Automatic Annotation of Visualizations

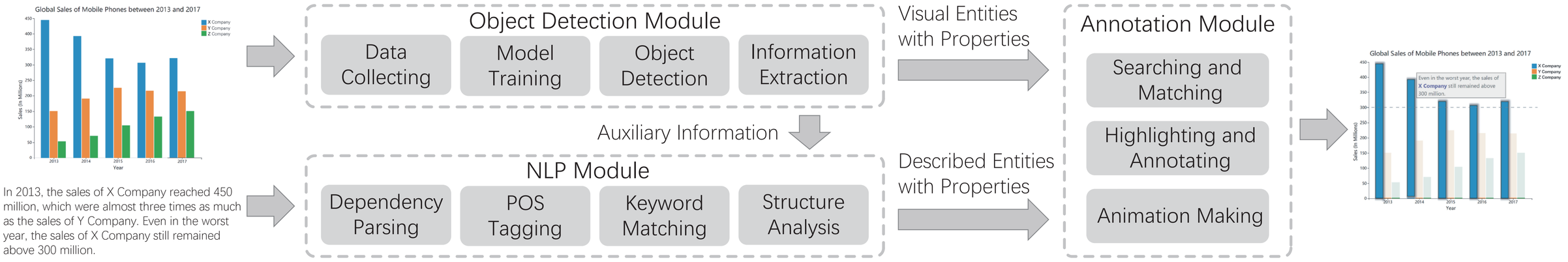
Chufan Lai[1] Zhixian Lin[1] Can Liu[1] Yun Han[1] Ruike Jiang[1] Xiaoru Yuan[1, 2, 3]

1) Key Laboratory of Machine Perception (Minister of Education), and School of EECS, Peking University, Beijing, China
2) National Engineering Laboratory for Big Data Analysis Technology and Application, Beijing, China
3) Beijing Engineering Technology Research Center of Virtual Simulation and Visualization, Peking University, Beijing, China

INTRODUCTION

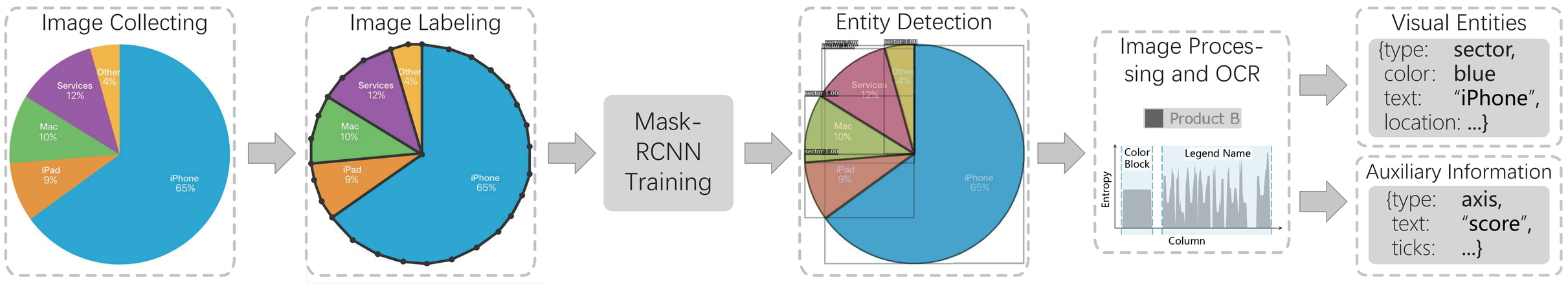
We propose a technique for automatically annotating a visualization based on the user's textual descriptions. In our approach, the annotating task is fulfilled by performing a series of automatic visual searches. Specifically, we extract all visual entities exhibited in the visualization along with their visual properties, using Object Detection and Image Processing techniques. Meanwhile, we parse the description into search requests for visual entities that have been mentioned, using Natural Language Processing (NLP). Knowing what exists there and what to look for, we fulfill the search requests by matching each descriptive sentence with the described area. Each sentence is displayed beside the corresponding area as an annotation. Different annotations are shown in different frames of an animation to promote a vivid presentation. We provide a prototype system that allows the user to upload a visualization image and its textual description, and generate the corresponding annotations.

