Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7699003

DYNAMIC REPRESENTATION AND SHOWN AS NETWORKS, DRIVEN BY ARTIFICIAL INTELLIGENCE

Sayyed Aziz Ahmed[1], Dr. Manoj Eknath Patil[2]
Research Scholar[1], Research Guide[2]
[1,2]Department of Computer Science & Engineering, Dr.A.P.J.Abdul Kalam University, Indore(M.P), aziz.dafi@gmail.com[1], mepatil@gmail.com[2]

Abstract
The ability to compare and contrast different sets of data visually has a lot of potential as a way to bring people together. It is important to have access to different ways of showing the same or similar information if you want to analyse data at different levels. Approaches that use network-based visualisation to find intrusions use graphs to show information like the source and destination addresses, as well as the port numbers and packets themselves. Graph-based methods of detection can be used to show that someone has broken into a network. This type of figure shows how the formation and growth of networks are fluid and always changing. Even though analysing anomalies in large-scale networks is very important, it might be hard to do because the dynamics are not linear and the graph gets more complicated as the size of the network grows. Using a lot of different kinds of complicated data is another problem that needs to be solved. Dealing with the many different types of data and file formats that come up when working with Big Data can be hard and take a lot of time. Using high-performance computing (HPC) and, more specifically, graphics processing units (GPUs) for Big Data analytics is a great way to speed up scientific computing, network analysis, and network visualisation. This is because GPUs are much better at handling graphics than CPUs. Future research may focus on Big Data analytics for streaming data, Big Data with complex structures, or Big Data with uncertainty.
Keywords: Big Data analytics , Visualization In Networking , data integration .

1. Introduction

Not only do high-quality data mining systems need highly trained data miners, but they also need advanced graphical user interfaces and visualisation tools to work well. As part of the visualisation of data mining, the data, the mining process, and the outcomes of the mining could all be shown. This would happen in the right order. The visualisation tools that are used can have a big effect on how friendly and easy to understand data mining systems are and how beautiful they look. Data mining and data visualisation are two ways to look at large datasets to find insights that weren't clear before. As part of an effort to standardise the different ways that data can be shown around the world, a new and effective K-means clustering algorithm that takes into account many different points of view was made. The strategy always leads to results that fit together in the best way possible. It is easy to parallelize and can run on computers with multiple processor cores because of this. This makes it a great choice for large-scale clustering applications that use visual data. Images, flowcharts, text, and application screenshots are all examples of visually focused HCI patterns that can hold a large amount of unstructured data. Some other examples are. These motifs would be shown in different ways, both in terms of how they looked and how they were shown. Meta-visualization is a subset of data visualisation in which visuals are used to show more information about the analysis technique. The word "meta-visualization" is used to describe this part of data visualisation. You can see things from a different point of view by using either of these two types of meta-visualizations: a) a presentation in chronological order; or b) an overly simple way to show, for example, the connections between multiple different data sets. Several programmes for visualising data have been used to look at data about telecommunications. Clustering, outlier visualisation, linkage visualisation, association visualisation, and OLAP (Online Analytical Processing) visualisation are just some of the visualisation technologies that can be quite useful for analysing telecommunications data. Association visualisation, linkage visualisation, and clustering are also types of visualisation technologies. Some people think that compressed graphs can still keep their topology even though they are smaller. These graphs could make it easier to show huge networks, which would make it much easier to find outliers in the data. Both the live talk and the movies showed that compressed graphs can be used for many things, such as network administration and security. When looking for possible security holes in packets, it's easy for analysts to get caught up in the details. It's possible that they won't be able to see the bigger picture because of this. The Time-based Network Traffic Visualizer (TNV) was made so that data loss would be less likely and so that existing intrusion detection systems could work better. Most methods for visualising intrusion detection focus on drawing attention to things that aren't normal instead of showing actual warnings. The audit data are shown to do this. These strategies are used a lot in different ways to find intrusions and see what they look like. If hosts need help figuring out where invasions are, they can use a visualisation strategy. For this method to work, you need to be familiar with how a certain set of guidelines or suite of programmes usually works. After that, you compare audit data and user profiles. An incursion is when the properties of an attack are different from those of a graph in its normal state. An incursion happens when diagnostic information is collected from a graph in order to assist in the identification of true anomalies. On the other hand, these methods don't do much besides show audit data instead of the real alarms. Because of this, they are useful for finding threats like malware and DDoS attacks that cause traffic. When it comes to attacks that don't cause too much trouble, these kinds of attacks don't give us any useful information.

