Future Direction of Information Visualization: Machine Learning & Big Data

line 1: Haley Waddell
line 2: *Computing and Informatics*
(of Affiliation)
line 3: *UNCC*
(of Affiliation)
line 4: Charlotte,NC
line 5: hwaddel2@uncc.edu

***Abstract*- Information visualization is utilized in almost all informative technological interactions. Currently big data is being gathered, quantifiably condensed, and then published for users. The idea is that these visual aids contain well-articulated features so that users may make sense of what is being presented. Accurate graphical perceptions and data analytics rely heavily on the proper formatting of visuals aid aspects so that end users can have strategic insight into the subject at hand-marketing, network mapping and urbanization strategies being just a few topics. As society continues to rely on machine learning to aid in this distribution of insight, it is important to maintain a standard level of construct. This paper will assess the ideologies and methods behind graphical perceptions and data analytics, while also highlighting some of the challenges that exist between merging machine learning and big data in the future realm of computing and informatics.**

Keywords—data analytics, machine learning, big data, information visualization, network mapping, marketing, visual aids

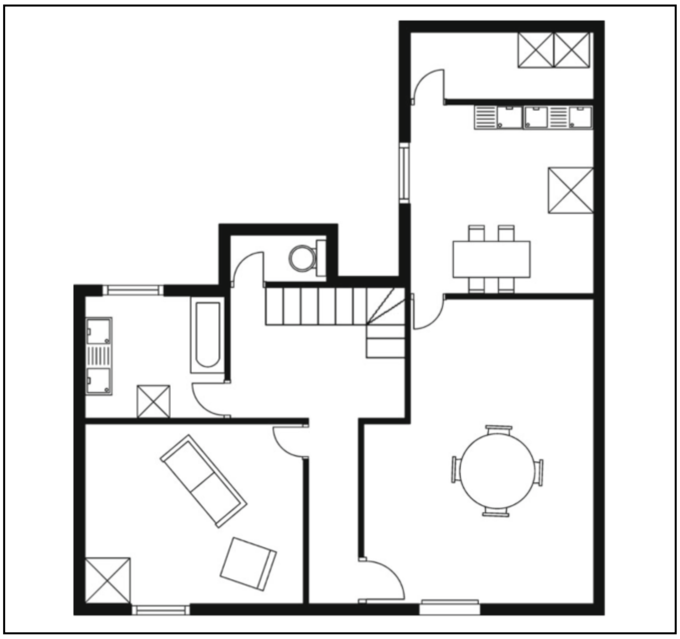
I. Introduction

Imagine a world of black and white packaging- no pops of color, depictive graphics, or branded logos to catch your eye. Information visualization has become a tool that is used not only in attracting attention and drawing in customers, but it is also used to transform the way we interact with comprehensive data. A person's ability to perceive quantitative data is useful in countless fields: medical, meteorology, astronomy and of course business marketing. The goal for the ongoing development in information visualization is to limit data misrepresentation and enhance the practices we currently use. If done correctly there is a promising future for the scalability of data analytics in combination with machine learning. "In visualization, there is increasing research interest in the computational analysis of graphs, charts, and visual encodings, for applications like information extraction and classification, visual question answering ("computer, which category is greater?"), or even design analysis and generation [2]. Often graphs that we see and use as visual aids are amassed through vast amounts of data which has been extrapolated, analyzed, and organized. As the capabilities of computers continue to grow, their role in this assembly of visual aids is going to change as machine learning becomes a more prominent incorporation. Generating useful data sets is an increasingly comprehensive process as we now have more data than ever. We need the power of the computer to firstly help us condense the data and secondly derive meaning through the creation of visual aids. Understanding which features in a graph provide perceptual information is key. Colors, metrics, and properties of visualization need to be considered at all stages of development as data analytics and machine learning combine their forces.

II. Background Research

A. The History of Graphical Perceptions

Visual cues are utilized for the interpretation and recognition of contextual information in the fields of engineering, architect design, maps, and logos to name a few. “To understand the importance of graphics recognition (or any meaningful shapes/parts/regions) has been the subject of several different projects: data acquisition, data preprocessing and data representation/description and recognition/classification” [4]. It seems that the usage of symbols is at the heart of producing a well-received graph. User’s will continue to be forced to develop their list of recognizable symbols as more and more are created. For now, however an example of universal graphical symbol spotting is seen below in this architectural floor plan of Fig.



