Extracting Information From Images Of Charts

Introduction

Computer Vision is one of the most significant subsets of Machine Learning due to its vast and absolute applications. Ranging from face detection technology to the detection of real-time objects for navigating self-driving cars, computer vision is experiencing a humongous growth in interest and investment. Computer vision can also be used to extract key data insights from graphs and charts. Charts (or data plots) have always been used as an effective and efficient way of communicating all forms of information. The various types of charts extend their usability even further to incorporate multiple parameters and attributes. Charts have been used repeatedly to convey and emphasise a particular point or irregularity that would have otherwise gone unnoticed if represented as a data table or text paragraph. Charts can incorporate much more information using numerous parameters and represent in their plot area than a paragraph or table of the same size. Conversely, interpreting and deducing conclusions from data-dense charts can often be a daunting task, especially when we consider charts that use some sort of referencing legends, convoluted relations, complex patterns, or multiple attributes. This paper intends to review published literature that describes all the different approaches that have been carried out to categorise, interpret and deduce a reasoned conclusion from various types of charts using computer vision models.

Literature Review

Since object detection in images has been an area of research for over a decade now, Ganguly et al. (2020) approach the problem of extracting information from charts by raising a very interesting proposition: How would the object detection models fare on scientific graphs? Given the amount of background research that has been conducted around object detection, they use popular object detection models and calculate the mean Average Precision (mAP) of these models over very high Intersection of Union (IOU) ratios. Their use of a high IOU of around 0.9 is justified wholly since scientific charts require a much higher overlap region as compared to natural images since even the slightest variation in these overlap regions can cause the prediction to change drastically. As expected, the traditional object detection models do not perform well on scientific charts since they are drastically different from natural images that require a much lower IOU setting (greater than or equal to 0.5). Ganguly et al. do acknowledge this finding and subsequently implement some minor modifications to existing models by combining ideas from different object detection networks, which does drastically improve the mAP values of these models, but these modifications cause some drawbacks: accuracy on text objects is low and the inference time is very high. These challenges motivate them to propose an architecture, named PlotNet, which contains '(i) a fast and conservative region proposal method based on Laplacian edge detectors, (ii) an enhanced feature representation of region proposals that include local neighbouring information, (iii) a linking component that combines multiple region proposals for better detection of longer textual objects, and (iv) a custom regression loss function that combines smooth l1-loss with an IOU-based loss designed for improving localisation at higher IOUs' (Ganguly et al., 2020). PlotNet is tested and trained on the PlotQA dataset which performed significantly better than other models that are tested on this dataset. Ganguly et al. do provide implementation, training, and testing details of their PlotNet model that conclusively prove its legitimacy and applicability in automated reasoning over charts. The key distinguishing factor between object detection models and chart reasoning models is well established in this work. This distinction clearly highlights the need for more accurate models that can analyse information from key elements in images of charts since the traditional object detection models fail drastically in this regard. By creating a model that consists of advantageous components from previously implemented object detection models, Ganguly et al. elicit the fact that researchers can be inspired by previous object detection algorithms and rely upon their robustness whilst implementing a new model to extract information from a chart.

Methani et al. (2020) have taken a strong stance against the previously established datasets and approaches to extracting information from charts. They outline in detail the reason existing datasets are not adequately tackling the key challenge of reasoning over data plots. They critique the established datasets for assuming that the answers to questions based on interpreting charts come from either a fixed size vocabulary or by extracting text labels from the original image of the chart. Once the shortcomings of these datasets are established, they move on to propose a PlotQA dataset with 28.9 million question-answer pairs over 224,377 plots on data from real-world sources and questions based on crowd-sourced questions templates. They establish the legitimate distinction between PlotQA and other datasets by stating that 80.76% of the out-of-vocabulary (OOV) questions in PlotQA have answers that are not in a fixed vocabulary. They also move forward onto analysing and critiquing the models that have been implemented in the past and subsequently propose a hybrid model to overcome some of the significant challenges and shortcomings of the formerly implemented models. The one significant shortcoming that is very prominently highlighted is the fact that existing models can only either read answers from the image or pick the answer from a fixed set of vocabulary. Methani et al. proposed model uses a question classifier to segregate the question and determine if the answer lies in a fixed vocabulary set or if it requires more complex reasoning. This classification allows them to implement a multi-staged model as a pipeline to answer these complex questions. Running this new model over established datasets, they achieve much higher accuracy than older models. This paper has some significant implications over the entire visual question answering (VQA) sub-section of computer vision. By proposing this idea that answers can also be out-of-vocabulary, they bring forward a very glaring shortcoming of all the previously implemented models and datasets. By accumulating a humongous dataset of millions of QA sets over hundreds of thousands of graphs, they have set forward an intuitive and pressing challenge for future researchers to come up with models that can achieve comparable accuracy percentages to be applicable in real-life usage. While concluding they do acknowledge that although their model outperforms the previously implemented models, this performance needs to be enhanced majorly to be significant in real-world scenarios.

