



Review of chart image detection and classification

Filip Bajić¹ · Josip Job²

Received: 7 November 2021 / Revised: 5 October 2022 / Accepted: 14 December 2022 / Published online: 12 January 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

This paper presents a complete review of different approaches across all components of the chart image detection and classification up to date. A set of 89 scientific papers is collected, analyzed, and enlisted into four categories: chart-type classification, chart text processing, chart data extraction, and chart description generation. Detailed information about problem formulation and a research field is provided, and an overview of used methods in each category. Each paper's contribution is noted, including the essential information for authors in this research field. In the end, a comparison is made between the reported results. The state-of-the-art methods in each category are described, and a research direction is given. We have also analyzed the open challenges that still exist and require the author's attention.

Keywords Chart classification · Chart image processing · Image processing and computer vision · Information visualization · Review and evaluation

1 Introduction

The first attempt to graphically display values dates back to the first millennium. It is a calculation of the movements of the sun, moon, and planets throughout the year, and the author is unknown [1]. The first graphical representation, "ligne de vie," was made by Christiaan Huygens in 1669 [2]. It depicts a continuous distribution function and demonstrates how to find the median for a person's remaining life. In 1765, Joseph Priestley drew up a historical timeline showing the life span of 2000 famous persons dating from 1200 BC to 1750 [1]. The above examples are a common feature of the authors recognizing the potential of additional visual elements when presenting complex information. Both contributed to statistics and the development of the visual presentation of data.

William Playfair (1759–1823), a Scottish engineer and political economist, brought the revolution in statistical graphics. Playfair was one of the first to use data not only

to educate but also to persuade people. He is widely considered the inventor of the first line, bar, pie, and circle chart [1]. Playfair understood that data visualizations could allow the brain to process information more efficiently by reducing memory use. Using data visualizations requires less attention from the reader to store important information in long-term memory. In the nineteenth century, statistics progressed, and statistical data became available, creating the need to display complex data. In the twentieth century, the need for information summarization grew. Image representations are used meaningfully to present important information relevant to the research. Data visualizations are one form of image representation that enables a clear and complete understanding of the relationships between the data. Graphic representation becomes helpful in understanding the obtained results and can be used to estimate values that are not directly determined by measurement using interpolation and extrapolation methods. The graphical presentation is essential for identifying unusual or unexpected results and makes it easier to compare different values, trends, and relationships. The human ability to notice quickly and easily visually is based on the brain's ability to detect regularities and imperfections. This ability happens subconsciously. The comparison happens before thinking about it.

In today's world of advanced Internet technologies, data and information play a significant role. The Internet is an interactive medium that processes vast amounts of data and information every second. Humans can hardly understand

✉ Filip Bajić
filip.bajic@srce.hr
Josip Job
josip.job@ferit.hr

¹ University of Zagreb, University Computing Centre, 10000 Zagreb, Croatia

² Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, 31000 Osijek, Croatia

data when it is piled up and unstructured. Tables containing a large amount of data are not easily readable and require mental effort to develop useful information. Statistics are often represented by numbers, which are sometimes hard to read and distinguish between important and unimportant information. Unique methods and tools have been developed to display the data in the graphical form [3]. This way of presenting data is called data visualization.

Data visualizations are graphs or diagrams created by the problem of deflating data into one visual piece that can be easily comprehended by the individual, allowing for easier dissemination, understanding of information, and decision-making. To better structure the data, customized charts have been developed. Each type of chart is designed to represent a specific type of data.

Data visualization is used in every aspect of life, from mathematics, statistics, and analytics to anywhere where there is a need to identify patterns in the dataset or to explain those patterns to a broader audience. Data visualizations can be found in various prints and digital documents and, as such, are not equally accessible to everyone. The problem arises when search engines need to include data visualization results or when blind and visually impaired people are trying to access data visualizations. Many authors and designers are unfamiliar with the accessibility problems that blind and visually impaired people face; hence, they do not consider making documents and visualizations accessible. As a result, most electronic documents do not include additional information (metadata, alt text, descriptive tags, or table), or the provided information is very general, short, and inadequate. According to Bajić et al. [4], a visualization dataset consisting of 2702 images collected using Google Image search contained 39 (1.44%) images with descriptive metadata. Metadata are useful for Internet browsers, but people using various screen readers require more information to understand the visualized content in question. As seen from history until today, such information is locked inside data visualizations. Data visualizations exploit human visual perception to transmit information efficiently and effectively. Such representations are not intended for use in the digital world. While people can easily decode data visualizations and create tables, computers cannot [5]. Due to these problems, various scientific types of research have been made to increase the availability of data visualizations, that is, to enable a more efficient classification of data visualizations and thus allow a more accurate and detailed interpretation of the content of data visualizations [6–9].

In Sect. 2, the problem formulation is presented. The details about the literature search and literature evaluation are given. Section 3 presents specifics about each collected scientific paper, such as a key approach, achieved results, and additional information. This section comprises four

subchapters: chart-type classification, chart text processing, chart data extraction, and chart description generation. Section 4 summarizes the conducted research and compares the achieved results in this research field. Finally, Sect. 5 shows the final remarks on this conducted research.

2 Related work

In this section, the primary problem formulation with summarized aims is presented. The literature overview is given, which includes detailed information on the process used for collecting the scientific papers. All collected scientific papers are visualized on a timeline, and the documents and the authors with the most contribution are pointed out.

2.1 Problem formulation

Printed documents can be digitized. A document contains heterogeneous and complex information linked together into one visual unit, significantly impacting readers. The detection or classification of content is formulated by decomposing a document into textual and graphical data. Text recognition is a well-known problem addressed by various optical character recognition (OCR) systems and is not the topic of this research. The document analysis, segmentation, and image extraction are also out of the scope of this work. Extracted image content can be classified into thousands of categories using various state-of-the-art methods. This paper integrates scientific research that uses the classification of a data visualization image into one or more categories as its first step. With this said, we isolate key related questions in our research. The answers are found in different chapters of this scientific paper and are well referenced. Some of the questions that we asked ourselves when creating the foundation of the research are as follows:

- Who are the involved authors, who among them stand out, and why?
- What are the benefits of chart-type classification, and where is it integrated?
- Is there a root process that is shared among the researchers?
- What are the numbers that authors are reporting, and can they be compared?
- Which scientific papers are considered state-of-the-art, and why?
- Where did the research begin, and where is the timeline going?

The summarized aims of scientific papers using the classification of data visualization are:

- To create a database using various methods of mining data visualizations [10, 11].
- To allow Internet search engines to include data visualizations in a user's query [12, 13].
- To create a summary description of the data visualization to allow access to blind people and all people with impaired vision [14, 15].
- To export data from a visual representation (creating a table with original data) [16, 17].
- To create a new data visualization from an existing one (e.g., convert bar chart into a pie chart) [18].
- To create a rating of data visualization (e.g., good, bad) [19].
- To retrieve similar topics based on information from data visualization [11].
- To create accessible data visualizations for various screen readers [20, 21].
- To create a Visual Question Answering (VQA) system (e.g., Natural Language Generation (NLG), Natural Language Processing (NLP)) [8, 13]

2.2 Literature search

Based on the problem formulation, a search of available literature was undertaken. All major databases were searched (Web of Science, Scopus, IEEE, etc.). The search extended beyond computer science, as this issue exists in medicine, mechanical engineering, natural sciences, and more. The search was conducted throughout the all-available period. When searching for a paper, the abstract, title, or keywords are required to contain terms like chart parsing, classification, detection, recognition, etc. The study was repeated by replacing the word chart with graph, diagram, and visualization. By the end of the literature search (October 20, 2021), the total number of scientific papers collected was 89.

Among the collected papers, there are three reviews. The first review was written by Lyu et al. in 2013 and published as a conference paper. The author covers earlier (traditional) methods for document segmentation, chart image classification, and chart data interpretation [7]. The second review was written by Davila et al. in 2020 and published in Transactions on pattern analysis and machine intelligence. The review is focused on the technical aspects of chart image mining. The author covers in detail methods for extracting charts from documents, classifying and interpreting data visualizations by type, and provides an overview of applications of chart mining and datasets for training and evaluation [8]. The work is organized around five main steps for automated chart mining: chart image extraction, multi-panel segmentation, image classification, chart data extraction, and the usage of extracted data. The work also analyzes the most

used chart types, the methods used for the classification process, and a quantitative evaluation of used datasets. In the end, the author discusses achieved results, future research, and persisting open challenges. The third review was written by Shahira and Lijija in 2021 and published in IEEE Access. The paper reviews the literature on chart image understanding and information extraction. The study focuses on data extraction from chart images to aid visually impaired people, and conventional and deep learning methods are discussed [9]. While we highly recommend reading them, none of the aforementioned reviews provide a graphical comparison of the achieved results and used methods in chart image processing. Our work also identifies the most cited scientific papers in chart-type classification, chart text processing, chart data extraction, and chart description generation. The provided tables in each subchapter contain information extracted from the literature.

All collected scientific papers are presented on the timeline in Fig. 1. The circle radius shows the total number of citations obtained from Google Scholar. PostGraphe [22] is the oldest scientific paper in this research field. In the last ten years, the research field has gained importance, evident from the number of circles in Fig. 1. Most papers come from conferences. Fewer scientific papers can be found in journals, technical reports, and PhD theses. At the time of writing this review, eight papers are in the pre-print state. The authors presented research results from 40 unique conferences, only six of which have had three or more scientific papers related to the above-mentioned problem. The conference with the most related papers is the International Conference on Document Analysis and Recognition (ICDAR), with ten papers. The International Conference on Image Processing (ICIP), The International World Wide Web Conference Committee (IW3C2), Special Interest Group on Accessible Computing (SIGACCESS), International Workshop on Graphics Recognition (GREC), and The Eurographics Conference on Visualization (EuroVis) each have three presented papers.

2.3 Evaluation

Across the period of 25 years, 268 authors were included in the research, of which 227 authors published only once, 26 authors published two times, seven authors published three times, and eight more than three times. Weihua Huang, Chew Lim Tan, Daniel Chester, Stephanie Elzer, C. Lee Giles, Sandra Carberry, Seniz Demir, and Yan Ping Zhou are the most active authors in this research field. These authors' total number of citations is shown in Fig. 2. All data for comparison is collected from the same source, Google Scholar. In Fig. 2. Jeffrey Heer is the only author in the group with only two publications. Many of these authors' publications can be considered state-of-the-art. The authors' publications in Fig. 3 are the most convincing and provide the most sig-

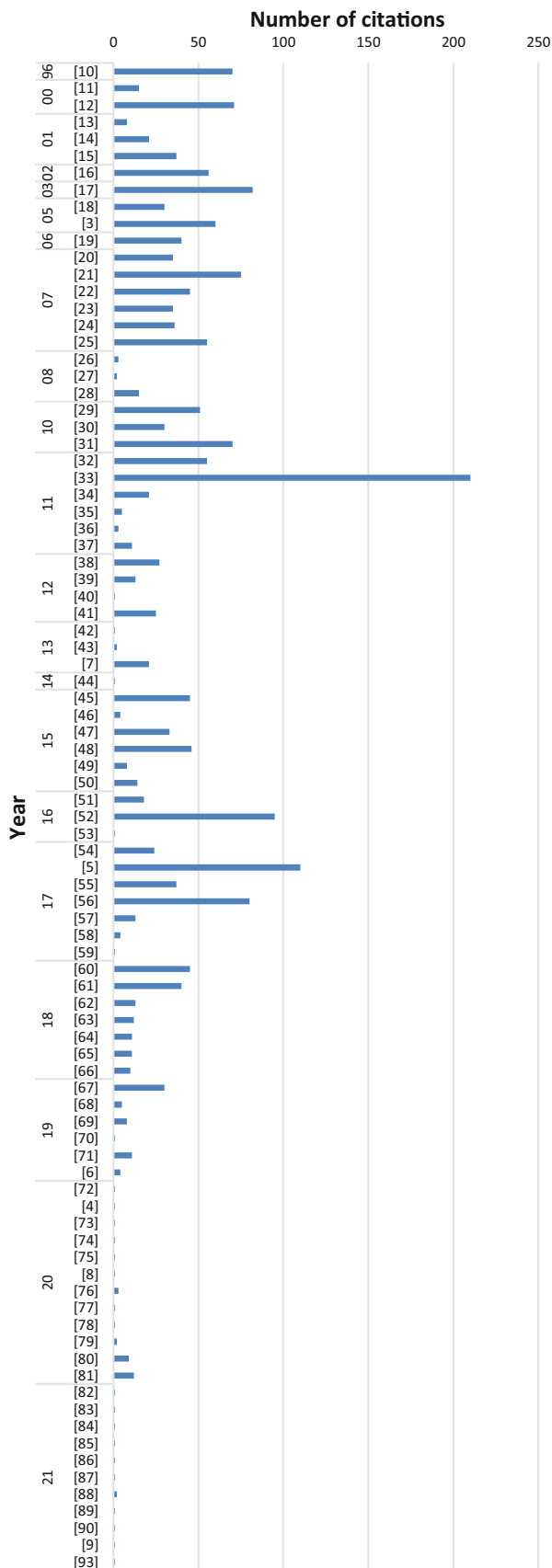


Fig. 1 The timeline of collected scientific papers organized from the oldest (1996) to the newest (2021)