2. Related Work

P. Cuddihy et al.[1] In two different types of industrial use cases, it has been shown that the time it takes to make a cache goes down at a rate that isn't linear with the number of data elements: I'm putting together information about current parts of a gas turbine by combining data from different sources and (ii) aligning siloed data from the electric grid's transmission and distribution networks to a common model. The Semantics Toolkit, which was made by GE, has a part called the open-source FDC Cache.

Y. Huang, et al.[2] In the next article, which can be found here, the structure and parts of TTIN's system are broken down in more depth. There are both use cases and technical details in the presentation.

O. Neretin et al.[3] This study takes a critical look at the different ways and tools that can be used to find out about vulnerabilities and then use that information to evaluate the security of AI systems. It also gives specific signs of how good this information is, especially how complete and reliable it is.

Y. He, et al[4] Importance This paper is important because it looks at the interdisciplinary approach of using AI technology to study the land use of urban central rail transit station core area and design the technical route. These two things are important for making the best use of urban land near rail stations and for the long-term growth of cities. In particular, this paper looks at how AI technology can be used to study the use of land near urban central rail transit station core areas.

S. Ban et al[5] This study shows a transfer learning strategy that uses spatiotemporal feature analysis and pictures of faces that are available to the public to automatically identify facial paralysis caused by an acute stroke. The goal of this research is to come up with such a label. The whole thing can be broken down into its three parts, which are: 1) a facial detection and alignment network to pull out the main parts of the face; 2) transfer learning with feature extraction networks and a gated recurrent unit; and 3) a classifier that checks for facial paralysis. In the first part of the process, a facial detection and alignment network is used to pull out the main parts of the face.

J. Ali, et al[6] In this paper, we show an SDN control plane architecture with a hierarchical control plane for cross-domain communication. Principal component analysis (PCA) is used in this architecture to reduce the number of dimensions in big data traffic, and a support vector machine (SVM) classifier is used to find DDoS attacks. We'll talk about these two ways in more depth below. PCA filters have a lot of false positives, but SVM has a high level of accuracy and a low number of false positives. As a direct result of this, both the performance of classification and accuracy and the rate of false positives are improved.

M. Barry et al.,[7] Include a lot of different parts of this process, from gathering information to actually putting it into action. Stream2Graph, a method based on streams, is what we recommend for building and updating the knowledge graph all the time. Then, we use an example to show how graph features can be added to online machine learning models in the future. Applications that use AI that is based on graphs will benefit from the solution because it speeds up the process of learning and getting knowledge from huge streams of data.

Xuan Yang et al[8] This study will look at how data is visualised using AI to learn more about how artificial intelligence could be used to improve transportation systems in cities. It does this by using a type of traffic control that makes predictions about traffic based on neural networks. It is an efficient way to dredge that can be done quickly, which helps with problems like the need for new routes and traffic jams. This study starts with a thorough look at AI, traffic control, and data visualisation techniques before moving on to the development of an AI-based data visualisation method that can be used in real-world traffic situations.

3. Proposed methodology.

One of the most basic ways to group different types of data visualisations is by the data that is being shown. Low dimensional data, high dimensional data, temporal data, hierarchical data, and network relational data are all diverse sorts of data. There are five main types of data visualisation: low-dimensional data visualisation, high-dimensional data visualisation, temporal data visualisation, hierarchical data visualisation, and network relational data visualisation.