Luo et al. (2021) propose a unified method called ChartOCR that combines two traditional predefined methods: rule-based method and deep neural networks. Their proposed approach tackles the chart components detection problem with key point detection methods which gives them the ability to expand the scope of ChartOCR to include various types of charts. Once they have accessed the key points of the chart, they can categorise the chart type. These key points are then passed onto a deep hybrid framework that combines the advantages of both deep neural networks and rule-based methods. Luo et al. have provided extensive research information to justify their use of both these methods together, by highlighting the shortcomings of both these methods individually. By stating that rule-based methods 'highly rely on hand-crafted rules and pre-defined features' (Luo et al., 2021) it is evident that this method can only be used for particular types of charts at a time and is not expandable to include all the types. Similarly, they deduce that deep neural networks are highly restrained to a certain chart type only which makes them inefficient given a dataset with various types of charts. Lou et al. build a framework consisting of two major layers: Common Information Extraction (the chart type and key points are detected), Data Range Extraction (range of the data represented by the chart is interpreted) and Type Specific Detection (the chart objects are extracted). They define the algorithms used at each of the layers to clarify the steps that are performed and propose evaluation metrics for three chart types: Bar Charts, Line Charts and Pie Charts. Having established their approach and methodology, Lou et al. provide detailed experimental analysis about the performance of their framework against a rule-based method, a deep-learning-based method, and an off-the-shelf commercial product. This experimental analysis conclusively proves that ChartOCR outperforms all these models in all the analyses. Although their ChartOCR model does have superior performance, it does not tackle a pressing issue that is very significant in interpreting information from charts i.e., answering out-of-vocabulary questions. The researchers do not run their model on larger datasets such as FigureQA or DVQA that have been used repeatedly by other researchers to establish the legitimacy and effectiveness of their implemented models. On comparing the results of ChartOCR on these more data-dense and complex datasets, we can observe and analyse if their approach is relevant and can extend towards real-world applications. Also, it would be interesting to observe how this approach can be extended to include the interpretation of more types of charts, especially ones that have more chart components.

Davila et al. (2021) very recently published an extremely detailed survey based on the contributions made by researchers for mining and automated reasoning of charts. This survey references over 175 previously published papers, spanning 15 years in the past, as it digs down deeper to analyse, critique and compare the contributions of these researchers. Davila et al. begin by stating the importance of charts in all sectors of human interactions and they acknowledge the importance of previously published surveys about automated chart reasoning methods. They highlight that although there is a significant amount of work being conducted to extract information from charts, there are still some major open challenges that need to be overcome before we can use automated chart reasoning methods in the real world. They list out and discuss these open challenges at the end of their survey before concluding their study. Davila et al. also study, analyse and discuss the various datasets that have been used and developed by previous researchers while also dipping into the methods used to accumulate these datasets. A fascinating section in this study focuses on handling the extraction of information from multi-panel charts. This section is very interesting because multi-panel charts pose a very different challenge than regular charts since there are multiple captions that need to be associated with the corresponding charts. This step requires caption analysis for not only interpreting the data but also mapping the correct caption to the correct chart. Once they have established and detailed the significance, challenges, and methods to extract charts, Davila et al. move forward to describe the various chart classification techniques. They analyse both high-level and low-level classification models across many classification techniques such as heuristic and deep learning. Another interesting section in this survey is the discussions around applications of automated chart analysis which reinstate the requirement for further research in this research area by listing down various applications of automatically extracting information from charts. These applications cover vast areas ranging from chart accessibility to chart retrieval and Visual Question Answering. Their study enables readers to understand the key challenges that previous researchers faced whilst extracting information from charts and sets forth the methods that they used to overcome said challenges. The detailed analysis published in this survey along with a set of challenges that are still relevant in this research sector gives future researchers a substantial starting point to build relevant models that can extract and reason information from charts.