1:

Fig. 1 Universal symbols used in architecture

B. Assesing the Integrity of Visual Aids

The Simplicity of symbols allowed graphical perceptions to be made in the past, however as we press onward in the development of more intricate graphics with multi layered features it is going to be important to maintain the integrity of visual marks and shapes for both the users and the computer. This is because we need a baseline when analyzing graphics with software. CNN’s or Convolutional Neural Networks are being used to perform many computers vision tasks on images [3]. Think of a sort of data bank that is fed a multitude of images, that then uses a higher -level abstraction methods for symbolic reasoning. Previously this method was highly successful in the natural image object recognition, however it is falling short with current higher-level abstractions (images that require the identification, estimation, and relation of new visual marks) [3]. Studies have been conducted which assess what visual cues aid humans in accurately mapping graphical elements to quantitative variables compared to other software networks, the results are calculated through a midmean logistics absolute error metric (MLAE). Looking Fig. 2 below is an analysis of multiple CNN’s and human perceptual tasks:

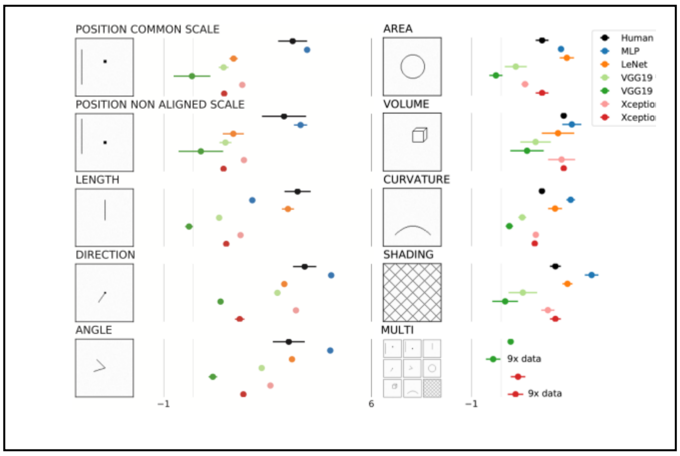


Fig. 2: “Elementary perceptual tasks results for the most complex task parameterization. In each column: *Left:* Example stimuli image. *Right:* MLAE and bootstrapped 95% confidence intervals for different networks. Lower MLAE scores are better. The * indicates fine-tuned ImageNet weights instead of weights trained from scratch. Bottom right shows ‘multi’ VGG19 and Xception networks trained on all perceptual tasks, combined with optional 9× increase of training data” [3].

The takeaway from this study was that humans can generalize the concept of space when analyzing various visual aids, with few examples, whereas the CNN’s need exhaustive training to be able to do this. CNN’S were not found to be a good model for graphical perception at the time of the study in 2019, therefore it warrants warning for “researchers and practitioners to be careful when applying and drawing conclusions from these models in their work" [3].

C. *Some Common Misunderstanding*

Some common misunderstandings in the field of graphical analysis as well as information visualization, is that computers have reached the summit of data analysis and data driven visualization. Let it be known that further innovations are still on the climb and there exists obstacles that must be overcome in this field. One brief example of an obstacle being that patents exists over large data sets: "Only the patented technologies of firms and only the firms that patent their technologies can be positioned in the total technology space and derive the foregoing analytics. Future research should consider the integration of a wider range of data sources to the current methodological framework" [8]. When dealing with users' personal data, which could potentially be used in a harmful manner, most people identify with the statement: Yes, please protect that information with utmost security measure. However, when dealing with data sets that could be useful in the proactive discovery of technologies and new knowledge that effects millions of people should we not consider universal access to data sets? Side stepping around marketing data pools and keeping this suggestion to technologies surrounding profound global topics such as internet infrastructure, automotive GPS, and medical advances- it is ever important to provide larger data sets to broaden future research and development in graphical perceptions, data analysis and machine learning.

III. PROPOSTION

Proposing how to best navigate the realm of machine learning and big data for a successful symbiotic relation in the future is a large assessment. Focus will be brought this proposition through discussion of data-driven visualization methodologies and how machine learning algorithms need to be tactfully leveraged in the field of competitive intelligence innovation.