Low-dimensional data structures include one-dimensional linear data, two-dimensional data with only two attributes, and three-dimensional data with only three attributes. Even though this type of data visualisation is often easy to use, it doesn't give a full picture of what's going on beneath the surface.

A data set is said to have a high dimensionality if it has more than three characteristics that are basically the same. When we look in any direction, we will always see high-dimensional data. Even the most basic things have different qualities, like their price, name, model, colour, and the year they were made. The main goal of high-dimensional imaging technology is to make it easier to study and understand high-dimensional data. This is done by using images as a way to represent data and by adding interactive media.

People often use the term "temporal data" to talk about information about the past, the present, and the future. Image technology makes collecting time data easier and more intuitive for users, which is another reason why time data is so important to people in their everyday lives.

When thinking about hierarchical data, the best comparison is to a tree. With the exception of the root node, every other node has at least one parent. The children of these parents are called "child nodes" (nodes that belong to the parent node). In the process of data visualisation, a method called "tree diagram visualisation" is used because it shows the hierarchical structure of the data in a more realistic way.

A node in a data network can have a relational connection to another node in the network if it is connected to that node in some way, whether directly or indirectly. Because the nodes in the network data structure are not limited by the other nodes that are linked to them, there is no fixed hierarchical structure in the network relationship data, and there may be many ways to get from one node to another. This means that the relationships between nodes will get more complicated as more attributes are added. In the field of data visualisation, a mechanical model for visualising network graphs has been shown to be a useful way to show the links between different sets of data. When we think of artificial intelligence, we no longer think of robot arms and forklifts that drive themselves. "Artificial intelligence" (AI) means using computer programmes to do certain tasks.

To give better service to clients, it's important to know who they are and how they interact with the company.

Companies try to improve their customers' experiences by keeping the promises they make to their customers and meeting or exceeding their high expectations.

In response to the growing number of cyber attacks, the company wants to improve its security protocols and take preventative steps.

The goal is to make people more aware of the brand and more loyal to it by using data analytics to find out what people think of it and then making changes to strategies based on what people say. in order to follow all of the rules and laws that have been set up by the government and the relevant industries. Giving customers experiences that are relevant and interesting to them, no matter where they are in the world, what kind of device they use, or what they want. Visualizing data is an important step in the development of artificial intelligence.

AI algorithms can only come up with accurate results if they are given a lot of data to process. Yet, just as it can be hard for humans to look at data presented in spreadsheets and figure out what's important, it can also be hard and time-consuming for AI systems to draw conclusions from a quick look at the numbers alone. Some of the things that could stop a company from moving forward are problems with data segmentation, overcoming bias, and using out-of-date analytics methods. This group includes things like pattern recognition software that doesn't work well, visuals that don't change, and ways of reporting data that are fundamentally wrong. The first step in the modelling process for machine learning is to make sense of the data. This sets the stage for making relevant models and figuring out what the results mean. It is an important part of the AI stack because it gives a visual summary of the most important parts of the data. It gives the data in easy-to-understand visual formats. This helps artificial intelligence algorithms understand the data before they start modelling. This is the first step in figuring out what the data means.

One of the most important things that data visualisation does is help AI developers figure out how well their algorithms or models work. Artificial intelligence (AI) modellers can use data visualisation to help make sense of the results, which can show a wide range of behaviours. Using synchronised views, which developers can easily add to their workflow, it is possible to look at and compare the results of complex and large-scale deep learning models.

Data visualization makes it easier for AI algorithms to:

Look at the data in formats that are easy on the eyes and make it easier to be precise, accurate, and understand.

Explore how the variables are related to each other and look for oddities, patterns, and general trends in the data.

Find out how the data is really put together, check your assumptions, and make better analysis models.