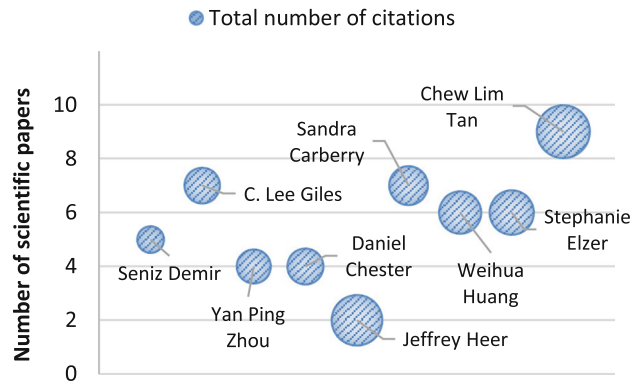


Fig. 2 Bubble chart showing a relation between the total number of citations and the number of scientific papers by author. Only authors with the highest number of citations are included

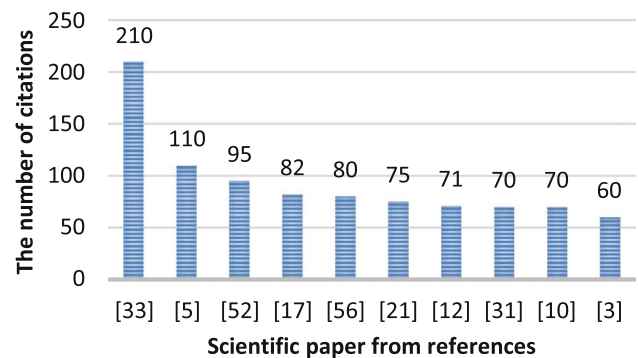


Fig. 3 Bar chart showing top 10 the most cited publications (source Google Scholar). ReVision [18] is the most cited paper in this research field

nificant scientific contribution. More details about their work and results, and other publications, which also provide a significant scientific contribution, will be provided in the next section.

2.4 Contributions

This review aims to survey the current literature on chart-type classification, chart text processing, chart data extraction, and chart description generation. The research started in 1996, and since then, many state-of-the-art methods, key approaches, and techniques have changed. By analyzing Fig. 1 from 2017 to 2021, it can be seen that the number of scientific papers almost doubled, which can be subjected to the spread of machine learning and ever-growing types of neural networks in all fields of chart image detection and classification. The main contributions of this paper are:

- The categorization of collected papers into chart-type classification, chart text processing, chart data extraction, and chart description generation

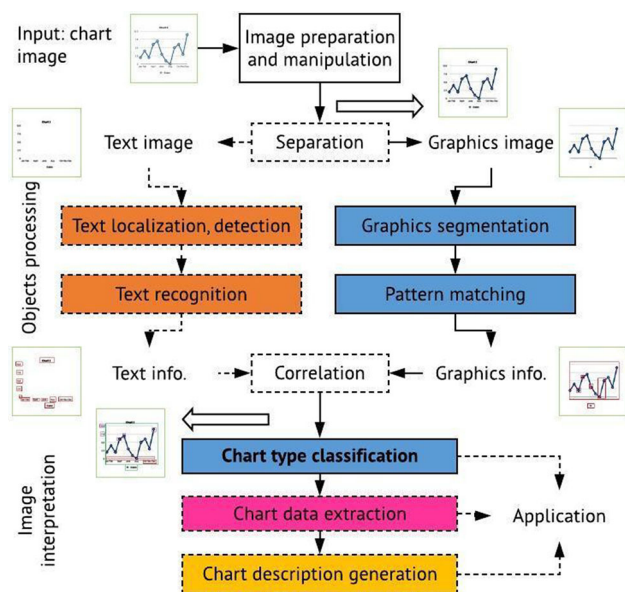


Fig. 4 The most basic concept of our research. Versions of this pipeline can be found in all collected scientific papers. The dashed blocks show the parts that are not obligatory for chart-type classification

- The graphical and analytical comparison of the achieved results in all four categories
- The discussion of different key approaches
- The known challenges and a direction for a future research

3 Research scope overview

When a chart is created with software (Microsoft Excel, Matlab, D3, Plotly) or drawn by hand, the chart elements remain accessible and can be modified. When that same chart is saved or digitized as an image, all the structural information about the chart becomes inaccessible. Our research and scientific papers about chart recognition and interpretation can be divided into four subchapters: chart-type classification, chart text processing, chart data extraction, and chart description generation. The above-noted information can be extracted from the chart itself using different approaches, methods, and algorithms. The most basic concept of our research can be seen in Fig. 4. Each of the following four subchapters consists of a table that summarizes existing scientific papers in that research field, Tables 1, 2, 3 and 4, and a pipeline description shown in Fig. 4. As noted in the previous chapter, all collected scientific papers should have chart-type classification as the first step in chart image processing. Analyzing the results, we noticed that 33 scientific papers do not provide information about chart image classification or use manual classification of chart images (human-annotated). Those scientific papers are not included in Table 1. The same logic is

applied to creating other tables in this chapter. If the scientific paper does not explicitly provide information about the problem-solving, it is not included in that specific table.

In contrast, many scientific papers deal with solving multiple problems and providing all the required information. Those scientific papers are listed in various tables with different information related to the results in that specific research field.

The average classification accuracy that the authors' report may vary from the one presented in Table 1, as this is calculated only on ten chart types. Some authors only include average classification accuracy without any classification accuracy by type, and we do not include those numbers. The last column in Tables 1, 2, 3 and 4 is a dataset size. This number is a sum of different datasets split into training, testing, and validation. We are using this number because while some of the authors report all three dataset sizes, most of the authors do not. Many newer scientific papers use multiple datasets and often combinations of datasets from other authors, making it challenging to distinguish the actual dataset split on training, testing, and validation. The most common dataset ratio for training ranges from 70 to 80%, and for testing from 20 to 30%. The validation dataset often is omitted, or the numbers depend on the author. In Tables 1, 2, 3 and 4, the following abbreviations are used: text and graphic separation (T&GS), text and graphic correlation (T&GC), algorithm (alg), and image (img).

3.1 Chart-type classification

Image classification is a well-studied process in computer vision, which refers to classifying an image according to the visual content. It is one of the most critical tasks in image processing.

(1) *An overview* In recent years, many algorithms for image classification have been proposed. Some algorithms can be used out-of-the-box for chart-type classification, and some need to be specially adapted. As seen in Table 1, most authors are focusing on classifying a few basic chart types. Some authors extend their work by classifying less common chart types such as bubble charts, flowcharts, heat maps, treemaps, and sunburst diagrams. Figureseer [23], DocFigure [24], and others [19, 25–29]. DocFigure [24] is a scientific paper that uses the largest number of chart types (28) and reports classification results by each type. Zhou and Tan [30, 31] wrote the earliest works on chart-type classification. According to our knowledge, these scientific papers are the first ones that represent one of the most popular and later used processes for chart-type classification presented in Fig. 4. The concept from Fig. 4 is used with all key approaches regardless of whether it is a custom algorithm, model-based approach, support vector machines (SVMs), or neural networks. As seen in Fig. 4, some blocks are created with a dashed line

Table 1 A summary of existing scientific papers in analyzing chart text processing

Ref.	Year	Key approach				Achieved results [%]											Image processing techniques									Dataset Sum of images															
		CNN	SVM	Model-based	Custom alg.	Other	Area	Bar	Line	Map	Pareto	Pie	Radar	Scatter	Table	Venn	Average	img. pre-processing	Feature extraction	Pattern processing	img. segmentation	T&GS	T&GC	Edge detection	Vectorization		Binarization	img. histogram	NLP	OCR											
[12]	2000																																								20
[11]	2000									85																															35
[13]	2001								73																																55
[14]	2001								85	96																															1290
[15]	2001								84	96																															1190
[18]	2005																																							53	
[3]	2005																																							-	
[23]	2007								95	4																														210	
[25]	2007								84	90	76																														653
[26]	2008								80	87																															75
[27]	2008								94	85																															200
[33]	2011								88	78	73	84	85	79	88	79	86	75																							2601
[34]	2011								96	96	100																														980
[35]	2011								93																															47	
[36]	2011								96	96	100																														980
[37]	2011								96	96	100																														980
[39]	2012																																							155	
[40]	2012								97	98																														600	
[41]	2012																																							300	
[44]	2014																																							722	
[46]	2015								74	68																														5000	
[48]	2015								85																															478	
[49]	2015								66	64	68	67	71	70	63	72	68																							1130	
[50]	2015								39	48	52	51	42	44	40	46	42																							265	
[52]	2016								85																															60,000	
[54]	2017								87																															213	
[5]	2017								98																															5125	
[5]	2017								95	97	94	96	89	98	93	92	98	91																						5659	
[56]	2017								67	93	78	88	85	92	86	86	94	67																						3377	
[57]	2017								92	96	92	99																													4837
[58]	2017																																							400	
[59]	2017								91																															33,778	
[60]	2018																																							50,442	
[61]	2018								90	90																														11,174	
[62]	2018																																							17,154	
[63]	2018								64	99	80																													30,000	
[64]	2018								100																															755	
[65]	2018																																							2398	
[67]	2019								96	98	99	97	100	96	94	98	92	97																						33,000	
[68]	2019								94																															1200	
[69]	2019																																							1592	
[70]	2019								100	97																														10,00,000	
[71]	2019																																							3002	
[6]	2019								77	70	87	87	93	100	93	40	100	70																						4073	
[72]	2020								88	97	91	97																												3002	
[4]	2020								93	90	87	80	83	100	87	87	83	100																						4249	
[73]	2020								59	86	88																													3000	
[74]	2020								53	50	38																													1920	
[75]	2020																																							4718	
[77]	2020																																							21,099	
[78]	2020								100	96	97																													10,000	
[79]	2020								96	99	100																													1400	
[84]	2021								85																															11,723	
[85]	2021																																							57,097	
[86]	2021																																							5560	
[87]	2021																																								

Achieved maximum: 100 100 100 99 100 100 94 98 100 100

[50-59]% [60-69]% [70-79]% [80-89]% [90-100]% TRUE

and mostly consist of text information extraction. Although it is the best practice, it is unnecessary to use both (textual and graphical) information when deciding on the chart type. Authors and scientific papers that report only on making a decision based on graphical information are Bajić et al. [4, 6], Image & graphic reader [32], Beagle [10], Chart-Text [15], and others [11, 24–28, 33–43].

(2) *Used image processing techniques* The first building block of the process mentioned above is image preprocessing or image preparation and manipulation. The main task of image preprocessing is to prepare the image for future analysis and feature extraction. All listed scientific papers use some image preprocessing. The most basic types of image preprocessing are image resolution normalization, image color space normalization, and image noise reduction. Image resolution normalization is used mostly with approaches that rely on convolutional neural networks (CNNs). The CNNs require large image datasets, and images collected from different sources usually have different resolutions. The neural network model expects an image with a fixed resolution and a fixed color mode; thus, image color space normalization is required. During the image conversion process, image noise may occur. Image noise reduction is removing or fixing unwanted pixels from the image. The type of noise in images is usually salt-and-pepper or amplifier noise. The scientific papers that report the use of this type of image preprocessing are Bajić et al. [4, 6], Reverse-Engineering Visualizations [5], Figureseer [23], Chart Decoder [44], Chart-Text [15], VizBy-Wiki [11], Visualizing for the Non-Visual [16], DocFigure [24], and others [17, 21, 26–29, 35, 37, 39–43, 45–51]. To explain the significance and impact of basic image preprocessing on chart-type classification, Bajić et al. conducted an experiment that consists of four datasets. All datasets share the same images, but each dataset has unique filters applied to its images. The experiment showed that the dataset with the heaviest image preprocessing achieved up to 10% better average classification accuracy than the original dataset [4].

Authors of Reverse-Engineering Visualizations [5], Image & graphic reader [32], ReVision [18], View [52], ChartSense [21], Chart Decoder [44], Chart-Text [15], Visualizing for the Non-Visual [16] and others are using advanced image preprocessing techniques, which enable them to extract various information from the graphic image or to more easily separate text and graphics. Advanced image preprocessing includes binarization, edge detection, and vectorization.