A. *Leveraging Big Data*

Limitations are always going to exist when dealing with computational developments, partly because technology is ever changing- it's a race with no finish line. A current trend in information visualization, that will likely not see retirement, is the leverage of big data. Industries are finding more and more uses for these data points, with the hope of contributing cutting-edge findings. One point that needs to be proposed when utilizing data in information visualization is the methodology of construction. Quantitative results should encompass descriptive, predictive, and comparative analysis just to name a few. One current study is utilizing the graph theory and network visualization methodology to assess space maps of firms' technology positions. Fig. 3 shows proximity of the technology positions created by InnoGPS.

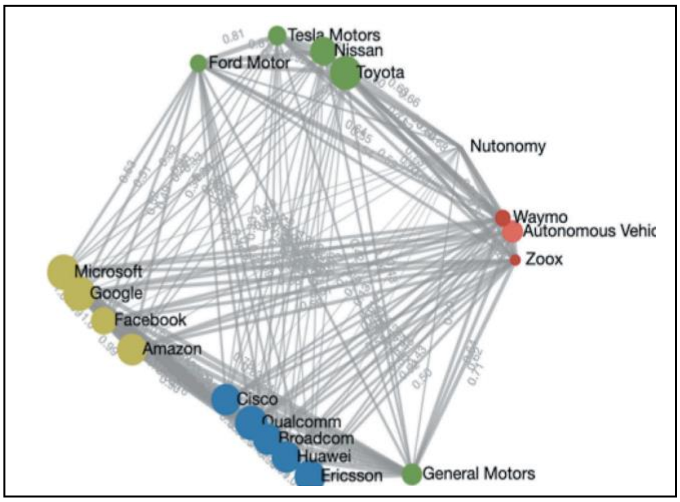


Fig.3:"The network of internet, telecommunication, automotive and autonomous vehicle firms, and autonomous vehicle technology, according to their proximities in the total technology space. The value next to an edge is the inter-firm proximity" [8].

The construction of this overlay follows the methodology of the Quick Technology Intelligence Processes or QTIP. These being: constant database access, analytical software, automated routines, and decision process standardization. Having this all-encompassing design should be the standard in future development of visual data analytics, as being able to set various proximities and select specific features is powerful for visual sense making. Fig. 4 shows the final product page of the interactive visual analytic tool.

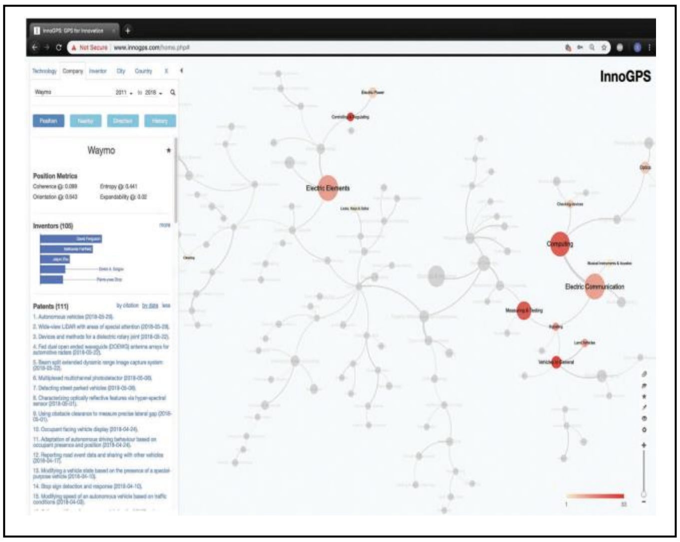


Fig. 4 InnoGPS [8]

B. *Machine Learning Algorthims*

The exploration of combining machine learning and data visualization starts with understanding how humans make sense of visual insights. Fig. 5 jumpstarts this proposition with a lovely depiction of how humans cognitively progress through the sense making process of information.

