ChartQC: Question Classification from Human Attention Data on Charts

Takumi Nishiyasu[*] [†]
University of Tokyo
Tokyo, Japan
nisiyasu@iis.u-tokyo.ac.jp

Tobias Kostorz[*]
University of Stuttgart
Stuttgart, Germany
st177908@stud.uni-stuttgart.de

Yao Wang[‡]
University of Stuttgart
Stuttgart, Germany
yao.wang@vis.uni-stuttgart.de

Yoichi Sato
University of Tokyo
Tokyo, Japan
ysato@iis.u-tokyo.ac.jp

Andreas Bulling
University of Stuttgart
Stuttgart, Germany
andreas.bulling@vis.uni-stuttgart.de

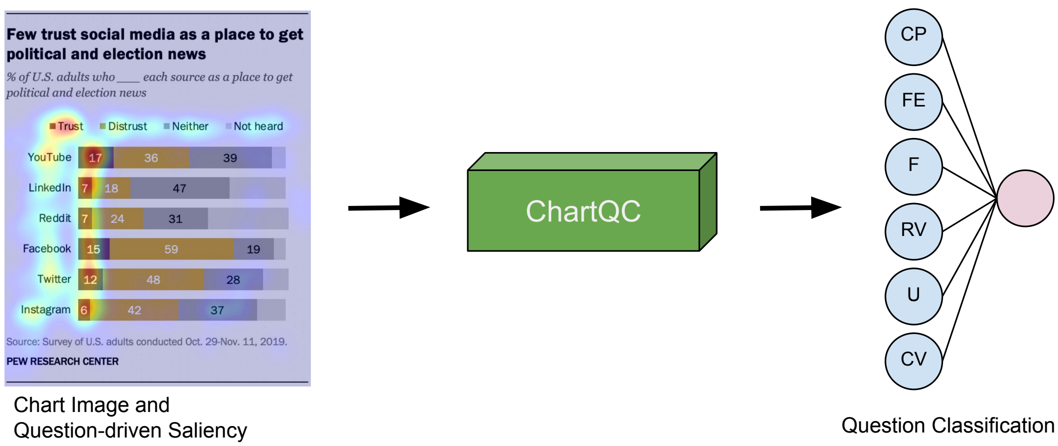

Figure 1: This work presents ChartQC, a novel approach for classifying question types based on human visual attention data on charts. It leverages image features, human attention maps, and metrics to predict visual analytical question types on charts. CP: Comparison, FE: Find Extremum, F: Filter Labels, RV: Retrieve Values, U: Context Understanding, CV: Compute Derived Value.

ABSTRACT

Understanding how humans interact with information visualizations is crucial for improving user experience and designing effective visualization systems. While previous studies have focused on task-agnostic visual attention, the relationship between attention patterns and visual analytical tasks remains underexplored. This paper investigates how attention data on charts can be used to classify question types, providing insights into question-driven gaze behaviors. We propose ChartQC, a question classification model leveraging spatial feature alignment in chart images and visual attention data. By aligning spatial features, our approach strengthens the integration of visual and attentional cues, improving classification accuracy. These findings help deepen the understanding of user perception in charts and provide a basis for future research on interactive visual analysis.

CCS CONCEPTS

• **Human-centered computing → Information visualization**; *HCI theory, concepts and models.*

KEYWORDS

data visualization, intention prediction, attention-based question classification

[*]Both authors contributed equally to this research
[†]A significant part of this work was conducted while at the University of Stuttgart
[‡]Corresponding author

ETRA '25, May 26–29, 2025, Tokyo, Japan
© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *2025 Symposium on Eye Tracking Research and Applications (ETRA '25), May 26–29, 2025, Tokyo, Japan*, https://doi.org/10.1145/3715669.3725883.

ACM Reference Format:
Takumi Nishiyasu, Tobias Kostorz, Yao Wang, Yoichi Sato, and Andreas Bulling. 2025. ChartQC: Question Classification from Human Attention Data on Charts. In *2025 Symposium on Eye Tracking Research and Applications (ETRA '25), May 26–29, 2025, Tokyo, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3715669.3725883

1 INTRODUCTION

Understanding how humans interact with charts (data visualizations) is essential for improving user experience and designing effective visualization systems. Charts serve as crucial tools for data interpretation, allowing users to grasp meaningful insights from complex data quickly [Tufte 1985]. Although a recent pioneering work [Wang et al. 2024b] explored the visual attention mechanisms on visualizations under tasks (questions), a deeper understanding of these interactions can help optimize visualization design, aligning better with user needs. Previous research has primarily focused on task-agnostic visual attention [Borkin et al. 2015; Matzen et al. 2017; Wang et al. 2023], and has neglected how users allocate attention when performing visual analytical tasks such as answering questions about chart data, for example, when finding the maximum value in the chart. Since these tasks require users to locate and interpret specific information from visualizations, it is critical to understand the corresponding attention patterns of questions.