Data visualisation is another tool that computer programmes that use artificial intelligence can use to:

By asking questions of the data, sharing the answers graphically, and then asking questions of the data again, you can help people find the answers they're looking for or get a sense of what a dataset might contain.

To figure out what people are saying and see how data looks, you can use written descriptions to make drawings that look like they were taken from real life.

To automatically build a core, you can have text or voice generated from visual images or whiteboard designs, or you can give engineers high-level component specs. These two choices are both possible.

Give your company something of value.

As the field of artificial intelligence (AI) grows, more and more work is being done to give AI algorithms better data. In the study of artificial intelligence, the idea of "data visualisation" is given a lot of attention. This is because, just like humans, AI models benefit greatly from being able to understand the relationships between variables rather than just the numbers themselves. Data visualisation and AI are now tied together, and when used together, they can help businesses reach new heights. AI-driven ideas and intelligent visual analytics are two examples of how these two technologies could help do this. When data is shown in a graphical format, it not only helps AI software understand the context of the problem at hand, but it also helps the model find connections, notice oddities, and find trends and patterns. AI software is visual by its very nature. At the same time, AI systems can make interesting and useful visual representations of data. This makes it easier for developers to understand the problems businesses face, find solutions to those problems, improve the efficiency of existing processes, and create brand new products.

4. Data: What Is Visualization Data

This article gives a formal explanation of the idea of showing data in a picture.

Visualization data will be broken down and put in order based on the format of its raw data and the many ways it can be shown. In we divide the different ways to display data into three groups: visuals, programmes, and a hybrid that combines the best parts of the first two. In a few of the systems we looked into, we found that visualisations were not only shown as raw data, but also in well-defined internal representation formats. This was yet another interesting thing we found. Many proposed changes to existing internal representations are meant to make computations easier by getting rid of data that isn't needed. For example, the VQL format only keeps track of how the data is changed and encoded. It does not keep track of how the data is formatted. Because of this, internal representations are almost never shown to the outside world (outputted and shared). We go into a lot of detail about feature presentations, covering everything from feature engineering to feature learning. In the past few years, feature presentations have become more important as a way to make machine learning problems easier to solve from a mathematical and computational point of view. These topics are being talked about because more and more people are interested in using machine learning in the process of data visualisation.

Data

There are many different kinds of things that can be used to hold data for visualisation. Graphics and code are just two examples. The choice of content format has a direct effect on any processing that may be done on visualisation data in the future. This is because each type of content has its own pros and cons.

In this section, we'll talk about the graphic, programmatic, and hybrid formats we found in our corpus. shows that three new research projects have recently come up with hybrid formats that use the power of both graphics and code.

Graphics

Since a data visualisation is just a graphical representation of the data, using visuals as a platform for information lets the data be shown in a more natural and expressive way. In fact, they make up the vast majority of the times they show up in our database. Bitmaps, which are another name for raster images, are often used to create and store visualisations so that they are easier to access and share. But because raster graphics are a stand-alone and lossy representation, the visualisation semantics (such as chart type, visual encoding, and underlying data) are lost in them. Before an automated analysis can be done, reverse engineering is often needed. This is the process of recreating lost data with the help of tools like computer vision and machine learning.

Even so, there is still a problem that needs to be solved about how to reverse engineer a system in a way that is both accurate and reliable. In the end, the fact that raster graphics are lossy makes it harder for machines to understand the images and change them .

Using vector graphics is one way to make sure there is no loss. Unlike raster graphics, vector graphics can be made any size without losing quality. Most of the time, visualisations are saved in a format called Scalable Vector Graphics (SVG). Visual elements can be shown in this format as shapes, like rectangles and text, that can have styles put on them (like locations and fill-color). When working with basic descriptions, for instance, there is no longer a need to use computer vision algorithms to recognise things like texts . With this format, it is also possible to add animation and interactive elements. Visual encoding and the underlying data are two examples of high-level visualisation semantics that have not yet been recovered despite the fact that a lot of work has been done to do so.