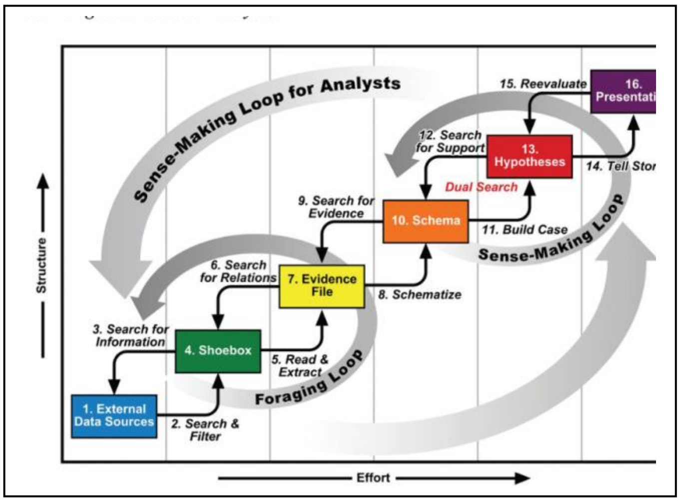


Fig. 5 Sense-Making Loop [10]

Having this foundational understanding of the psychology behind human sense making has been a critical aspect of developing computational analytic processes. Additionally, two scientific methodologies have been coined when it comes to ML.

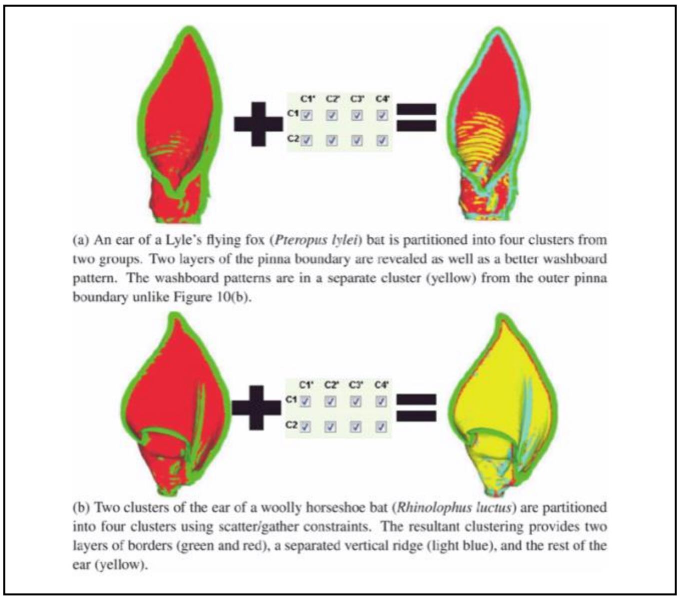
When referencing machine learning algorithms there exist two divisions of learning: supervised and unsupervised. This paper sheds light on the variation of semi-supervised methods as its important to understand how "The supervision is frequently provided by external data in an automated way, but those methods can lead to principled ways of integrating user feedback" [10.] Interactive visualization remains to be untested in the context of Machine Learning. This is because there is much work to be done in the integration of both visualization and analytic systems. Moving forward into development there are two perspectives to take into consideration of this integration: the ML algorithms and the user interaction [10].

C. *Integration*

Different types of ML algorithms are being used when considering visual analysis: dimension, reduction, clustering, classification, and regression/correlation analysis. It is observed that "ML algorithms [used] to tackle these tasks are frequently adopted in visual analytics applications since these analytical tasks often require the joint capabilities of computation and user expertise" [10]. So not only do innovators of ML need to be aware of conventional visualization methods, but they also need to focus on the interaction intent of the user. For a user to be able to directly manipulate the parameters and computational domain measurements is critical as analyst may need to run a subset of data- being able to refine and interact with the data is the goal of this integration. If done correctly this will allow future development of ML and data visualization to become more precise.

IV. Results

The steering of computational resources that are being developed with multi-dimensional and multivariate interfaces is well underway. ML is using the basis of clustering results to now "enable users to interactively work on topic clusters through operations such as splitting, merging and also refining clusters by pointing to example instances or keywords" [10]. Being able to work with an encompassing structure that is rigid yet also malleable is where the success is beginning to amount for users. An example of how ML and interactive features are being used in biological science is seen in Fig. 6



(a) An ear of a Lyle's flying fox (*Pteropus lylei*) bat is partitioned into four clusters from two groups. Two layers of the pinna boundary are revealed as well as a better washboard pattern. The washboard patterns are in a separate cluster (yellow) from the outer pinna boundary unlike Figure 10(b).