Image binarization is when an image is converted from input color space to a black-and-white image. The process enables the reduction of information contained within the image. This task is used when extracting objects from an image is needed. The authors usually call this process the histogram of grey levels, Reverse-Engineering Visualizations [5], View [52], ChartSense [21], and others [3, 45]. This

process should not be mistaken for the histogram of oriented gradients (HOG), which is also a method for extracting objects from the graphic image but used in conjunction with machine learning [25, 27, 36, 53].

Image edge detection is an important step for image feature detection and extraction. The edge in the image represents a local change of intensity that can occur on the boundary between two different regions. The result of edge detection is an image of objects described by the lines, curves, and corners. The two most popular methods for edge detection in chart images are canny and thinning. The scientific papers and authors that report on using edge detection are Reverse-Engineering Visualizations [5], Zhou and Tan [30, 31], Image & graphic reader [32], ReVision [18], Mishchenko and Vassilieva [54–56], ChartSense [21], Chart Decoder [44], Chart-Text [15], Visualizing for the Non-Visual [16], and others [3, 19, 25, 27, 33, 36, 42, 45, 52].

Image vectorization is a process that can be used after the image edge detection process to convert a raster image to a vector image automatically. It enables the extraction of graphical primitives such as straight lines and arcs. The straight-line vector has three values: start point, endpoint, and line width. The arc vector includes an additional parameter, the arc center. The vectorization process is used by Mishchenko and Vassilieva [54–56] and others [19, 33, 57, 58].

(3) *Used key approaches* Authors have used different approaches to obtain chart-type information throughout the years. One of the most popular approaches is based on image feature extraction followed by model design. In Table 1, this approach is labelled as custom algorithm and is used in Image & graphic reader [32], View [52], Beagle [10], ChartFuse [59], and others [3, 13, 19, 25, 36, 45, 53, 57, 60]. The model-based approach creates a model for each chart type. The drawback of this approach is that the model can only recognize charts with the same features as the model. The usage of the model-based approach is best described by Mishchenko and Vassilieva [55, 56]. All extracted features can be grouped into low level, middle level, and high level. The extraction of low-level features can be paired with the Hough transform as explained by Zhou and Tan [30, 31] or with hidden Markov model as explained by Zhou and Tan [61, 62] for chart-type classification. Until ChartFuse [59] in 2020, this approach was only used on three chart types: line, bar, and high-low-close. An extraction of middle-level features can be used with multiple-instance learning [33] and for classifying images based on the shape with the help of using HOG and scale-invariant feature transform (SIFT) descriptors [25, 36, 38]. The extraction of high-level features is used with the aforementioned model-based approaches. A comparison of existing feature extraction methods is shown in ChartFuse [59]. The extracted features can be handed to SVMs for classification. The basic idea of SVMs is to find a line

(or hyperplane) that will separate the data into two classes. SVMs are used in Reverse-Engineering Visualizations [5], ReVision [18], View [52], ChartFuse [59], and others [13, 25, 34, 38, 53]. With a small image dataset, SVMs can achieve state-of-the-art results in a chart-type classification where classification accuracy can reach up to 97.00%, as stated in View [52].

Recent research uses neural networks, especially CNNs, for chart-type classification and feature extraction. Chagas et al. compared traditional classifiers such as HOG + Naïve Bayes, HOG + K-nearest neighbor, HOG + Random Forest, and HOG + SVM with CNN's VGG-19, Inception-V3, and Resnet-50. The results show that CNNs outperform all traditional classifiers by roughly 20% [27]. Other advantages of using CNNs are that they can be used out-of-the-box, and some are available as pre-trained models. Pre-trained models are already trained on datasets that consist of millions of images (e.g., ImageNet), and to use them, only the final layers need to be changed and retrained, Reverse-Engineering Visualizations [5], Figureseer [23], Chart-Text [15], VizBy-Wiki [11], Visualizing for the Non-Visual [16], DocFigure [24], and others [17, 27, 29, 41, 47]. The drawbacks of using CNNs are the need for computing power, large datasets, and an explanation since the whole process is treated as the "black box." The CNN architectures that authors use in this field are: LeNet (named after Yann LeCun et al. in 1989) [26, 37], AlexNet (named after Alex Krizhevsky et al. in 2012) [23, 41, 44, 50], VGG (named after Visual Geometry Group in 2014) [4, 6, 16, 17, 27, 39, 41, 44, 47, 48, 50, 51], GoogLeNet (named after Google in 2014) [21, 44], Residual neural network (ResNet) [16, 23, 27, 44, 47, 49, 50], Inception (developed by Szegedy et al. in 2014) [11, 15, 27, 47, 50], and Mobile computer vision network (MobileNet) [15, 29, 47]. The comparison of multiple CNN architectures is available in Chart Decoder [44, 47, 50], and it shows that all architectures perform similar up to 5% divergence, depending on the used dataset.

(4) *The research direction* The latest research is based on combining CNNs with SVMs, where CNNs are used for feature extraction, and SVMs are used as classifiers. This approach is documented in VizByWiki [11], Visualizing for the Non-Visual [16], DocFigure, [24], by Kaur and Kiesel [29], and in ChartFuse [59]. The mixed combination achieves state-of-the-art results in chart-type classification. In 2019 and 2020, competitions in automatic chart recognition, ICDAR The Competition on Harvesting Raw Tables (CHART-Infographics), were held, and to our knowledge, that are the only competitions that include tasks such as chart image classification, text detection and recognition, text role classification, axis analysis, legend analysis and data extraction from chart images [63, 64].

(5) *Discussions* The methods in chart-type classification changed over the years. All methods used before 2015 can

be considered traditional methods. Although some methods yielded classification accuracy above 80%, the chart images had to follow predefined rules. The ReVision [18] is the first state-of-the-art paper that introduced multiclass classification using SVM. Before neural network architectures, the features for SVMs or any other custom algorithm were manually extracted. The authors used many different image processing techniques to create the best features for their models. Using CNNs reduced the number of required image processing techniques, increased the diversity in chart images, and increased the need for publicly available datasets. All CNN architectures can achieve classification accuracy above 90% without using any additional algorithms. To achieve an accuracy of 100%, authors combine different approaches which help CNN extract image features and focus on relevant areas of an image. Although CNN achieves state-of-the-art results, the research field still lacks a unified dataset showing which architecture and model perform the best in classification accuracy and feature extraction.

3.2 Chart text processing

Chart image is an image that consists of multiple information. This information can be split into three categories: graphical, textual, and semantical (semantic information will be explained in the last subchapter).

(1) *An overview:* As shown in Table 1, the system or the process does not need to know anything about textual information within an image to decide the chart type. The scientific papers and authors that are using both textual and graphical information are Reverse-Engineering Visualizations [5], ReVision [18], Mishchenko and Vassilieva [54, 55], ChartSense [21], Chart-Text [15], Visualizing for the Non-Visual [16], and DocFigure, [24]. Using both information, authors can achieve significantly better results than the authors using only graphical information. All the listed scientific papers show average classification accuracy or classification accuracy of a specific chart type of at least 90%. Table 2 summarizes all scientific papers that deal with textual information. As evident from Table 2, the same process used to separate graphical information from the image can separate textual information. In other words, the opposite of what was classified as a graphic can be assumed as text. The preprocessing used on the graphic image also benefits the text image, View [52], and [58, 65].

Regarding text processing in chart images, text localization, classification, and recognition should be considered. Most chart types can be sorted into one of two groups: those with a Cartesian coordinate system (e.g., bar, line, scatter) and those without a Cartesian coordinate system (e.g., pie, donut, map). The coordinate lines provide index and scale information of data. In his PhD thesis, W. Huang explained that the chart image should include text blocks like chart title,

Table 2 A summary of existing scientific papers in analyzing chart text processing

Ref.	Year	Key approach					Accuracy [%]			Provided algorithm	Image processing techniques										Dataset Sum of images		
		CNN	SVM	Model-based	Custom alg.	Other	Text localization	Text classification	Text recognition		Img. pre-processing	Feature extraction	Pattern processing	Img. segmentation	T&GS	T&GC	Edge detection	Vectorization	Binarization	Img. histogram		NLP	OCR
[12]	2000					TRUE	-	-	-		TRUE	TRUE	TRUE			TRUE							20
[11]	2000					TRUE	-	-	-		TRUE	TRUE	TRUE										35
[14]	2001					TRUE	-	-	-		TRUE	TRUE	TRUE										1290
[15]	2001					TRUE	-	-	-		TRUE	TRUE	TRUE										1190
[17]	2003			TRUE			-	-	-		TRUE	TRUE	TRUE			TRUE	TRUE						27
[18]	2005					TRUE	-	-	91		TRUE	TRUE	TRUE			TRUE						TRUE	53
[3]	2005					TRUE	-	-	-		TRUE	TRUE	TRUE							TRUE		TRUE	-
[19]	2006					TRUE	-	-	-		TRUE	TRUE	TRUE			TRUE	TRUE						115
[24]	2007					TRUE	-	-	-		TRUE	TRUE	TRUE										200
[26]	2008					TRUE	-	-	-		TRUE	TRUE	TRUE									TRUE	75
[27]	2008					TRUE	-	-	-		TRUE	TRUE	TRUE										200
[28]	2008					TRUE	-	-	-		TRUE	TRUE	TRUE							TRUE	TRUE		-
[33]	2011		TRUE				-	-	-		TRUE	TRUE	TRUE	TRUE								TRUE	2601
[34]	2011			TRUE			-	-	-		TRUE	TRUE	TRUE							TRUE		TRUE	980
[35]	2011					TRUE	-	-	-		TRUE	TRUE	TRUE									TRUE	47
[36]	2011			TRUE			-	-	-		TRUE	TRUE	TRUE									TRUE	980
[37]	2011			TRUE			-	-	-		TRUE	TRUE	TRUE									TRUE	980
[40]	2012					TRUE	-	-	-		TRUE	TRUE	TRUE										600
[41]	2012		TRUE			TRUE	-	-	-		TRUE	TRUE	TRUE			TRUE					TRUE	TRUE	300
[43]	2013					TRUE	-	-	-		TRUE	TRUE	TRUE										200
[45]	2015					TRUE	-	-	-		TRUE	TRUE	TRUE							TRUE		TRUE	18
[48]	2015		TRUE			TRUE	-	-	-		TRUE	TRUE	TRUE			TRUE					TRUE	TRUE	478
[49]	2015					TRUE	-	91	-		TRUE	TRUE	TRUE			TRUE				TRUE		TRUE	1130
[50]	2015		TRUE			TRUE	-	-	-		TRUE	TRUE	TRUE									TRUE	265
[51]	2016					TRUE	-	-	-		TRUE	TRUE	TRUE									TRUE	300
[52]	2016	TRUE					60	-	76		TRUE	TRUE	TRUE									TRUE	60,000
[54]	2017	TRUE				TRUE	-	92	-		TRUE	TRUE	TRUE									TRUE	213
[5]	2017	TRUE	TRUE				88	98	99		TRUE	TRUE	TRUE			TRUE						TRUE	5125
[55]	2017	TRUE					-	-	91		TRUE	TRUE	TRUE									TRUE	25,600
[62]	2018	TRUE					-	82	93		TRUE	TRUE	TRUE			TRUE						TRUE	11,174
[64]	2018	TRUE					94	-	-		TRUE	TRUE	TRUE									TRUE	30,000
[67]	2019	TRUE					98	97	-		TRUE	TRUE	TRUE									TRUE	2398
[71]	2019	TRUE					-	-	-		TRUE	TRUE	TRUE										10,00,000
[72]	2020	TRUE					-	-	-		TRUE	TRUE	TRUE									TRUE	4073
[84]	2021	TRUE					95	-	91		TRUE	TRUE	TRUE			TRUE						TRUE	1400
[87]	2021	TRUE			TRUE		-	-	-		TRUE	TRUE	TRUE			TRUE						TRUE	5560
[88]	2021	TRUE					89	80	-		TRUE	TRUE	TRUE							TRUE		TRUE	30,480
Achieved maximum:							98	98	99														

[50–59]%
[60–69]%
[70–79]%
[80–89]%
[90–100]%
TRUE

x-axis title, x-axis values, y-axis title, y-axis values, legend title, and additional description [57].