Semi-supervised Clustering

This clustering method goes beyond just using the similarity parameter. Instead, it changes or directs the information associated with the domain to get better clustering results. It's possible that there are restrictions between the target variables and the observation variables that act as guides or change the domain for some of the observations.

Support vector machines, which are a type of binary classifier, and decision trees are two other ways to classify data based on its features. There are also some other methods that can be used.

If you're new to categorising big data, you should know the following strategies, as they are some of the most important ones. As we keep saying, you have to have experience in the "real world" in order to understand these kinds of problems. If you haven't already, you should sign up as soon as possible for a course in artificial intelligence or machine learning.

This matrix consists of 4 main elements that show different metrics to count a number of correct and incorrect predictions. Each element has two words either as follows:

If the predicted and truth labels match, then the prediction is said to be correct, but when the predicted and truth labels are mismatched, then the prediction is said to be incorrect. Further, positive and negative represents the predicted labels in the matrix.

Working of proposed model

Semi-supervised learning needs less labelled training data than supervised learning, so it uses pseudo labelling to train the model. In contrast, supervised learning requires that all of the training data be labelled. During the process, it's possible that different neural network models and training methods will be mixed. The following sections go into more detail about how semi-supervised learning works:

Step 1: Because it uses data efficiently at the start, it can train a model with a smaller sample size than supervised learning techniques can. This is a clear advantage over methods of learning without supervision. The model is tweaked until it can come up with results that can be trusted.

Step 2: In the second step, the algorithms use the unlabeled dataset with fake labels, which could lead to wrong results.

Step 3: It has been found that there is a link between the labels in the labelled training data and the fake labels in the pseudo-labels data.

Step 4: There is a connection between the input data used in labelled training data and the input data used in unlabeled training data.

Step 5: Do the model training again, but this time use the data from stages (a) and (b) as input (b). By taking these steps, the model's accuracy will improve and the number of mistakes will go down.

There are four metrics combinations in the confusion matrix, which are as follows:

True Positive: How often does a model correctly tell you that a positive sample is positive?

False Negative: This combination helps show how often a model gets it wrong and mistakes a positive sample for a negative one.

False Positive: This combination shows how often a model gets it wrong and mistakes a negative sample for a positive one.

$$Precision = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)}$$

True Negative: With this combination, you can figure out how often an analytical model correctly recognises a negative sample as just that.?

$$Recall\ = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)}$$

The first thing that has to be done is to look at the final results of an experiment that was carried out in an environment with a lot of data. The purpose of this research was to evaluate a variety of approaches to data classification.

$$\text{Accuracy} = \frac{\textbf{True Positive + True Negative}}{\textbf{(True Positive + False Positive + True Negative + False Negative)}}$$

After constructing five baseline models using 10-fold cross validation as a starting point, we then use an equation to evaluate how well each model performs individually.

$$F1\ Score = \ 2 * \frac{Precision * Recall}{Precision + Recall}$$

After that, the results of the classification tasks performed by the base classifiers are evaluated using TPR, FPR, FNR, precision, accuracy (in percent), F-measure, and MCC.

