. For example, a rule designed for extracting vertically aligned bars can not be used to find horizontally aligned bars. Methods like ChartSense[11] try to solve this problem by adding user interaction and ask users to correct the mistakes during the process. Although adding user input can achieve better performance, it also increases the time cost significantly.

2.2. Deep Neural Networks

Some works try to solve the chart data extraction problem with deep neural networks. For Bar Chart, [6, 17] adopt the idea of general object detection to detect the bar components by treating each bar as an object. For pie Chart, Liu[17] proposes to use the recurrent network and feature rotation mechanism to extract the data. Despite of the great improvement of time efficiency, the deep neural networks are highly restrained to a certain chart type as well. The bounding box detection on bar charts can not be adapted to other chart types like line charts. In the meanwhile, intermediate result like the data range and plot area information cannot be learned with deep methods.

2.3. Keypoint-based Object Detection

Keypoint-based idea has been adopted in many complicated object detection tasks like pose detection[19], face detection[25] and general object detection[15, 7]. Instead of generating the bounding boxes directly, key point methods output the semantically important key points of the target components. For example, in pose detection, the key points are the critical joints of the human body. In face detection, they are the landmark points of the ear, eyes, lip, nose, and mouth. In the chart understanding task, comparing to directly detecting the object bounding boxes, the key points based methods are more flexible for the detection of various chart objects. There is no need to design specific networks for each chart object (e.g. bars, lines and sectors) anymore. Instead, we only extract the keypoints that defines the object. Although the chart objects vary in shape, they are highly structured. This enables us to reconstruct the chart components based on only key points. After the extraction of key points, some task specific rules are applied to group the key points to form complete objects.

3. Our Method

As shown in Figure 2, the framework can be divided into three major parts: common information extraction, data range extraction and type specific detection. The common information extraction includes the key point detection and the chart type classification. Data range extraction determines the range of the numerical values in the plot area. Type specific detection uses the type-dependent rules to obtain the data components (e.g. sectors for pie chart) of each type of chart. Finally by combing the data components and the data range, we can obtain the numerical chart values.

3.1. Common Information Extraction

Key Point Detection In this step, we extract key points of chart components independent of the chart style. With the universal key point detection model, we no longer need to train separate object detection modules for different charts. For chart images with unseen style, we only need to fine-tune the existing key point detection model by adding more samples that reflect the new chart style.

The key points are defined slightly differently depending on the chart type. For the **bar** chart, the key points are the top-left and bottom-right corner of each separate bar. For the **line** chart, the key points are the pivot points on the line. For the **pie** chart, the key points are the center points plus the intersection points on the arc that segment the chart into multiple sectors. As shown in Figure 3, we adopt a modified version of CornerNet[15] with Hourglass Net[19] backbone