(2) *Text localization* For chart text localization, the traditional method provides sufficiently good results regardless of the diversity of chart images. The method uses a binarized image with or without further preprocessing. The key is to detect black pixels in the image, as they are candidates for text characters or image noise. The concentration of black pixels in the surrounding space is calculated to distinguish noise from text character candidates. The algorithm looks for the next candidate characters in certain directions when the first character is located. A word can be detected with the amount of white space (white pixels) between characters. Since the white space between words is greater than the white space between characters, a sentence can be detected. The same logic can be applied to multiline strings. Once all candidate words are detected, the bounding box is drawn to indicate a text box for further processing. This method or similar method that works on pixel-level is used in Reverse-Engineering Visualizations [5], Zhou and Tan [30, 31, 61, 62], Mishchenko and Vassilieva [54–56], ReVision [18], Gao et al. [52, 65], Figureseer [23], Chart Decoder [44], and others [3, 13, 19, 36, 45, 53, 57, 58, 60, 66–69]. Reported text localization F1-scores are in Figureseer [23] is 60.30% and 80.00–88.00% in Reverse-Engineering Visualizations [5].

(3) *Text classification* Text classification is less commonly used. Like text localization, text classification also has a traditional method. This method uses certain assumptions about localized text and often draws parallels with a previously extracted graphic image. Higher black pixel density (bigger font) and a position closer to the top edge can be assumed to determine that something is a chart title. The axis values usually consist of numbers closer to the vertical or horizontal black line. The x-axis label is below the x-axis values or close to the bottom edge, and the y-axis label has a tall and narrow bounding box. The legend is usually somewhere in the corner and has multiline words. Under these assumptions, Choudhury et al. [36] and Al-Zaidy and Giles [45] achieved impressive results in text classification accuracy. Text classification is also used in Reverse-Engineering Visualizations [5], Chart Decoder [44], Visualizing for the Non-Visual [16, 48, 57, 70].

(4) *Text recognition* After the text has been localized and classified, it must be processed by a text recognition engine or OCR. The accuracy of OCR greatly depends on the quality of the image. Chart images cannot be compared with documents or natural images since they contain mixed content [54]. They have strings of different rotation (horizontal, vertical), sizes (often tiny), font styles, and various special characters. Since this is a problem studied on a global level, most of the authors choose to use out-of-the-box solutions, such as Microsoft OCR used in Figureseer [23], Amazon's Rekognition [51], or Google's open-source Tesseract OCR engine

used in Reverse-Engineering Visualizations [5], ReVision [18], Mishchenko and Vassilieva [54, 55], Chart Decoder [44], Chart-Text [15], Scatteract [53, 70, 71]. Figureseer [23] noted that Microsoft OCR achieves an overall accuracy of 75.60%. The Tesseract OCR can achieve overall accuracy from 90 to 99%, as indicated in Reverse-Engineering Visualizations [5], Chart Decoder [44], and Scatteract [71].

(5) *The research direction* The latest scientific papers use Darknet or PixelLink to predict text pixels. With Darknet, used in Reverse-Engineering Visualizations [5], the average text localization accuracy F1-score is 80–88%. PixelLink in Visualizing for the Non-Visual [16] achieves an average text localization accuracy F1-score of 77.10–97.50%. For text localization, it is also possible to use Faster region proposal convolutional neural network (R-CNN) [15, 17, 70], which achieves state-of-the-art results in Chart-Text [15] and [70].

Text classification can also be done using SVMs. Reverse-Engineering Visualizations [5], Visualizing for the Non-Visual [16], and Kaur Kiesel [29] achieve an average text classification accuracy of 95.00–100.00%. Kaur and Kiesel [29] is the only author that uses chart caption information in the classification process.

(6) *Discussion* Text localization is finding text areas in chart images and isolating them, text classification is an association of text with graphics, and text recognition is a process of turning words on images into machine-encoded text. The advancement of machine learning increased the object detection accuracy of images. Today's state-of-the-art model for object detection is Faster R-CNN, which is used in many computer vision tasks. The Faster R-CNN consists of a feature extractor, region proposal network, classifier, and a bounding box regressor. The network accepts an input image (which can be unprocessed) and returns bounding boxes of the detected objects. Object detection accuracy is above 90% for textual and graphical elements, and real-world chart images cause a reduction in performance. Text classification is the most challenging field in chart text processing. Authors use structural and geometric information of detected objects or bag-of-words (certain words are expected in certain places). Both approaches need to be manually adapted for each new chart type. Although text recognition can be a challenging task, authors are mainly using publicly available OCR models on text objects and rotate them in four different directions (0°, 90°, 180°, and 270°) to encode text. Despite the remarkable results in chart text processing, real-world charts and charts other than a bar, line, and pie remain a problem. An image caption is the least explored area that can hold vital information for further steps.

3.3 Chart data extraction

Graphical and textual information can be processed separately, or as previously shown, one type of chart image processing can be omitted.

(1) *An overview* Graphical and textual information is required to obtain higher-level information on chart images. Chart data extraction cannot be achieved if one type of information is missing [57]. The scientific papers that deal with this problem are listed in Table 3. The main goal of chart data extraction is to recover the original data table from which the chart image has been created. The authors use two approaches for chart data extraction: automatic and interactive. The automated approach is a complete system that requires the user only to select an image for the input. This type of approach is in the majority and is well studied. The interactive approach requires actions from the user, allowing users to select the data needed for export (points, lines, bounding boxes). An interactive approach is studied in ChartSense [21] and by Yang et al. [66].

(2) *Used techniques* Chart data extraction supports fewer chart types than chart-type classification. Due to the diversity in chart types, the authors focus chart data extraction on three basic chart types: bar, line, and pie chart. As for the other chart types, scatter plot data extraction is researched in Scatteract [71] and [72], and area and radar chart data extraction are analyzed in ChartSense [21]. The developed algorithms cannot be shared among chart types but need to be specifically designed for each chart type and thus are very restricted, and each chart type must comply with a listed set of rules [18, 44, 57, 69, 73, 74]. All traditional algorithms work on a pixel level; therefore, restrictions exist. Because of this, a different dataset is used for validating the chart data extraction. The algorithms count the pixel distance of detected objects from the x- and y-axis and use appropriate scaling values obtained with text processing. They try to pair labels with detected objects if the values are unavailable. The object's dimensions or presence is detected by pixel color change. This applies to the bar and line charts. Pie charts do not contain an axis, and this approach is not valid for them. The goal is to fit the circle/ellipse inside the pie chart using random sample consensus (RANSAC) regression and sample a random number of pixels. With the change of color between two adjacent pixels, the pie slice can be detected, as seen in ReVision [18], ChartSense [21], and Chart-Text [15]. The pie slices can also be detected by the proportions of each color, which leads to counting pixels between slice boundaries [16, 75].

(3) *The research direction* The latest research uses a single deep learning model for data extraction, combining text detection model, text recognition model, and pairwise matching of components, which creates a bar bounding box. They also propose a recurrent neural network (RNN) model to detect angles in pie charts. This single deep neural network

for training images developed by Liu et al. achieves bar data extraction of 79.40% and pie data extraction of 88.00%, but charts outside the training corpus degrade for 57.50% and 62.30% [17]. A similar method is used by Chen and Zhao, where state-of-the-art results are achieved [72]. The two other types of research focus only on bar chart data extraction [48] and [70]. At the same time, Dadhich et al. cover multiple bar charts such as simple, grouped/clustered, stacked horizontal and vertical orientation, and histograms. To extract bar data, regions of interest are labeled, and all other elements that are not chart objects are removed using image processing techniques. After the canvas with chart objects is extracted, the local geometric descriptor (tensor field computation) is used [48]. Zhou et al. use an encoder–decoder framework [70]. The CNN is an encoder that extracts features from images, and RNN is used as a decoder for processing and generating sequence data. This approach results in bar chart data extraction that ranges from 71 to 91%, depending on the used dataset. Sohn et al. extensively researched the line charts, where a line slope, partition, minimum, maximum, range, and other knowledge can be extracted using CNN [76].

(4) *Discussions* The traditional methods require image processing techniques that depend on pre-defined rules and values, and these methods are limited to well-structured chart images. CNN can improve the extraction of chart images, but the best results are achieved when both traditional and deep learning methods are used. CNN can significantly benefit from image processing techniques, and data extraction accuracy can increase further. Data extraction is an important research field that can help blind people and people with impaired vision. It can also help in VQA systems, image searching, NLG, or any textual description. The limitations still exist, and authors are still trying different key approaches for data extraction. The most significant limitation is the natural diversity of chart images and the deviation of real-world charts from synthetic images. The improvements can still be made in detecting multi-lines, crossing lines, stacked and grouped elements, and multiple colors. As for the reported accuracies, it is required to create unified metrics that enable authors to compare their work. While some authors report tables with the accuracy of extracted regions of interest, many authors report only the average value and exclude descriptive values (legends, axis, title, caption), and others report only data point extraction accuracy.

3.4 Chart description generation

The last building block of Fig. 4 is chart description generation. All previous steps must be done to create a textual description of the chart image.

1) *An Overview*: As this is the most complex research field, the least number of scientific papers are associated with it.

Table 3 A summary of existing scientific papers in analyzing chart data extraction

Ref.	Year	Key approach					Accuracy [%]				Type	Image processing techniques										Dataset Sum of images				
		CNN	SVM	Bayesian network	Custom alg.	Other	Bar	Line	Pie	Scatter		Auto. [A] / Inter. [I]	Provided algorithm	Img. pre-processing	Feature extraction	Pattern processing	Img. segmentation	T&GS	T&GC	Edge detection	Vectorization		Binarization	Img. histogram	NLP	OCR
[16]	2002						-	-	-	-	-															-
[17]	2003						-	-	-	-	-															27
[3]	2005						-	-	-	-	A														-	
[19]	2006						-	-	-	-	I														115	
[20]	2007						-	86	-	-	A														177	
[22]	2007						-	-	-	-	A														110	
[27]	2008						-	-	-	-	-														200	
[28]	2008						-	-	-	-	-														-	
[29]	2010						-	-	-	-	-														215	
[30]	2010						-	-	-	-	-														-	
[33]	2011						79	-	62	-	-														2601	
[32]	2011						-	-	-	-	-														110	
[34]	2011						-	-	67	-	A														980	
[36]	2011						-	-	-	-	A														980	
[41]	2012						-	-	-	-	-														300	
[45]	2015						83	-	-	-	A														18	
[47]	2015						-	-	-	-	A														-	
[48]	2015						-	-	-	-	A														478	
[49]	2015						-	-	-	-	A														1130	
[50]	2015						-	-	-	-	A														265	
[51]	2016						86	-	-	-	A														300	
[52]	2016						-	-	-	-	A														60,000	
[54]	2017						98	-	-	-	A														213	
[55]	2017						-	-	-	84	A														25,600	
[56]	2017						-	-	-	-	I														5659	
[58]	2017						-	-	-	-	-														4837	
[62]	2018						83	-	-	-	A														11,174	
[64]	2018						79	-	79	-	A														30,000	
[66]	2018						-	-	99	-	A														1900	
[67]	2019						99	72	92	-	A														2398	
[71]	2019						79	-	88	-	A														10,00,000	
[80]	2020						47	-	-	-	A														28	
[81]	2020						-	-	-	-	A														3700	
[84]	2021						-	-	-	-	A														1400	
[85]	2021						-	-	-	-	A														11,723	
[87]	2021						72				A														5560	
[88]	2021						85	-	-	-	A														30,480	
[89]	2021						-	-	-	-	A														1,87,059	
[90]	2021						-	-	-	-	A														75,000	
[91]	2021						91	89	-	90	A														14,892	

Achieved maximum: **99 89 99 90**

[50–59]% [60–69]% [70–79]% [80–89]% [90–100]% TRUE

Table 4 A summary of existing scientific papers in analyzing chart description generation

Ref.	Year	Key approach				Chart types			Accuracy of creating corresponding text. description [%]	BLEU score	Short summary [SS] / Intended message [IM]	Image processing techniques										Dataset Sum of images		
		CNN	Bayesian network	Custom alg.	Bar	Line	Pie	Scatter				Img. pre-processing	Feature extraction	Pattern processing	img. segmentation	T&GS	T&GC	Edge detection	Vectorization	Binarization	img. histogram		NLP	OCR
[10]	1996							-	-	IM														-
[21]	2007							-	-	SS														-
[22]	2007							79	-	IM														110
[28]	2008							-	-	IM														-
[29]	2010							-	-	IM														215
[30]	2010							-	-	IM														-
[31]	2010							-	-	SS														-
[32]	2011							79	-	IM														110
[38]	2012							79	-	IM														107
[42]	2013							-	-	SS														347
[51]	2016							83	-	IM														300
[53]	2016							-	-	IM														115
[54]	2017							-	-	IM														213
[64]	2018							79	-	SS														30000
[82]	2020							51	19	SS														8305
[83]	2020							-	-	SS														3000
[84]	2021							-	-	SS														1400
[91]	2021							-	93	SS														10232
[92]	2021							-	41	SS														3000

■ TRUE

Table 4 shows scientific papers whose result is a generation of descriptive text for input chart images. The chart types are further limited here, and the most supported types are the basic bar, line, and scatter chart. Only Chart-Text [15] deals with additional chart types: pie, horizontal bar, vertical bar, stacked horizontal bar, and stacked vertical bar chart.