(b) Two clusters of the ear of a woolly horseshoe bat (*Rhinolophus luctus*) are partitioned into four clusters using scatter/gather constraints. The resultant clustering provides two layers of borders (green and red), a separated vertical ridge (light blue), and the rest of the ear (yellow).

Fig. 6 ScatterGather is using its feature techniques to gather feedback from analyst in response to algorithms output and refine user-generated constrains to improve clustering [10].

A. *Past, Present and Future*

The development of visuals aids has an understandable progressive history, they started simple as users only had needs for low variable measurements and have begun to be remodeled so that they constitute a multi-faceted collection of features as data variability has increased in the last decade.

Presently, researchers need for visual representations that are capable of interpretation, configuration and interaction are leading the way in design. With the rise data pools current work is being conducted to provide visual aids that are all encompassing when it comes to their features and the ability to systematically categorize theses data features. An example of current work that encompasses this design is a visual approach that helps determine energy usage of various household appliances. Fig. 7 illustrates a combination of various patterns, colors, and shading help to create a visual aid that is useful in determining energy usage year-round [7].

Future aspirations to be able to have computer algorithms specifically tailor

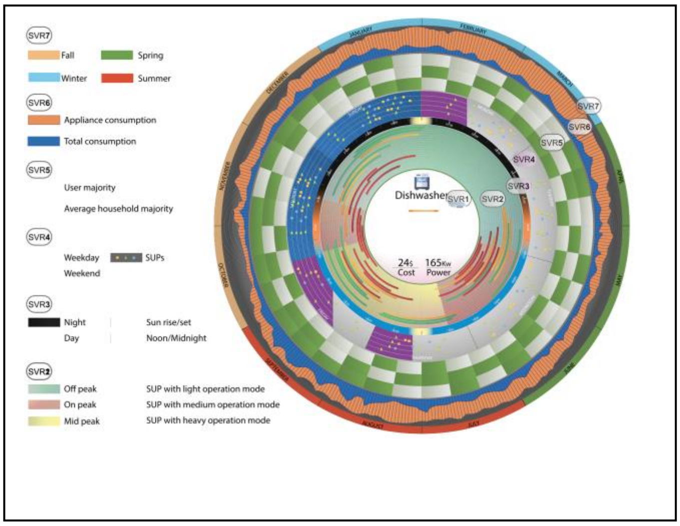


Fig 7. Smart Home Energy Visualizer [7].

visualization as a whole are in the works. If done correctly these results will allow us to start to think about how design tools might account for complex perceptual phenomena in the advancement of visualization [2].

B. *Results of Current Urban Visual Analytics*

Urban visual analytics is tackling urban problems such as traffic predication, air quality forecasting and transit route planning. Analysts in this field are researching how to find feasible approaches for designing visual analytics solutions for the challenges in urban domain areas, visualization techniques, computational methods and the actual visual analytics systems being computed. The researchers are looking to obtain the big picture using primarily mobility data from users. “Given clean mobility data, various visual analytics approaches have been developed for studying mobility patterns, co-occurrences, and mobility semantics. These analyses support in-depth understanding of how citizens move within the urban space” [12]. The road network has been assessed on various levels- macro-, meso- and micro-levels which ultimately assists in determining traffic flow and congestion. Fig. 8 provides some examples of current map visualizations.

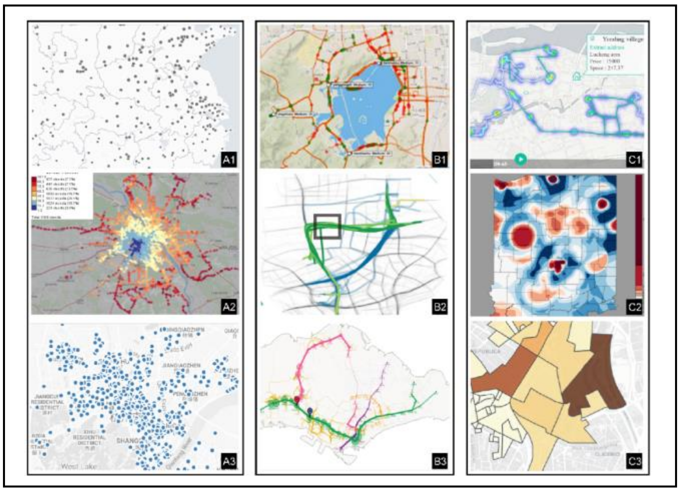


Fig. 8 (A) Examples of dot-on-the-map visualization. (B) Examples of line-on-the-map visualization. (C)Examples of heatmap visualization. [12]

While machine learning is still rarely used in urban visual analytics, the hope is that computational methods will be able to eventually prevail in urban visual analytics [12].