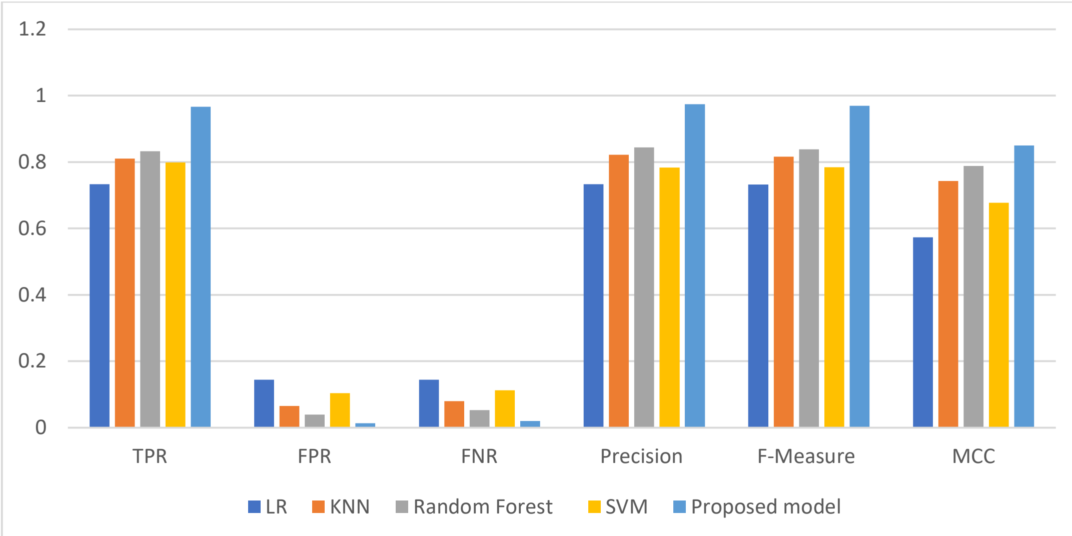

Figure 1: Classification results using base classifiers

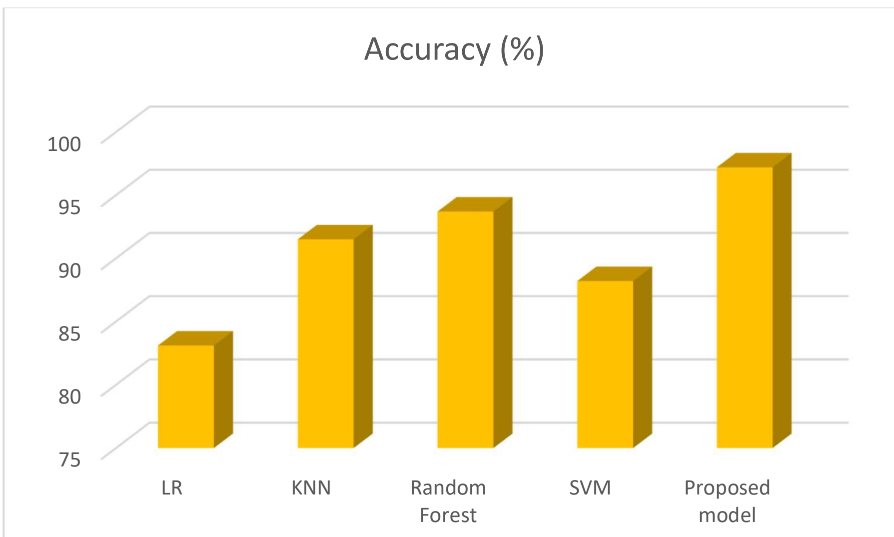

Figure 2: Comparative analysis proposed model and existing model in term of accuracy

5. Conclusion

AI is used by the modern digital enterprise to improve every part of the business, such as reducing the amount of work employees have to do, making processes more accurate and precise, automating boring tasks, reducing the chance of human error, making decisions faster, making the business more flexible and easy to use, and a lot more. Even though artificial intelligence has a lot of potential in today's fast-paced world, it isn't very useful if the data it uses isn't shown in the right way. When data is shown visually, not only can business users easily see the connections between high-dimensional datasets, but so can AI models, which often use a variety of methods to deal with situations. Let's take a look at how the visual representation of data could give AI algorithms the images they need to better understand data and drive more profitable business results.

Reference

[1]. P. Cuddihy, J. W. Williams, V. S. Kumar, K. S. Aggour, A. Crapo and S. Dixit, "FDC Cache: Semantics-driven Federated Caching and Querying for Big Data," 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 2020, pp. 1493-1502, doi: 10.1109/BigData50022.2020.9377876.