2) *Used Techniques*: Regarding chart image description, authors are divided between two points of view. One point of view is to create a short summary of chart image, as in Chart-Text [15], Chart-to-Text [77], AutoCaption [78], AutoChart [79], by Ferres et al. [20, 80], Schwartz et al. [81], and others [48, 72], whereas the other point of view is to try to understand the intended message of the chart which is presented by Al-Zaidy et al. [14, 45], Schwartz et al. [82], and others [22, 83–88]. The difference between the two is that chart summary generation contains the same information in the image in terms of data values and labels. The process uses object-oriented data from previous steps and string templates. The intended message is the information humans perceive when they see the chart image for the first time and understand what is presented through the graphical objects. This message highly depends on the design of a chart. As stated in the papers, the chart image consists of three communicative signals. The first signal is the relative effort needed for various perceptual and mental tasks. A second signal in the chart is the salience of the objects. A third signal is locating and isolating code verbs, nouns, and adjectives in the chart’s title.

This raw output data from previous steps are presented to a decision tree, Bayesian network, or any other node-based structure. After the probabilities are calculated, the top-level message is generated using string templates.

The scientific papers differ in the way the information is presented to the user. The automatic way is when the chart image is presented to the system, and the system automatically creates a textual description, Al-Zaidy et al. [14, 45], Chart-Text [15], Chart-to-Text [77], AutoCaption [78], AutoChart [79], and others [48, 72, 81, 84, 86–88]. The interactive way is when the user can explore or navigate through the visualization and ask the system (with input commands) to display only the most relevant information, Ferres et al. [20, 80], and others [22, 83, 85].

3) *The Research Direction*: Chart description generation’s success depends on OCR’s quality and other previously explained methods for chart data extraction. The CNNs have found entry into this research field but have yet to be fully adopted. The latest research uses an NLG model based on NLP. The model uses long- and short-term memory (LSTM) network to generate a description. The authors also compare their model to previous models and show that it achieves a higher Bilingual Evaluation Understudy (BLEU) score. The model improves the quality of the generated description, and it is at the time of writing the model with the highest BLEU score [72]. The BLEU score is used to evaluate the quality of translated text in machine translation. The method evaluates the text by comparing the similarity

between machine-translated and human-translated texts; the higher the similarity, the higher the quality. This score is also reported in Chart-to-Text [77] and AutoChart [79].

4) *Discussions*: Chart description generation is essential in creating accessible charts. The chart data extraction step can create a data table from chart images, but that table is not helpful for blind people and people with impaired vision. Instead, a brief description of the data table can make a difference. The generated description should accurately present the data table and provide basic information about trends and values. The number of challenges that need to be addressed in this step further increases and some of them are the lack of large datasets that consists of pairs of chart images and description text, the requirement for human, the NLG methods lack generality and style, and interactivity that will allow users to explore chart images part by part. The automatic dataset construction, which consists of a chart generation and an analytical description generation, is proposed in AutoChart [79] but is limited to scatter, line, and vertical and horizontal bar charts.

4 Research summary

The highest number of authors deal with chart-type classification problems, while the least number of authors try to generate an appropriate textual description of a presented chart image. CNN's are the most used in chart-type classification, where they achieve state-of-the-art results. Though CNNs are used for chart text processing and data extraction, the highest number of authors uses custom algorithms. The comparable results are achieved using both key approaches, and a fine-tuning of specific approaches makes the paper stand out. The SVMs were utilized before the occurrence of CNNs, and now they are returning since they can be used as classifiers for CNNs output.

Until 2004, authors mainly used custom algorithms or modified existing ones, such as Hough transform. From 2005 until 2015, chart image detection and classification research was the most active, and many authors experimented with different key approaches. This was also when authors attempted to solve multiple problems regarding the chart image in their scientific papers. In 2015, the first scientific paper that used CNN appeared, written by Liu et al. [69]. The author simplified the process of image preparation and manipulation and showed that basic image preprocessing combined with CNN could achieve superior results compared to the traditional approaches. Image processing techniques followed the evolution of key approaches. Until 2004, image preparation and manipulation were mainly done manually for each image or each chart type, which was time-consuming. From 2005 up until 2015, image processing techniques grew in number. However, not every image processing technique was suitable

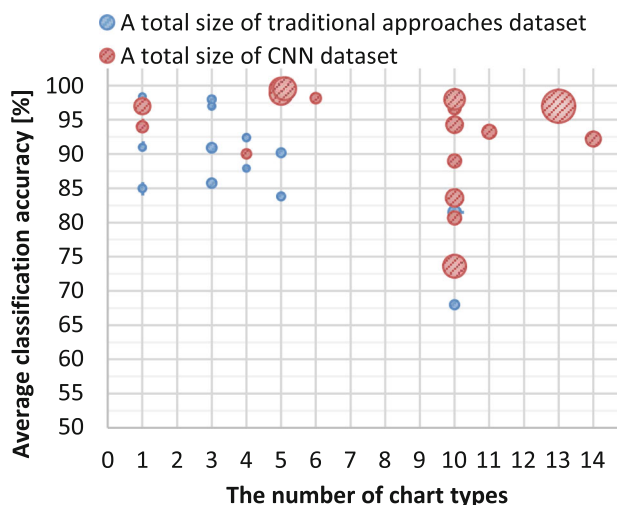


Fig. 5 Bubble chart showing the average classification accuracy in comparison with the number of chart types to be identified. The radius of a bubble presents the total size of the dataset that the authors used

for every key approach, and authors needed to try different "new" techniques. From 2015 to today, authors started using CNNs in chart-type classification, chart text processing, chart data extraction, and chart description generation for the main task and subtasks (e.g., object detection). In any way to be used, CNN achieves competitive results.

Comparing the results in the presented scientific papers is challenging as the authors have no shared variables. Although some authors use out-of-the-box solutions, they had to adjust the input parameters to suit their needs and dataset. The dataset is the most important input variable that enables the comparison of different methods and key approaches. In the upcoming subchapters, the size of the dataset is used in all figures. As all used datasets are qualitatively and quantitatively different, the presented numbers should be considered additional information and not a strong reference point where to achieve X results, with the Y method, the Z size of the dataset is required. In rare cases, authors use the same publicly available datasets and report the same split on the training, testing, and validation, but the images used in each split are still unknown.

1) *Chart-Type Classification*: Fig. 5 shows the summary of chart-type classification from Table 1. We manually removed scientific papers whose results stood out from the group. Scientific papers that were removed are the ones that achieve high average classification accuracy with a small dataset or vice versa. For example, Figureseer [23] is removed. The authors report average classification accuracy from 84 to 86% across seven chart types, using 60,000 images. Other papers report higher average classification accuracy using three to four times fewer data. As a general rule of thumb, we removed papers that achieve average values but use datasets larger than 50,000 images or achieve high accuracy values

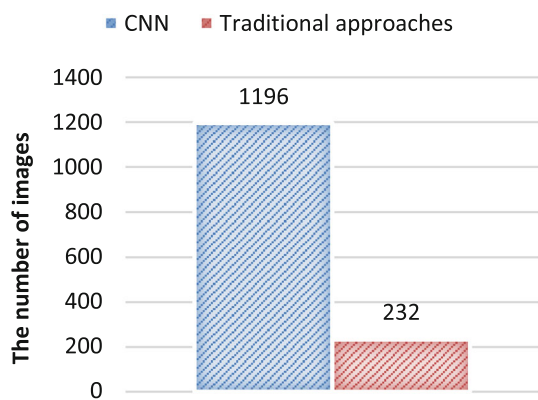


Fig. 6 Bar chart showing average size of dataset per one chart type. The CNNs are using five times more images for chart-type classification

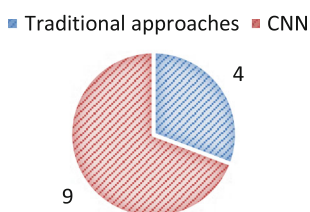


Fig. 7 Pie chart of average number of chart types used for classification with CNN and traditional approaches. The CNNs are used on twice as much chart types

but use less than 200 images. The horizontal axis presents the number of chart types authors report, and the vertical axis presents the average classification accuracy. The bubble’s radius presents a combined total of images used in the process. This dataset includes the sum of training, testing, and validation dataset. The bubbles are presented with blue and orange colors. The same color scheme is applied in Fig. 6 and Fig. 7. The orange color presents scientific papers that use CNNs as a key approach, and the blue color presents traditional (all other) key approaches. When comparing the bubbles, the size of the dataset and the number of chart types should be noted. Figure 6 answers the question about the used dataset size for one chart-type classification, and Fig. 7 shows that CNNs can be used on a much larger number of chart types than traditional approaches. The latter does not require large datasets and is used mostly as binary classifiers (true/false) regardless of the number of chart types. Their complexity increases when introduced with a larger number of chart types. On the other hand, CNNs require large datasets and are used as multiclass classifiers. Most authors are focused on ten or more chart types. The CNNs’ classification results depend on the type of image preprocessing and the dataset. The CNN architecture is also important, but any can be used for any purpose (e.g., feature extraction, object detection, or classification). Authors in [50] evaluated nine neural networks with different training dataset sizes. The

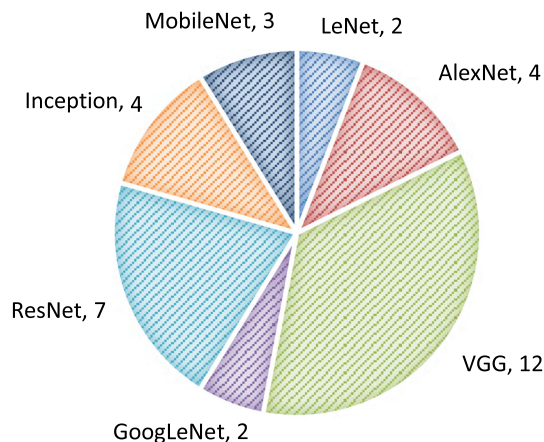


Fig. 8 Pie chart of CNN architectures used in chart-type classification. VGG and ResNet are among the most used ones. The shown architectures include their subtypes, e.g., VGG includes VGG-16 and VGG-19

best-performing CNN in classification accuracy is Xception (The extreme version of Inception). The VGG is the second-best, with a performance reduction of < 1%. Although the VGG might not be the best neural network, it is the most used by the authors. The modular design and the simplicity of modifying it to better suit required needs make it used for any purpose previously explained, as shown in Fig. 8. While most authors are creating their datasets, using the same publicly available dataset for comparison purposes is recommended. The publicly available datasets are the ones of DeepChart [35], Figureseer [23], ReVision [18], and DocFigure [24], as seen in DocFigure [24]. From Fig. 5, we draw two recommendations. To understand in detail how traditional approaches work and how feature extraction works, we recommend Shukla and Samal [19], ReVision [18], Mishchenko and Vassilieva [54], and ChartFuse [59].

ReVision [18] is the most cited scientific paper. Though it was published in 2011, it is still considered state of the art. The authors presented in detail the process of classifying chart images using low-level image features, textual information for improving classification results, bar and pie chart data extraction, and applying perceptually based design principles to redesign data visualizations.

Shukla and Samal [19] described low-level graphic and text processing features. They presented a framework for a quality measure of graphics and the structural elements of data charts. Text processing involves text-parsing, text-stemming, text extraction, and accumulating the keywords in the reference text. The authors also present an algorithm for line processing.

An overview of related work based on the type of extracted features is best described by Mishchenko and Vassilieva [54], who proposed an algorithm for chart-type classification, graphical and textual components, and extraction of data. The algorithm is model-based and uses high-level image features.

The author also proposed an algorithm for text detection regardless of text orientation, size, and style. The experiment procedure is shown for each part of the proposed algorithm, and the results are compared with other approaches.

A novel image classification method is proposed in Chart-Fuse [59]. The authors use a heterogeneous feature extractor, heterogeneity index (HI), fused with a local penta pattern. The HI uses colors, textures, structural layout, and illumination details with extracted features for image classification. With this method, authors can classify charts into subcategories. The comparison is made with existing state-of-the-art feature extractors and different CNN models on three other datasets. The proposed method results in accuracy between 95 and 98%, depending on the used dataset.

Since up-to-date scientific papers use CNNs, good starting points for this type of approach are Amara et al. [26], Chart Decoder [44], and Visualizing for the Non-Visual [16].

Amara et al. use modified LeNet CNN to classify input images into 11 chart types [26]. With the custom dataset's help, the average classification accuracy is 93.27%. The details about the used dataset and average classification accuracy for each type are shown. The model description, experiment setup, and comparison between classic LeNet pre-trained LeNet, and modified LeNet model are also given.