OVERVIEW



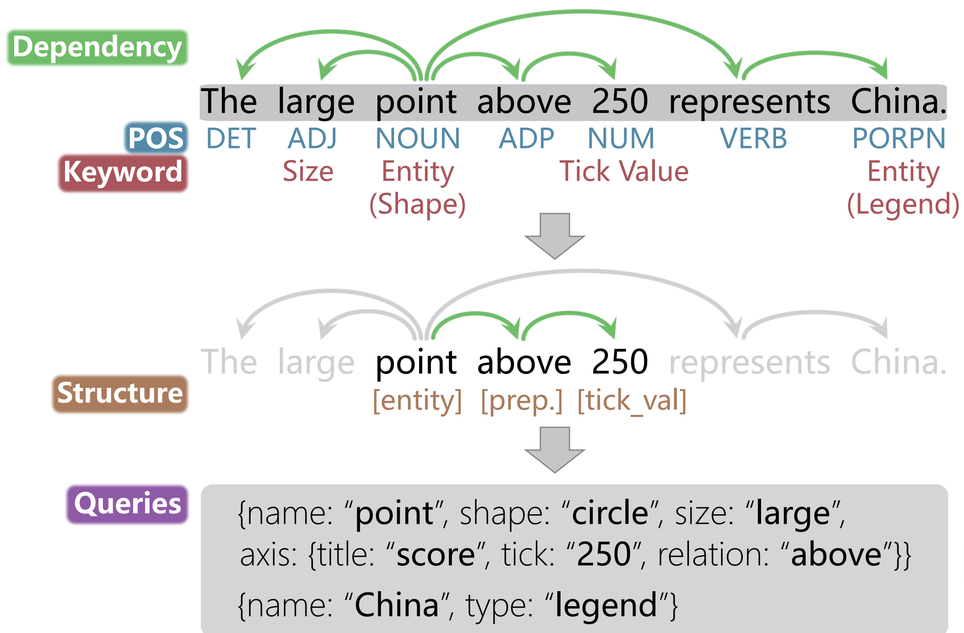
OBJECT DETECTION

Object detection is a class of deep learning techniques originated from computer vision. A trained model can identify not only the type but the exact contour of each object in the image. One state-of-the-art model is called Mask-RCNN. We adopt it here to identify visual entities in a visualization, including both the data entities (e.g. the bars in a barchart) and the auxiliary entities (e.g. axes and legends). We also extract visual properties (e.g. color and size) and other auxiliary information (e.g. legend categories, tick values on the axis) using Image Processing and Optical Character Recognition (OCR).



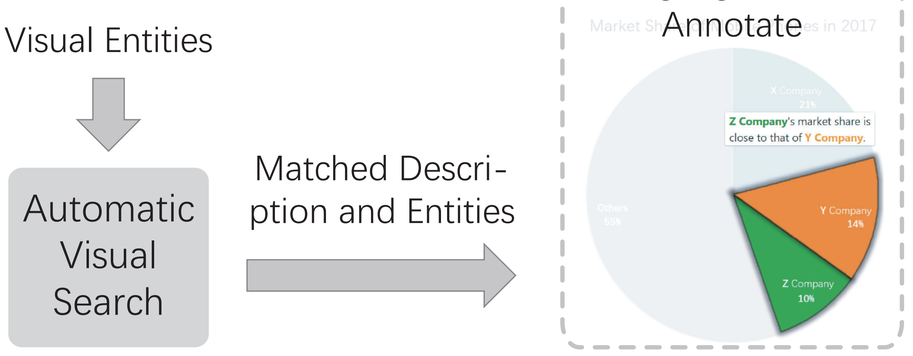
NATURAL LANGUAGE PROCESSING

First, we identify dependencies and part-of-speech (POS) of words. Then we extract the visual properties of any mentioned entity. A dictionary is built to interpret common descriptions (e.g. "large" describes size). The auxiliary information helps interpret descriptions specific to the current visualization (e.g. tick values). A structure library is defined to handle expressions in different forms.

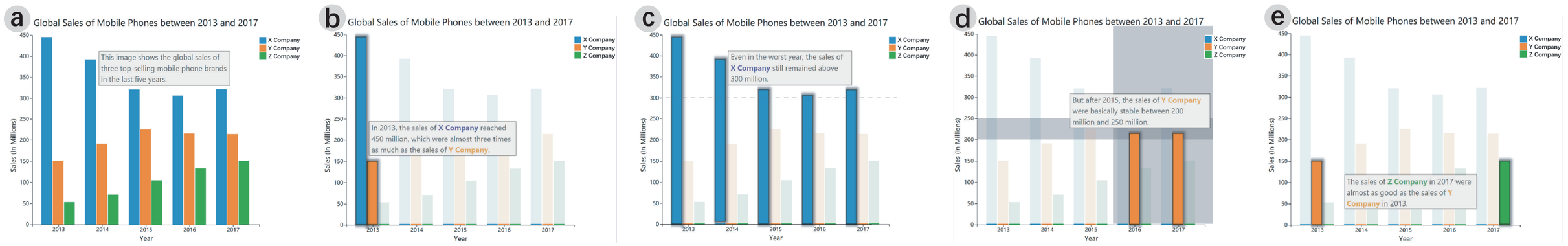


ANNOTATION

We search for any entity mentioned in the description by its ID (like labels and legend categories) or visual properties. Once found, it will be highlighted with the source sentence displayed aside as an annotation. The final animation shows different sentences in their original orders.



CASE STUDY



This is a bar chart depicting the sales of three companies (X, Y, and Z) between 2013 and 2017. The first sentence introduces the background. Since no entity has been described, it is shown without highlighting (a). The second sentence compares the sales of X and Y in 2013. Entities are found by crossfiltering using legends and the axis range (b). The next two sentences describe the two companies respectively (c, d). They differ in whether the axis ranges are specified. The last sentence compares Z in 2017 with Y in 2013 (d). The annotations are always positioned close to the targets without occluding them.

ACKNOWLEDGEMENTS

This work is supported by NSFC No. 61872013 and the National Key Research Development Program of China (2016QY02D0304). This work is also supported by PKU-Qihoo Joint Data Visual Analytics Research Center.

VisPKU

CONTACT

xiaoru.yuan@pku.edu.cn
http://vis.pku.edu.cn