Kafle et al. (2018) have clearly enclosed the differences of VQA and DVQA. They have defined the limitations of VQA and solved them by presenting DVQA. One basic issue with VQA was the use of natural images which they solved in DVQA using bar charts. Some other issues with VQA were the use of a fixed vocabulary dictionary that makes it difficult to answer questions about a variety of images and its language that possesses fixed semantic concepts. A huge amount of data is compacted and visualised using data visualisations such as bar charts, plots, and pie charts. DVQA demands answering questions about bar charts. There were 3 major contributions made by them. 1) The DVQA dataset that Kafle et al. described contained over 3 million image-question pairs which tested the different forms of understanding their diagrams, a) structural understanding; b) data retrieval; and c) reasoning. 2) They also depicted that both the baseline and state-of-the-art VQA algorithms could not answer many questions in DVQA. 3) Kafle et al. proposed two models that could handle words that were unique to a specific image. One was the end-to-end neural network that could read answers from the bar charts and the other was the one that could encode generic questions and answer them. From these contributions, they showed that DVQA needed more than just data extraction. They needed the DVQA questions to interpret the context and the meaning of the question, read them into and from the diagrams and correctly answer them. Kafle et al. treated DVQA as an open-ended question answering task. Talking about the DVQA dataset images, they used the tool, Matplotlib to generate charts with varieties in both appearance and style, such as number of bars, orientation, difference in colour, width, spacing and location of labels and legends. They also used three types of data in their charts such as linear, percentage and exponentials. Their dataset contained three types of questions such as a) structural understanding; b) data retrieval; and c) reasoning. Kafle et al. also minimised the bias by randomising the generation of charts so they could ensure there was no correlation between styles, colours, and labels. Once they generated the DVQA dataset, they began training by previously designed baseline models. One of their models was closely related to one the baseline model. Kafle et al. proposed two models namely MOM and SANDY. Multi-Output Model (MOM) used a dual-network architecture where one of its subnetworks was able to generate *bar chart specific* answers and the other subnetworks was able to generate *generic* answers. They trained a classifier to classify which question belongs to either of the subnetworks generating the most accurate answers. SANDY on the other hand (Stacked Attention Network for DYnamic Encoding Model) could directly generate chart specific answers to chart specific questions. They established their findings by backing it up with relevant results. Although DVQA acknowledges that many of the questions based around charts are out of vocabulary (OOV) types, they don't make any significant advancements in attempting to resolve this parity between real world scenarios and experimental data. Kafle et al. mentioned that they will enhance their dataset by including other types of charts which would really help the diversity of the dataset since currently it is just bar charts.

Zhou et al. (2020) adopted a neural network-based object detection model to work on reverse-engineering bar charts. Their model works only on bar charts since they are the most used chart types over the Internet. Extracting data from charts seemed to be a tedious job since they were either extracted manually or had automated tools extract them. This is where reverse-engineering bar charts falls into place. Reverse-engineering bar charts extract chart information in two types: textual and numeric information. Neural networks, on the other hand, possess the feature of self-learning ability that is more efficient than other models. Zhou et al. implemented the idea of using neural networks to extract textual and numeric information from bar charts. While collecting bar charts, Zhou et al. has made 7 assumptions with respect to different chart forms and their design styles. After specifically choosing and studying real-world bar charts, they discovered that a large amount of synthetic data was clustering the charts and that they needed to be transformed into small amount of relevant and real-world data. Zhou et al. used a Python script to generate quality charts which were flexible to their assumptions and with important data. These bar charts that were generated had specific considerations such as figure style, design style, font size, number of bars, their lengths, heights, and orientation and so on. For extraction of textual information, they used a neural network-based model to localise and classify textual information simultaneously. This ensured efficiency and accuracy of the extraction process of previous works done in this field. It followed a bottom-up approach that contained three steps, a) localising the elements of the chart; b) recognizing them for the fundamental text; and c) classifying them into their text types. For numeric information extraction, they used a two-step framework along with an attention mechanism to achieve robustness and accuracy. The two-step framework was the encoder-decoder framework that put together two neural networks such as conventional neural networks (CNN) and recurrent neural networks (RNN). Zhou et al. displayed some results that indicated their method to surpass some previous work already done but there were some degradations as they mentioned. They should've included some more assumptions present in while choosing bar charts that could have led to better accurate outcomes and probably lessened the degradations. Another limitation is the use of only bar charts, while pie charts, plots and other data visualisation methods should be worked on too.