Chart Decoder [44] is a system that can classify input images into five chart types, extract graphical and textual features and generate textual and numeric information. The authors extensively researched chart-type classification, chart text processing, and chart data extraction. Four CNN models are trained from scratch, achieving an average classification accuracy of 99%. The achieved results are compared with ChartSense [21]. For text detection, Darknet is used, and for text recognition Tesseract OCR. Data extraction is supported only for bar charts, and a detailed method overview is given.

To enable visually impaired users to access data visualizations, Choi et al. presented a fully automated system in Visualizing for the Non-Visual [16]. The system can classify input images into 10 data visualization types, detect and classify graphical and textual objects, extract shapes into vector format and extract data from three chart types. The authors achieved average classification accuracy with ResNet CNN of 96.70%. A detailed overview of the dataset is presented, and a comparison is made with Reverse-Engineering Visualizations [5], ChartSense [21], and ReVision [18]. The authors used multiple CNNs for a specific task and achieved state-of-the-art results, as seen in Tables 1, 2 and 3. For data extraction, the authors presented custom algorithms. All system parts are tested separately, and numeric tables of achieved results are noted. This scientific paper is the most complete one.

2) *Chart Text Processing*: Chart text processing is essential in understanding chart images. The text contained within

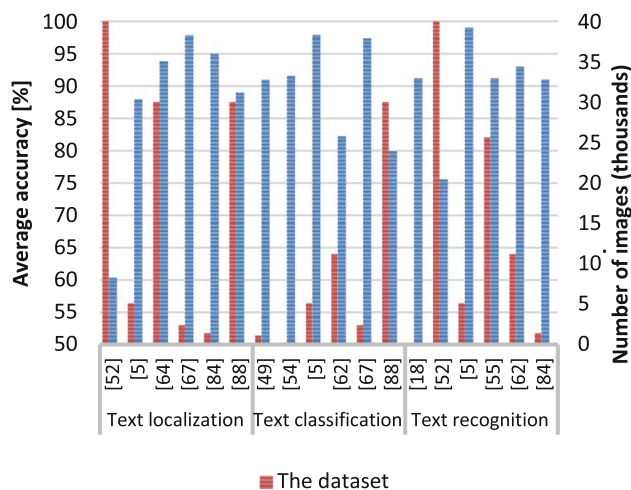


Fig. 9 Summary of chart text processing from Table 2. Reverse-Engineering Visualizations [5] is the only scientific paper that reports all three values. In each category, papers are organized from the oldest to the newest

the image can help in the chart-type classification process. The extracted text is the most used with chart data extraction and summary generation. The text can be of any length, font, size, orientation, or language. It can also be scattered in many different places and thus many possible regions of interest. The overlapping of text, graphics, and color-coded text presents an additional challenge in creating regions of interest. Although this research field is well-studied, only a few authors report accuracy values. The summary of Table 2 is presented in Fig. 9. The figure includes all scientific papers that report three values: text localization, text, classification, and text recognition. Average accuracy is presented on the left side of the figure, and on the right side, there is a number of images in the dataset. Only Reverse-Engineering Visualizations [5] reports all three values. The competitive results are presented in Visualizing for the Non-Visual [16] and Zhou et al. [70]. The author of Reverse-Engineering Visualizations is the same as the most cited scientific paper in chart-type classification, ReVision [18]. The author presents a process for automatically obtaining a visual specification of a chart image. Visual specification includes image dimensions, title, type, labels, and details about the axis. An overview of text processing is given, showing how to obtain textual information from a binarized image. The process includes localization, classification, recognition, and word merging. Darknet is used for text detection, and Tesseract OCR for text recognition. The authors also provide information about chart-type classification, which is done using pre-trained AlexNet CNN. The system is tested on multiple datasets, and ReVision [18] and ChartSense [21] are compared.

In Visualizing for the Non-Visual [16], the authors use pretrained PixelLink based on VGG architecture for text localization. The text is then cropped, and two types of

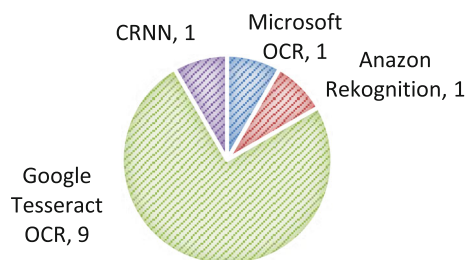


Fig. 10 Pie chart of used OCR engines in chart text recognition. The Google's Tesseract is the most commonly used

OCR models are used, Tesseract and Convolutional Recurrent Neural Network (CRNN). The authors state that CRNN outperforms Tesseract. The exact process is used for text classification as in Reverse-Engineering Visualizations [5]. For evaluation, Intersection-Over-Union (IOU) is used. The IOU calculates the regions of interest from the ground-truth bounding box and predicted bounding box. Although the text localization achieves state-of-the-art results, the error still exists when low-resolution images are used or when the image contains long sentences.

Zhou et al. [70] use Faster R-CNN to detect regions of interest and classify text. The regions are then cropped, and on every cropped image, OCR is applied. The coordinates of cropped regions match the axis, labels, and legend. The authors provide a detailed explanation of the used algorithm. The experiments are done on a synthetic dataset and a real-world dataset. Common Objects in Context (COCO) and Pascal Visual Object Classes (VOC) are used for evaluation. The reported numbers are averaged over ten IOU over all text classes. The results show that the proposed method achieves competitive results on a real-world dataset.

Many authors do not report the type of used OCR engine, and half do not report the achieved accuracy. From Fig. 10, it is evident that the most used OCR engine is Google's Tesseract. Throughout the history, the architecture of the Tesseract changed significantly, and today the architecture is based on neural networks. The engine supports Unicode characters and more than 100 languages and can be used with various output formats; thus, it is still considered one of the best text recognition engines.

3) *Chart Data Extraction*: Chart data extraction is a process whose primary goal is to create an original data table from which the chart was made. Data table creation success depends on chart-type classification and chart text processing. The summary of all scientific papers from Table 3 is presented in Fig. 11. The scientific papers are organized from the oldest (left) to the newest (right). The figure shows all scientific papers that report data extraction for specific chart types. As mentioned, bar, line, pie, and scatter charts are the most processed charts for data extraction. When comparing Table 3 with Fig. 11, it should be noted that the reported

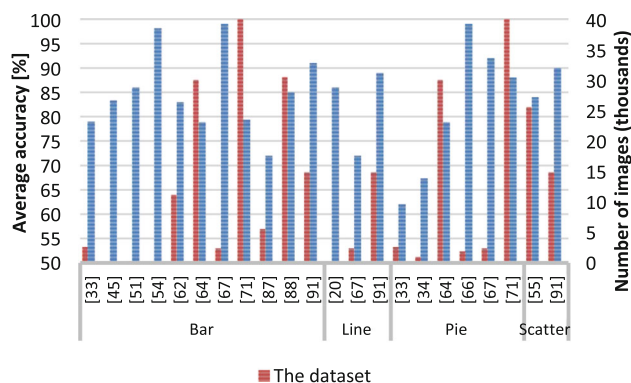


Fig. 11 The achieved average data extraction accuracy for bar chart. The scientific papers are organized from the oldest (left) to the newest (right)

values are only for scientific papers with automatic data extraction from chart images. While interactive models can achieve competitive results, they require the human selection of data points for extraction, which can be prone to errors. Another drawback is the impossibility of processing large datasets, as this process is time-consuming. Visualizing for the Non-Visual [16] reports data extraction for bar, line, and pie chart, where bar data extraction achieves 99%. The bars are detected with the You Only Look Once (YOLO) object detection model. Each bar is processed separately, and the bar's top left and bottom right positions are detected. The axes are processed using CRNN OCR, and the tick span is detected. The height conversion of bars into values is done by calculating the chart's scale as the ratio of the number of pixels and extracted values from the y-axis. The authors also provide an algorithm for line and pie charts and quantitative evaluation of achieved results. Different datasets support the analysis, and results are compared with ReVision [18]. Chen and Zhao separate the process of text extraction and data point extraction for bar, line, and scatter charts [72]. The CornerNet and HourglassNet are used for proposing data points. The probability map is calculated with marked pixels in data point locations. This map is then used in predict module, and thermal, embedded, and offset feature maps are calculated. The maps calculate the top left and bottom right corner points. This method allows the extraction of points regardless of chart type. The evaluation comparison is made with ReVision [18] and ChartSense [21], and higher average accuracy is achieved. The state-of-the-art algorithm that works with 2D and 3D pie charts is presented in automatic data extraction from 2 and 3D pie chart images [75]. The authors use RCNN for chart-type classification, and the detection of pie slices and data extraction is conducted using a series of image processing techniques. First, the image is denoised, converted to greyscale, and binarized. The binarized image consists of separated pie slices whose number of pixels can

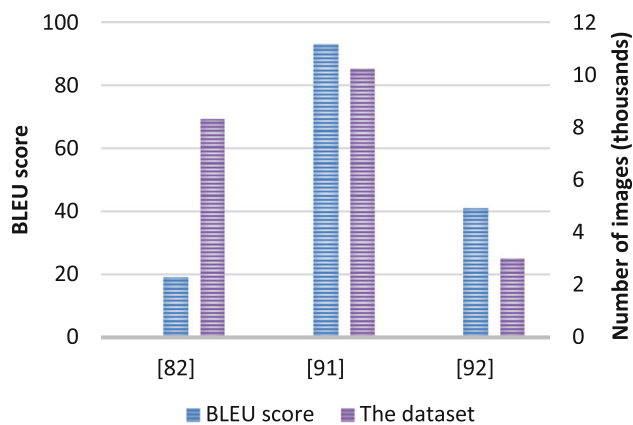


Fig. 12 BLEU score of autogenerated summaries

be calculated. The proposed algorithm extracts data information with an error of 0.14 for 2D pie charts. For 3D pie charts, the error is 0.93, which cannot be compared since no other work focuses on 3D pie charts.

The proposed methods are state-of-the-art for each chart type and achieve the highest data extraction accuracy. The results are for simple synthetic chart images.

4) *Chart Description Generation*: Chart description generation is the research field with the highest number of known challenges, limitations, and least researched. The importance of this field grows with the number of assistive technologies. The vast majority of information is locked in the image, and because of that, the chart description is of great importance for blind people and people with impaired vision. The generated description can differentiate in many ways, but it can be placed in two categories: short summary or intended message. The comparison between categories cannot be made, as both pieces of information differ in the message they convey. Recent research shows the BLEU score used to determine the quality of a short summary, as shown in Fig. 12. The usage of this score is still new in this field. In AutoChart [79], authors also report Recall-Oriented Understudy for Gisting Evaluation (ROUGE) and Bilingual Evaluation Understudy with Representations from Transformers (BLEURT) scores. BLEU is a word-based untrained metric. This metric takes two sentences and assigns a score to the input sentence based on the word overlap between an input and a reference. It is one of the first and the most used metric as it is precision-based. ROUGE is another word-based untrained metric, but unlike BLEU, it is recall-based. The metrics can contain learnable parameters, which can be trained using the input and a reference sentence. BLEURT is a trained metric based on a neural network. It captures non-trivial semantic similarities between two sentences, and it is pre-trained on a public dataset [89]. The scores can range from 0 to 100, where value 0 means no overlap between an input and a reference sentence, and value 100 represents a perfect overlap with a reference sentence.

The higher the value, the higher the quality of the generated text. From Fig. 12, the highest BLEU score is achieved in [91], where the authors state that generated chart descriptions can meet users' needs in terms of quality and correctness. The advantage of using automatic metrics is that they do not depend on human interpretation and judgment. If automated metrics cannot be used, human evaluation should include metrics for naturalness, informativeness, quality, conciseness, coherence, and fluency, as stated in AutoChart [79] and Chart-to-Text [77]. The column "Accuracy of creating corresponding text. description" in Table 4 is evaluated on human subjects and is susceptible to human error.

While the evaluation of automatic chart descriptions improved over time, the evaluation metrics are still lagging. Simple metrics like BLEU and ROUGE are widely used as they do not require any training data; they are consistent and can be very accurate. On the other hand, these metrics give basic and often incomplete knowledge of the presented data, which usually does not capture semantic information like trained metrics. The automatic chart description is limited by chart data extraction and includes simple bar, line, pie, and scatter charts. Of the three mentioned, the bar charts are the most processed ones.