This paper explores the relationship between human visual attention and visual analytical tasks, aiming to predict question types based on attention data. By analyzing how users distribute attention when responding to different questions, we seek to uncover patterns that distinguish question-driven gaze behaviors. Identifying these patterns will contribute to developing intelligent systems that can predict user intent and dynamically adjust visualization layouts to enhance readability and usability. To achieve this, we propose a model that integrates chart images with human attention data, leveraging spatial feature alignment to improve question classification accuracy. Unlike conventional approaches that rely solely on textual or numerical data, our model explicitly incorporates visual and attentional features to enhance classification performance. The model can better capture the relationships between visual elements and user focus areas by aligning spatial features between attention maps and chart structures, improving accuracy in question classification. To validate the effectiveness of our approach, we conduct evaluation experiments on question classification using attention data and analyze the impact of different model components on classification performance. In particular, we perform an ablation study to assess the contribution of various features, such as human attention data and spatial feature fusion, to the overall model accuracy. This analysis provides valuable insights into which aspects of our method are most beneficial for improving classification outcomes.

Our contributions are twofold. First, we propose ChartQC, a transformer-based model that integrates image features, visual attention maps, and human attention metrics to classify question types on charts. Second, we validate the effectiveness of our approach through comprehensive experiments and ablation studies. Our findings contribute to a deeper understanding of user intent in visual analytics and lay the foundation for future research in interactive visualization and gaze-based user modeling.

2 RELATED WORKS

2.1 Task-(Question-)driven Attention on Charts

Research on task-driven visual attention has examined how users observe visualizations across different tasks. Studies have shown that visual attention patterns vary based on cognitive processes, task complexity, and visualization design [Gomez et al. 2016; Huang 2007; Lallé et al. 2016]. For example, Borkin et al. [2015] found that attention falls on visual elements that correlate with memorability. Gomez et al. [2016] demonstrated that gaze behavior is strongly influenced by task demands and visualization structures. Polatsek et al. [2018] found significant differences in gaze behaviors between three visual analysis tasks, attesting that people read completely different regions of charts when handling different tasks. Wang et al. [2024b] collected the first large-scale question-driven attention dataset on charts, and proposed a transformer-based model to predict question-driven saliency. In this work, we leverage the question-driven saliency maps as features for predicting question types.

2.2 Gaze-based User Modeling

Numerous studies in eye-tracking research and cognitive science have shown that human eye movements offer valuable insights into cognitive behavior [Bulling and Roggen 2011; Bulling and Zander 2014], inspiring a growing body of work in gaze-based user modeling [Hu et al. 2021; Pfleging et al. 2016; Wang et al. 2019]. Models incorporating human attention data have been developed to infer user intent, playing a critical role in task recognition and search behavior analysis. These include gaze-based search target prediction models [Barz et al. 2020; Borji et al. 2015; Nishiyasu and Sato 2024; Sattar et al. 2015; Stauden et al. 2018], as well as models for action recognition [Fathi et al. 2012] and intent estimation [Lethaus et al. 2013; Sattar et al. 2020]. In reading behavior modeling, prior research has leveraged eye movements to estimate participants' levels of text comprehension [Ahn et al. 2020] and mind-wandering tendencies [Huang et al. 2019]. In information visualization, a strong correlation between human attention and visualization recallability was found [Wang et al. 2022], and recallability could be predicted from human visual scanpath [Wang et al. 2024a]. Complementing these prior works, we focus on the problem of predicting question types from human visual attention.

3 CHARTQC : QUESTION CLASSIFICATION ON CHARTS

ChartQC models the relationship between visual attention maps, image content, and question representation in the Chart Question Answering (CQA) [Masry et al. 2022] context. Specifically, our method takes visual attention maps and image features as inputs to predict question types using a deep learning framework. The model architecture of ChartQC is depicted in Figure 2.

3.1 Problem Settings

Our approach aims to classify six common visual analytical question types in charts: **Comparison (CP)**: The user has to compare two visual elements with each other. **Find Extremum (FE)**: The user has to find maximum or minimum values. **Retrieve Values (RV)**: The user has to read data values from the visualization. **Filter Labels (F)**: The user has to find labels that match the given conditions. **Context Understanding (U)**: The user has to understand the contexts of charts, such as questions on title, legend, or description. **Compute Derived Value (C)**: The user has to calculate a value (e.g. count, sum, median, ratio).