C. *Merger of All Things Good*

Extending existing approaches is what is going to allow for the successful merge of machine learning and visual analytics. What this will entail is focusing on a user-centric system that can rely on computational methods which can provide seamless interaction and be steerable by the user. It is important to note that urban visual analytics is the path to developing smart cities. Collaboration from researchers in the fields of data and visual analytics, ML and graphic design will be paramount in the success merging.

V. Conclusion

We have no time to waste when it comes to implementing better data representation in information visual analytics. The successful execution of this will allow for user interaction with important features and data sets. Leveraging machine learning to help create visual aids and provide new properties and metrics for user to interact with can only help advance technological practices as we know today. The phrase "out with the old and in with new" is an encouraging stance for the powerful developments that will take place in the future of direction of information visualization.

References

[1] El-Khoury Lesueur, F., Bolze, C., Gomajee, R., White, V., & Melchior, M. (2019). Plain tobacco packaging, increased graphic health warnings and adolescents' perceptions and initiation of smoking: DePICT, a french nationwide study.*Tobacco Control, 28* doi: https://doi.org/10.1136/tobaccocontrol-2018-054573

[2] Szafir, Danielle Albers. "Modeling Color Difference for Visualization Design." IEEE Transactions on Visualization and Computer Graphics, vol. 24, no. 1, 2018, pp. 392–401, https://doi.org/10.1109/TVCG.2017.2744359.https://ieeexplore.ieee.org/document/8017604

[3] Haehn, Daniel, et al. "Evaluating 'Graphical Perception' with CNNs." IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 1, 2019, pp. 641–50, https://doi.org/10.1109/TVCG.2018.2865138.

[4] Santosh, K. C. Document Image Analysis Current Trends and Challenges in Graphics Recognition. 1st ed. 2018., Springer Singapore, 2018, https://doi.org/10.1007/978-981-13-2339-3.

[5] Negrón, Adriana Peña Pérez, et al. Virtual Reality Designs. Edited by Adriana Peña Pérez Negrón et al., CRC Press, Taylor and Francis Group, 2020.

[6] Ondov, Brian, et al. "Face to Face: Evaluating Visual Comparison." IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 1, 2019, pp. 861–71, https://doi.org/10.1109/TVCG.2018.2864884.

[7] Jaradat, Abdelkareem, et al. "Smart Home Energy Visualizer: A Fusion of Data Analytics and Information Visualization." Canadian Journal of Electrical and Computer Engineering, vol. 45, no. 1, 2022, pp. 77–87, https://doi.org/10.1109/ICJECE.2021.3134499.

[8] Sarica, Serhad, et al. "Data-Driven Intelligence on Innovation and Competition: Patent Overlay Network Visualization and Analytics." Information Systems Management, vol. 37, no. 3, 2020, pp. 198–212, https://doi.org/10.1080/10580530.2020.1696583.

[9] Garcia, Rafael, et al. "A Task-and-Technique Centered Survey on Visual Analytics for Deep Learning Model Engineering." Computers & Graphics, vol. 77, 2018, pp. 30–49, https://doi.org/10.1016/j.cag.2018.09.018.

[10] Endert, A., et al. "The State of the Art in Integrating Machine Learning into Visual Analytics." Computer

Graphics Forum, vol. 36, no. 8, 2017, pp. 458–86, https://doi.org/10.1111/cgf.13092.

[11] Xu, Haowen, et al. “An Overview of Visualization and Visual Analytics Applications in Water Resources Management.” Environmental Modelling & Software : with Environment Data News, vol. 153, no. C, 2022, p. 105396–, https://doi.org/10.1016/j.envsoft.2022.105396.

[12] Deng, Zikun, et al. “A Survey of Urban Visual Analytics: Advances and Future Directions.” Computational Visual Media (Beijing), vol. 9, no. 1, 2023, pp. 3–39, https://doi.org/10.1007/s41095-022-0275-7.

IEEE conference templates