5 Conclusion

To our knowledge, this is a complete review of chart image detection and classification. This extensive research includes 89 scientific papers. We have defined the problem formulation and extracted the aims where this research can be applied. The related works are classified into four categories (subchapters): chart-type classification, chart text processing, chart data extraction, and chart description generation. Each subchapter consists of a table, a brief description of the used process, a research direction, and a discussion. A table enlists the related works and the most important information, including the achieved goal or results. For research direction, we observed the recent related work, which includes the newest methods. These methods are not always state-of-the-art, but sometimes they are a proof of concept which could serve as a foundation for future research. In the discussions, we pointed at the most widely used methods and persisting open challenges. In the end, we have compared related works in the same category and provided additional information for all researchers whose results stand out.

Given all evidence, we conclude that chart-type classification is a very well-researched problem that can achieve almost 100% average accuracy. The highest results are achieved using both graphical and textual information. As for the used key approach, neural networks are the best solution. The architecture of a neural network depends on the author, but the best practice is using the standard models

pre-trained on large datasets. Though other approaches can be considered, Faster R-CNN for object detection and Tesseract OCR for text recognition are commonly used for textual processing. Image filtering and pre-processing can make a notable difference in the model's performance. Comparing results with datasets of different sizes and images can be seen as comparing apples to oranges. While we recommend using publicly available datasets to compare different models and methods, many authors use private datasets, which often lack real-world charts' complexity and diversity. There is also a substantial difference between datasets manually collected from the Internet and those generated using pre-defined parameters. The model trained on real chart images often performs worse on computer-generated chart images and vice versa.

The number of chart types for classification ranges from 1 to 20 or more, which does not represent a problem for neural networks. However, charts are usually limited to bar, line, pie, and scatter charts for chart data extraction and chart description generation. While the achieved results are competitive, the lack of automation in the process and a suitable technique for quality measurement is still present. Compared with chart-type classification, chart data extraction and chart description generation models cannot be used on any chart image. Authors often use different datasets, which are much more limited in size, complexity, and diversity. Chart description generation is the most complex process of all four, which highly depends on the quality of all previous steps. This process enables the creation of accessible chart images, a common end goal for all authors.

Through our research, it can be seen that neural networks have been successfully adapted in the chart-type classification field. With their further development and increased attention of the chart community, we believe that neural networks will become the standard in all areas and give more state-of-the-art results, fully applicable to the real world. In the near future, we expect more studies that will evaluate their results on real-world charts and whose results will equalize with the results of synthetic chart images.

References

- Chen, C., Härdle, W., Unwin, A., Friendly, M.: A brief history of data visualization. In *Handbook of Data Visualization*, pp. 15–56. Springer Handbooks Comp.Statistics. Springer, Berlin, Heidelberg (2008). https://doi.org/10.1007/978-3-540-33037-0_2
- Spence, I.: William playfair and the psychology of graphs. In: *JSM - Proceedings of the American Statistical Association*, pp. 2426–2436 (2006). Accessed 01 May 2020
- Schwartz, S.E., Chester, D., Elzer, S.: Getting Computers to See Information Graphics So Users Do Not Have to, *Foundations of Intelligent Systems. ISMIS 2005. Lecture Notes in Computer Science* (2005), Springer, Berlin, Heidelberg, vol. 3488 LNAI, pp. 660–668 (2005). https://doi.org/10.1007/11425274_68
- Bajić, F., Job, J., Nenadić, K.: Data visualization classification using simple convolutional neural network model. *Int. J. Elect. Comput. Eng. Syst.* (2020)
- Poco, J., Heer, J.: Reverse-engineering visualizations: recovering visual encodings from chart images. *Comput. Graph. Forum* **36**(3), 353–363 (2017). <https://doi.org/10.1111/CGF.13193>
- Bajić, F., Job, J., Nenadić, K.: Chart classification using simplified VGG model. In *Proceedings of the 2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 229–233. ISSN: 2157–8702. <https://doi.org/10.1109/IWSSIP.2019.8787299>
- Liu, Y., Lu, X., Qin, Y., Tang, Z., Xu, J.: Review of chart recognition in document images. *SPIE*, vol. 8654 (2013). <https://doi.org/10.1117/12.2008467>
- Davila, K., Setlur, S., Doermann, D., Kota, B.U., Govindaraju, V.: Chart mining: a survey of methods for automated chart analysis. *Trans. Pattern Anal. Mach. Intell.* (2020). Accessed 30 Aug 2020
- Shahira, K.C., Ljijija, A.: Towards assisting the visually impaired: a review on techniques for decoding the visual data from chart images. *IEEE Access* **9**, 52926–52943 (2021). <https://doi.org/10.1109/ACCESS.2021.3069205>
- Battle, L., Duan, P., Miranda, Z., Mukusheva, D., Chang, R., Stonebraker, M.: Beagle: automated extraction and interpretation of visualizations from the Web. In *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2018-April, pp. 1–8 (2018). Accessed 26 Sept 2021
- Lin, A.Y., Ford, J., Adar, E., Hecht, B.: VizByWiki: mining data visualizations from the web to enrich news articles. In *The Web Conference 2018 - Proceedings of the World Wide Web Conference, WWW 2018*, pp. 873–882 (2018). <https://doi.org/10.1145/3178876.3186135>
- Chen, Z., Cafarella, M., Adar, E.: DiagramFlyer: a search engine for data-driven diagrams. In *WWW 2015 Companion - Proceedings of the 24th International Conference on World Wide Web*, pp. 183–186 (2015). <https://doi.org/10.1145/2740908.2742831>
- Choudhury, S.R., Giles, C.L.: An architecture for information extraction from figures in digital libraries. In *WWW 2015 Companion - Proceedings of the 24th International Conference on World Wide Web*, pp. 667–672 (2015). <https://doi.org/10.1145/2740908.2741712>
- Al-Zaidy, R.A., Choudhury, S.R., Giles, C.L.: Automatic summary generation for scientific data charts. In *Workshops at the Thirtieth AAAI Conference on Artificial Intelligence, 2016*. Accessed 26 Sept 2021
- Balaji, A., Ramanathan, T., Sonathi, V.: Chart-text: a fully automated chart image descriptor (2018). [arXiv:1812.10636](https://arxiv.org/abs/1812.10636). Accessed 26 Sept 2021
- Choi, J., Jung, S., Park, D.G., Choo, J., Elmqvist, N.: Visualizing for the non-visual: Enabling the visually impaired to use visualization. *Comput. Graph. Forum* **38**(3), 249–260 (2019). <https://doi.org/10.1111/CGF.13686>
- Liu, X., Klabjan, D., NBless, P.: Data Extraction from Charts via Single Deep Neural Network. *arXiv preprint* (2019). Accessed 26 Sept 2021
- Savva, M., Kong, N., Chhajta, A., Li, F.F., Agrawala, M., Heer, J.: ReVision: automated classification, analysis and redesign of chart images. In *UIST'11 - Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, pp. 393–402 (2011). <https://doi.org/10.1145/2047196.2047247>
- Shukla, S., Samal, A.: Recognition and quality assessment of data charts in mixed-mode documents. *Int. J. Doc. Anal. Recogn.* **11**(3), 111–126 (2008). <https://doi.org/10.1007/s10032-008-0065-5>
- Leo, F., Gitte, L., Livia, S., Bruce, T.: Evaluating a tool for improving accessibility to charts and graphs. *ACM Trans. Comput.-Human Interact. (TOCHI)* **20**(5), 1–32 (2013). <https://doi.org/10.1145/2533682.2533683>

21. Jung, D., Kim, W., Song, H., Hwang, J., Lee, B., Kim, B.H., Seo, J.: ChartSense: interactive data extraction from chart images. In Conference on Human Factors in Computing Systems - Proceedings, vol. 2017-May, pp. 6706–6717 (2017). <https://doi.org/10.1145/3025453.3025957>
22. Fasciano M., Lapalme, G.: PostGraphe: a system for the generation of statistical graphics and text. In International Natural Language Generation Conference (1996). Accessed 26 Sept 2021
23. Siegel, N., Horvitz, Z., Levin, R., Divvala, S., Farhadi, A.: FigureSeer: parsing result-figures in research papers. Lecture Notes in Computer Science, vol. 9911 LNCS, pp. 664–680 (2016). https://doi.org/10.1007/978-3-319-46478-7_41
24. Jobin, K.V., Mondal, A., Jawahar, C.V.: DocFigure: a dataset for scientific document figure classification. In 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW), pp. 74–79 (2019). <https://doi.org/10.1109/ICDARW.2019.00018>
25. Prasad, V.S.N., Siddiquie, B., Golbeck, J., Davis, L.S.: Classifying computer generated charts. In CBMI'2007 - 2007 International Workshop on Content-Based Multimedia Indexing, Proceedings, pp. 85–92 (2007). <https://doi.org/10.1109/CBMI.2007.385396>
26. Amara, J., Kaur, P., Owonibi, M., Bouaziz, B.: Convolutional neural network based chart image classification.” 25th International Conference in Central Europe on Computer Graphics (2017)
27. Chagas, P., Akiyama, R., Meiguins, A., Santos, C., Saraiva, F., Meiguins, B., Morais, J.: Evaluation of convolutional neural network architectures for chart image classification. In Proceedings of the International Joint Conference on Neural Networks, vol. 2018 (2018). <https://doi.org/10.1109/IJCNN.2018.8489315>
28. Shahira, K.C., Lijiya, A.: Document image classification: towards assisting visually impaired. In IEEE Region 10 Annual International Conference, Proceedings/TENCON, vol. 2019-October, pp. 852–857 (2019). <https://doi.org/10.1109/TENCON.2019.8929594>
29. Kaur, P., Kiesel, D., Combining image and caption analysis for classifying charts in biodiversity texts. VISIGRAPP 2020 - Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, vol. 3, pp. 157–168 (2020). <https://doi.org/10.5220/0008946701570168>
30. Zhou, Y.P., Tan, C.L.: Bar charts recognition using hough based syntactic segmentation. Lecture Notes in Computer Science, pp. 494–497 (2000). https://doi.org/10.1007/3-540-44590-0_45
31. Zhou, Y.P., Tan, C.L.: Hough technique for bar charts detection and recognition in document images. IEEE International Conference on Image Processing 2, 605–608 (2000). <https://doi.org/10.1109/ICIP.2000.899506>
32. Redeke, I.: Image & graphic reader. IEEE International Conference on Image Processing 1, 806–809 (2001). <https://doi.org/10.1109/ICIP.2001.959168>
33. Huang, W., Zong, S., Tan, C.L.: Chart image classification using multiple-instance learning. In Proceedings - IEEE Workshop on Applications of Computer Vision, WACV 2007 (2007). <https://doi.org/10.1109/WACV.2007.17>
34. Karthikeyani, V., Nagarajan, S.: Machine learning classification algorithms to recognize chart types in portable document format (PDF) files. Int. J. Comput. Appl. **39**(2), 1–5 (2012). <https://doi.org/10.5120/4789-6997>
35. Liu, X., Tang, B., Wang, Z., Xu, X., Pu, S., Tao, D., Song, M.: Chart classification by combining deep convolutional networks and deep belief networks. In Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, vol. 2015-November, pp. 801–805 (2015). <https://doi.org/10.1109/ICDAR.2015.7333872>
36. Choudhury, S.R., Wang, S., Mitra, P.: Automated data extraction from scholarly line graphs. GREC 2015 (2015). Accessed 26 Sept 2021
37. Chagas, P., Freitas, A.A., Akiyama, R.D., Miranda, B.: Architecture proposal for data extraction of chart images using convolutional neural network. In Proceedings - 2017 21st International Conference Information Visualisation, iV 2017, pp. 318–323 (2017). <https://doi.org/10.1109/IV.2017.37>
38. Shi, Y., Wei, Y., Wu, T., Liu, Q. Statistical graph classification in intelligent mathematics problem solving system for high school student. ICCSE 2017 - 12th International Conference on Computer Science and Education, pp. 645–650 (2017). <https://doi.org/10.1109/ICCSE.2017.8085572>
39. Kavasidis, I., Palazzo, S., Spampinato, C., Pino, C., Giordano, D., Giuffrida, D., Messina, P.: A Saliency-based Convolutional Neural Network for Table and Chart Detection in Digitized Documents. Lecture Notes in Computer Science, vol. 11752 LNCS, pp. 292–302 (2018). Accessed 26 Sept 2021
40. Gokhan, A.: “DeepGraphNet: grafiklerin sınıflandırılmasında derin öğrenme modelleri”. Avrupa Bilim ve Teknoloji Dergisi, pp. 319–329 (2019). <https://doi.org/10.31590/ejosat.638256>
41. Huang, S.: An Image Classification Tool of Wikimedia Commons. Berlin (2020). Accessed 26 Sept 2021
42. Kosemen, C., Birant, D.: Multi-label classification of line chart images using convolutional neural networks. SN Appl. Sci. **2**(7), 1–20 (2020). <https://doi.org/10.1007/S42452-020-3055-Y>
43. Ishihara, T., Morita, K., Shirai, N.C., Wakabayashi, T., Ohyama, W.: Chart-type classification using convolutional neural network for scholarly figures. Lecture Notes in Computer Science, vol. 12047 LNCS, pp. 252–261 (2020). https://doi.org/10.1007/978-3-030-41299-9_20
44. Dai, W., Wang, M., Niu, Z., Zhang, J.: Chart decoder: Generating textual and numeric information from chart images automatically. J. Vis. Lang. Comput. **48**, 101–109 (2018). <https://doi.org/10.1016/J.JVLC.2018.08.005>
45. Al-Zaidy, R.A., Giles, C.L.: A machine learning approach for semantic structuring of scientific charts in scholarly documents. Twenty-Ninth IAAI Conference (2017)
46. Vougiouklis, P., Carr, L., Simperl, E.: Pie chart or pizza: identifying chart types and their virality on Twitter. In Proceedings of the International AAAI Conference on Web and Social Media, vol. 14, pp. 694–704 (2020). Accessed 26 Sept 2021
47. Araújo, T., Chagas, P., Alves, J., Santos, C., Santos, B.S., Meiguins, B.S.: A real-world approach on the problem of chart recognition using classification, detection and perspective correction. Sensors 2020, vol. 20, no. 16 (2020). <https://doi.org/10.3390/S20164370>
48. Dadhich, K., Daggubati, S., Sreevalsan-Nair, J.: BarChartAnalyzer: digitizing images of bar charts. IMPROVE, pp. 17–28 (2021). <https://doi.org/10.5220/0010408300170028>
49. Ma, W., Zhang, H., Yan, S., Yao, G., Hiang, Y., Li, H., Wu, Y., Jin, L.: Towards an efficient framework for Data Extraction from Chart Images (2021). Accessed 26 Sept 2021
50. Thiyam, J., Singh, S.R., Bora, P.K.: Challenges in chart image classification. In Proceedings of the 21st ACM Symposium on Document Engineering, pp. 1–4 (2021). <https://doi.org/10.1145/3469096.3474931>
51. Rane, C., Subramanya, S., Endluri, D., Wu, J., Giles, C.L.: ChartReader: automatic parsing of bar-plots. Accessed 26 Sept 2021
52. Gao, J., Zhou, Y., Barner, K.E.: View: Visual Information Extraction Widget for improving chart images accessibility. In Proceedings - International Conference on Image Processing, ICIP, pp. 2865–2868 (2012). <https://doi.org/10.1109/ICIP.2012.6467497>
53. Nair, R.R., Sankaran, N., Nwogu, I., Govindaraju, V.: Automated analysis of line plots in documents. In Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, vol. 2015, pp. 796–800 (2015). <https://doi.org/10.1109/ICDAR.2015.7333871>