[2]. Y. Huang, S. Liu, C. Zhang, X. You and H. Wu, "True-data testbed for 5G/B5G intelligent network," in Intelligent and Converged Networks, vol. 2, no. 2, pp. 133-149, June 2021, doi: 10.23919/ICN.2021.0002.

[3]. O. Neretin and V. Kharchenko, "Model for Describing Processes of AI Systems Vulnerabilities Collection and Analysis using Big Data Tools," 2022 12th International Conference on Dependable Systems, Services and Technologies (DESSERT), Athens, Greece, 2022, pp. 1-5, doi: 10.1109/DESSERT58054.2022.10018811.

[4]. Y. He, H. Yuan, Q. Yao and Z. Wang, "Research and Exploration of Land Use in Core Area of Urban Central Rail Transit Station Based on AI Technology," 2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2021, pp. 645-650, doi: 10.1109/ICAIBD51990.2021.9459099.

[5]. S. Ban, H. S. Nam and E. Park, "Detecting Paralysis of Stroke Symptom in Video: Transfer Learning with Gated Recurrent Unit using Public Big Data of Facial Images," 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 2022, pp. 6587-6589, doi: 10.1109/BigData55660.2022.10020374.

[6]. J. Ali, B. -h. Roh, B. Lee, J. Oh and M. Adil, "A Machine Learning Framework for Prevention of Software-Defined Networking controller from DDoS Attacks and dimensionality reduction of big data," 2020 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), 2020, pp. 515-519, doi: 10.1109/ICTC49870.2020.9289504.

[7]. M. Barry et al., "Stream2Graph: Dynamic Knowledge Graph for Online Learning Applied in Large-scale Network," 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 2022, pp. 2190-2197, doi: 10.1109/BigData55660.2022.10020885.

[8]. Xuan Yang, Zhiyan Pu, Lei Zhang, "Designing Data Visualization Using Artificial Intelligence for Urban Intelligent Transportation Scenarios", *Mobile Information Systems*, vol. 2022, Article ID 3967267, 11 pages, 2022. https://doi.org/10.1155/2022/396726.

[9]. Y. Luo, X. Qin, N. Tang, and G. Li, “Deepeye: Towards automatic data visualization,” in Proc. of the International Conference on Data Engineering (ICDE). IEEE, 2018, pp. 101–112

[10]. Y. Luo, X. Qin, C. Chai, N. Tang, G. Li, and W. Li, “Steerable self- driving data visualization,” IEEE Transactions on Knowledge and Data Engineering, 2020 (Early Access.

[11]. E. Kim and K. F. McCoy, “Multimodal deep learning using im- ages and text for information graphic classification,” in Proc. of the ACM International Conference on Computers and Accessibility (SIGACCESS), 2018, pp. 143–148

[12]. H. Lin, D. Moritz, and J. Heer, “Dziban: Balancing agency & automation in visualization design via anchored recommendations,” in Proc. of

the ACM Conference on Human Factors in Computing Systems (CHI), 2020, pp. 1–12.

[13]. D. Moritz, C. Wang, G. L. Nelson, H. Lin, A. M. Smith, B. Howe, and J. Heer, “Formalizing visualization design knowledge as constraints: Actionable and extensible models in draco,” IEEE Transactions on Visualization and Computer Graphics , vol. 25, no. 1, pp. 438–448, 2018

[14]. J. Choi, S. Jung, D. G. Park, J. Choo, and N. Elmqvist, “Visualizing for the non-visual: Enabling the visually impaired to use visualization,” in Computer Graphics Forum , vol. 38, no. 3. Wiley Online Library, 2019, pp. 249–260

[15]. M. Oppermann, R. Kincaid, and T. Munzner, “Vizcommender: Comput- ing text-based similarity in visualization repositories for content-based recommendations,” IEEE Transactions on Visualization and Computer Graphics, 2020 (Early Access