An underappreciated application of automatically reasoning data from charts is assisting the section of the human population that are visually impaired. Charts and graphs are very useful representations of data provided the reader has the ability to parse and contemplate the information encrypted in the chart. Readers that do not possess the ability to visually see the charts find themselves at a definite disadvantage. Models that can interpret chart data efficiently and effectively will provide much-needed accessibility assistance to visually impaired readers. Shahira and Lijiya (2021) discussed this application of extracting visual information from charts in detail whilst conducting an in-depth survey of all the implemented methods and datasets that have been worked upon by various researchers. They start off by restating the current assistive technology that is enabling visually impaired people to access text content and image caption (or alt text). They provide their motivation for conducting such a survey by quoting statistics from the World Health Organisation that clearly reinstate the importance of implementing models to read data from charts. In a very structured table, Shahira and Lijiya describe the solutions introduced in the chart, data understanding and future possibilities from a few related works. This table compared the merits and limitations of all the previously implemented methodologies to extract information from a chart. This is a good jumping-off point for them to build on and introduce the various methods that can be used for identifying charts. These methods have been broadly classified into three categories: Modality Based Approaches, Traditional Methods and Deep Learning Based Methods. Each of these categories has been studied in detail but since a lot of recent advancements have been made in Deep Learning Based Models, Shahira and Lijiya produce a quantitative analysis of various Deep Learning Based models where they denote if a particular model handles the different types of charts. This analysis is presented in a table that is very concise and helpful for researchers that would like to build upon previously implemented models. Another striking feature of this survey is their section where Shahira and Lijiya have discussed the various evaluation metrics (IOU, precision, recall, accuracy and F1) used to measure the success of implemented models. These metrics have been used by numerous researchers in the past and understanding their capabilities will enable future researchers to build and compare new models with the existing ones. This compilation of relevant information in a survey by Shahira and Lijiya provides a key starting point for any future researcher in this area. Their detailed analysis and list of open challenges provide a clear background summary and futuristic scope of extracting information from images in charts.

Mathew and Meena (2021) presented a detailed survey about various implementation methods for object detection from scientific plots. Charts have been the most effective way of representing data in a compact yet accurate way. Data extraction from charts has been the most common method to be able to analyse, interpret and then answer questions related to them. The major limitations of these object detection procedures are low accuracy and efficiency. Computer vision might solve these problems. Mathew and Meena focus on the functionality of detection of elements from charts, value extraction and its analysis. During the survey, they reviewed 10 different datasets of various data visualisation types such as scatter plots, pie charts, donut graphs and line graphs and their algorithms along with their performance. In this analysis, almost 70% algorithms have success ratios over 90%, along with their individual set of limitations. Hence, Mathew and Meena focused on feature detection and extraction methodology using Deep Learning Feature Detection and Extraction techniques. They studied some conventional neural network (CNN) architectures that proved to be the solution to high performance image classification. After thoroughly examining them, Mathew and Meena found some gaps in the existing research. One of the most important gaps was the use of only linear plots such as bar charts. A new methodology is yet to be discovered for nonlinear plots such as scatter plots, pie charts, donut charts and line charts. Mathew and Meena concluded that if given the time, the methods used for scientific plot analysis will become more efficient and accurate.

We are currently reviewing more papers and will update this document in due time.

Summary

Computer Vision encompasses a wide range of subsections including but not limited to object detection in images. Although initial object detection studies and datasets have been criticised severely for being plagued with bias and their inappropriate categorization of images, recent research has worked largely in favour of negating the aforementioned problems. While object detection algorithms for natural images have been widely worked on, there are still relatively fewer attempts to automatically interpret data from images of charts. Algorithms possessing the ability to reason and contemplate data encoded within charts can have enormous real-world applications. This review of existing literature around automatically interpreting charts reveals that although there have been some significant advancements in this sector, there are still many challenges that need to be overcome if we hope to use these techniques in the real world. The concept of object detection from images has been largely researched in the last decade. Following the large advancements made in object detection, we can be certain that with sufficient investments, researchers in the near future possess the potential to implement highly accurate models that can interpret, analyse, and deduce reasoned conclusions from charts.

Bibliography

Davila, K., Setlur, S., Doermann, D., Bhargava, U.K. and Govindaraju, V., 2020. Chart mining: a survey of methods for automated chart analysis. IEEE transactions on pattern analysis and machine intelligence.

Ganguly, P., Methani, N., Khapra, M.M. and Kumar, P., 2020. A Systematic Evaluation of Object Detection Networks for Scientific Plots. arXiv preprint arXiv:2007.02240.

Luo, J., Li, Z., Wang, J. and Lin, C.Y., 2021. ChartOCR: Data Extraction from Charts Images via a Deep Hybrid Framework. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1917-1925).

Methani, N., Ganguly, P., Khapra, M.M. and Kumar, P., 2020. Plotqa: Reasoning over scientific plots. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1527-1536).

Kafle, K., Price, B., Cohen, S. and Kanan, C., 2018. DVQA: Understanding data visualisations via question answering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5648-5656).

Mathew, J., 2021, May. A Survey on Object Detection from Scientific Plots. In 2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP) (pp. 94-99). IEEE.

Shahira, K.C. and Lijiya, A., 2021. Towards Assisting the Visually Impaired: A Review on Techniques for Decoding the Visual Data from Chart Images. IEEE Access.

Zhou, F., Zhao, Y., Chen, W., Tan, Y., Xu, Y., Chen, Y., Liu, C. and Zhao, Y., 2021. Reverse-engineering bar charts using neural networks. Journal of Visualisation, 24(2), pp.419-435.