54. Mishchenko, A., Vassilieva, N.: Chart image understanding and numerical data extraction. In 6th International Conference on Digital Information Management, ICDIM 2011, pp. 115–120 (2011). <https://doi.org/10.1109/ICDIM.2011.6093320>
55. Mishchenko, A., Vassilieva, N.: Model-based recognition and extraction of information from chart images. *J. Multim. Process. Technol.* **2**(2), 76–89 (2011)
56. A. Mishchenko and N. Vassilieva, “Model-based chart image classification,” *Lecture Notes in Computer Science*, vol. 6939 LNCS, no. PART 2, pp. 476–485, 2011. https://doi.org/10.1007/978-3-642-24031-7_48
57. Weihua, H.: *Scientific chart image recognition and interpretation*, Singapore (2008). Accessed 26 Sept 2021
58. Karthikeyani, V., Nagarajan, S.: Scientific chart image property identification by connected component labeling in PDF files. *ICECT 2011–2011 3rd International Conference on Electronics Computer Technology*, vol. 4, pp. 209–212 (2011). <https://doi.org/10.1109/ICECTECH.2011.5941888>
59. Mishra, P., Kumar, S., Chaube, M.K.: ChartFuse: a novel fusion method for chart classification using heterogeneous microstructures. *Multim. Tools Appl.* **80**(7), 10417–10439 (2021). <https://doi.org/10.1007/s11042-020-10186-z>
60. Huang, W., Tan, C.L., Leow, W.K.: Associating text and graphics for scientific chart understanding. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR 2005*, 580–584 (2005). <https://doi.org/10.1109/ICDAR.2005.54>
61. Zhou, Y., Tan, C.L.: Chart analysis and recognition in document images. In *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, vol. 2001-January, pp. 1055–1058 (2001). <https://doi.org/10.1109/ICDAR.2001.953947>
62. Zhou, Y., Zhou, Y., Tan, C.L.: Learning-based scientific chart recognition. In 4th IAPR International Workshop on Graphics Recognition, GREC2001, vol. 4, pp. 482–492 (2001). Accessed 26 Sept 2021
63. Davila, K., Kota, B.U., Setlur, S., Govindaraju, V., Tensmeyer, C., Shekhar, S., Chaudhry, R.: “ICDAR 2019 Competition on Harvesting Raw Tables from Infographics (CHART-Infographics).” In 2019 International Conference on Document Analysis and Recognition (ICDAR), pp. 1594–1599 (2019). Accessed 30 Aug 2020
64. Davila, K., Tensmeyer, C., Shekhar, S., Singh, H., Setlur, S., Govindaraju, V.: ICPR 2020 - Competition on harvesting raw tables from infographics. *Lect. Notes Comput. Sci.* **12668**, 361–380 (2021). https://doi.org/10.1007/978-3-030-68793-9_27
65. Gao, J., Zhou, Y., Sensing, K.B.: Classifying chart images with sparse coding. *Compressive Sensing*, vol. 8365 (2012). Accessed 26 Sept 2021
66. Yang, L., Huang, W., Tan, C.L.: Semi-automatic ground truth generation for chart image recognition. *Lecture Notes in Computer Science*, vol. 3872 LNCS, pp. 324–335 (2006). https://doi.org/10.1007/11669487_29
67. Liu, R., Huang, W., Chew, L.T.: Extraction of vectorized graphical information from scientific chart images. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR 1*, 521–525 (2007). <https://doi.org/10.1109/ICDAR.2007.4378764>
68. Svendsen, J., Albu, A.B.: Document segmentation via oblique cuts. *Document Recognition and Retrieval XX*, vol. 8658 (2013). Accessed 26 Sept 2021
69. Al-Zaidy, A. Rabah, and C. L. Giles, “Automatic extraction of data from bar charts,” *Proceedings of the 8th International Conference on Knowledge Capture, K-CAP 2015*, pp. 1–4, Oct. 2015. <https://doi.org/10.1145/2815833.2816956>
70. Zhou, F., Zhao, Y., Chen, W., Tan, Y., Xu, Y., Chen, Y., Licu, C., Zhao, Y.: Reverse-engineering bar charts using neural networks. *J. Visual.* **24**, 419–435 (2021)
71. M. Cliche, D. Rosenberg, D. Madeka, and C. Yee, “Scatteract: Automated extraction of data from scatter plots,” *Lecture Notes in Computer Science*, vol. 10534 LNAI, pp. 135–150, Apr. 2017. https://doi.org/10.1007/978-3-319-71249-9_9
72. Chen, L., Zhao, K.: An approach for chart description generation in cyber–physical–social system. *Symmetry* **13**(9), 1552 (2021). <https://doi.org/10.3390/SYM13091552>
73. Huang, W., Tan, C.L., Leow, W.K.: Model-based chart image recognition. *Lect. Notes Comput. Sci.* **3088**, 87–99 (2003). https://doi.org/10.1007/978-3-540-25977-0_8
74. Lu, X., Wang, J.Z., Mitra, P., Giles, C.L.: Automatic extraction of data from 2-D plots in documents. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR 1*, 188–192 (2007). <https://doi.org/10.1109/ICDAR.2007.4378701>
75. De, P.: Automatic data extraction from 2D and 3D pie chart images. In *Proceedings of the 8th International Advance Computing Conference, IACC 2018*, pp. 20–25 (2018). <https://doi.org/10.1109/IADCC.2018.8692104>
76. Sohn, C., Choi, H., Kim, K., Park, J., Noh, J.: Line chart understanding with convolutional neural network. *Electronics* **10**(6), 749 (2021). <https://doi.org/10.3390/ELECTRONICS10060749>
77. Obeid, J., Hoque, E.: Chart-to-Text: generating natural language descriptions for charts by adapting the transformer model. *arXiv preprint* (2020). Accessed 26 Sept 2021
78. Liu, C., Xie, L., Han, Y., Wei, A., Yuan, X.: AutoCaption: an approach to generate natural language description from visualization automatically. *IEEE Pacific Visualization Symposium*, vol. 2020, pp. 191–195 (2020). <https://doi.org/10.1109/PACIFICVIS48177.2020.1043>
79. Zhu, J., Ran, J., Lee, R.K., Choo, K., Li, Z.: AutoChart: A dataset for chart-to-text generation task (2021). Accessed 26 Sept 2021
80. Ferres, L., Verkhogliad, P., Lindgaard, G., Boucher, L., Chretien, A., Lachane, M.: Improving accessibility to statistical graphs: The iGraph-lite system, ASSETS’07: *Proceedings of the Ninth International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 67–74 (2007). <https://doi.org/10.1145/1296843.1296857>
81. Demir, S., Schwartz, S., Burns, R., Carberry, S.: What is being measured in an information graphic? In *International Conference on Intelligent Text Processing and Computational Linguistics*, vol. 7816 LNCS, no. PART 1, pp. 501–512 (2013). https://doi.org/10.1007/978-3-642-37247-6_40
82. Elzer, S., Schwartz, E., Carberry, S., Chester, D., Demir, S., Wu, P.: Accessible bar charts for visually impaired users. In *Fourth Annual IASTED Intl. Conf. on Telehealth and Assistive Technologies*, pp. 55–60 (2008)
83. Elzer, S., Schwartz, E., Carberry, S., Chester, D., Demir, S., Wu, P.: A Browser Extension for Providing Visually Impaired Users Access to the Content of Bar Charts on The Web,” *WEBIST*, pp. 59–66 (2007). Accessed 26 Sept 2021
84. Wu, P., Carberry, S., Elzer, S., Chester, D.: Recognizing the intended message of line graphs. *Lecture Notes in Computer Science*, vol. 6170 LNAI, pp. 220–234 (2010). https://doi.org/10.1007/978-3-642-14600-8_21
85. Demir, S., Oliver, D., Schwartz, E., S. Elzer, S. Carberry, and K. F. McCoy, “Interactive SIGHT into information graphics,” *W4A 2010 - International Cross Disciplinary Conference on Web Accessibility Raleigh 2010*, pp. 1–10, 2010. <https://doi.org/10.1145/1805986.1806009>

86. Elzer, S., Carberry, S., Zukerman, I.: The automated understanding of simple bar charts. *Artif. Intell.* **175**(2), 526–555 (2011). <https://doi.org/10.1016/J.ARTINT.2010.10.003>
87. Demir, S., Carberry, S., McCoy, K.F.: Summarizing information graphics textually. *Comput. Linguist.* **38**(3), 527–574 (2012). https://doi.org/10.1162/COLI_A_00091
88. Balawejder, E., Traub, T., Burns, R.: Exploring the automatic recognition of pie chart information messages, ericbalawejder.com. Accessed 26 Sept 2021
89. Sai, A.B., Mohankumar, A.K., Khapra, M.M.: A survey of evaluation metrics used for NLG systems (2020). <https://doi.org/10.1145/0000001.0000001>
90. Telea, A.C., Maccari, A., Claudio Riva: An open toolkit for prototyping reverse engineering visualizations—Eindhoven University of Technology research portal. In *Proceedings of the symposium on Data Visualization*, vol. VisSym'02, pp. 241–249 (2002). Accessed 26 Sept 2021
91. Hamraz, H.: *Classification of chart images*. Lexington (2014). <https://doi.org/10.13140/RG.2.2.32589.23527>
92. Carderas, A., Yuan, Y., Livnat, I., Yanagihara, R., Saul, R., Oca, G., Zheng, K., Browne, A.W.: Automated data extraction of bar chart raster images. arXiv preprint (2020). Accessed 26 Sept 2021
93. Sreevalsan-Nair, J., Dadhich, K., Daggubati, S.C.: Tensor fields for data extraction from chart images: bar charts and scatter plots. arXiv, no. Figure 1, pp. 1–17 (2020). Accessed 26 Sept 2021
94. Huang, D., Wang, J., Wang, G., Lin, C.-Y.: Visual style extraction from chart images for chart restyling. In: *International Association of Pattern Recognition*, pp. 7625–7632 (2021). Accessed 26 Sept 2021

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.