Our method integrates human attention data to explore how users interact with visual elements in charts and graphs, aiming to improve question type classification. Formally, given an image $I \in \mathbb{R}^{h\times w\times 3}$ representing a chart and human attention data H, which captures user interactions with the visualization, the goal is to predict the question type label y. From H, we compute a visual attention map $A \in \mathbb{R}^{h\times w}$ that aggregates the spatial distribution of fixations, as well as a set of statistical features (metrics) $S \in \mathbb{R}^d$ extracted from human attention data, including fixation density, dispersion, inter-fixation time, and visual entropy. The prediction function is defined as:

$$f : (I, A, S) \to y, \quad y \in \{\mathrm{CP}, \mathrm{FE}, \mathrm{RV}, \mathrm{F}, \mathrm{U}, \mathrm{C}\}$$

Our approach leverages a deep learning framework to integrate the visual features from I, the visual attention map from A, and statistical characteristics from S to predict the question type label on charts. The model is trained using cross-entropy loss.

3.2 Model

The proposed ChartQC model integrates three key components to predict question types from charts. First, we extract image features from the visualization, capturing structural and semantic information. Second, we derive attention-based features from human visual attention maps, leveraging regions of interest that users focus on during question answering. These two feature sets preserve spatial correspondence and are concatenated to maintain alignment between visual content and human attention. Finally, we incorporate statistical metrics computed from human attention data, such as the number of fixations, to provide additional context. The combined features are processed through a neural network to classify the input into question types.

Image Feature Extraction. We employ a pre-trained Vision Transformer (Swin [Liu et al. 2021]) to extract deep visual features from the input chart image $I \in \mathbb{R}^{h\times w\times 3}$. The transformer processes the visualization as a sequence of patches and learns meaningful representations that encapsulate its structural and semantic information. These extracted features contribute to understanding how different chart elements influence question answering.

Visual Saliency Encoding. To enhance the ChartQC with human visual attention, we incorporate visual saliency maps $A \in \mathbb{R}^{h\times w}$, which represent the distribution of user attention within the chart given certain questions. The attention maps can be predicted from charts and questions using the pre-trained weights of VisSalFormer [Wang et al. 2024b], which extracts spatial features of question-driven visual attention.

Human Attention Metrics. Beyond spatial information, we incorporate statistical metrics $S \in \mathbb{R}^d$ derived from human attention data H to enhance the model's predictive ability, which includes the number of fixations, fixation percentage across chart elements (axes, legend, title, label, data) [Shi et al. 2025], fixation density [Wang et al. 2023], inter-fixation time, saliency coverage [Wang et al. 2024b], image Shannon Entropy [Bruce and Tsotsos 2005], chart type, and normalized average center distance of fixations. These metrics provide additional contextual signals, capturing behavioral trends in human interaction with visualizations.

Neural Network Architecture. All extracted features—image features, attention-based features, and statistical metrics—are concatenated and fed into a fully connected neural network. The network consists of multiple dense layers with nonlinear activation functions, followed by a final classification layer that outputs the predicted question type.

4 EXPERIMENT

4.1 Setup

Implementation Details. We used the SalChartQA dataset [Wang et al. 2024b] for training and evaluation. SalChartQA includes 6,000 charts with question-driven human visual attention maps. The input of ChartQC consists of an image I of size $(c, h, w) = (3, 224, 224)$, a visual attention map A of size $(1, h, w) = (1, 224, 224)$ derived from human attention data H, and statistical features S of size d extracted from H (e.g., click density, dispersion, entropy). The model is trained with a batch size of 64 for 75 epochs using the Adam optimizer (weight decay = 1×10^{-4}). The initial learning rate is 2×10^{-5} and decays by a factor of 0.5 every 5 epochs. Cross-entropy loss is used during the training process.

Baselines. To evaluate the performance of our proposed approach, we compared it against several baseline methods. The *Random* classifier randomly assigns a question type to each instance, serving as a chance-level performance. The *Major* classifier predicts the most frequent question type in the dataset. The Support Vector Machine (SVM) is trained on the same handcrafted features derived from human attention data and visualization characteristics. Additionally, we finetuned the VisRecall [Wang et al. 2022] model to the question classification setting. The VisRecall model extracts image features with Xception [Chollet 2017], followed by a global average pooling and a linear layer.

4.2 Quantitative Evaluation

Metrics. We used accuracy and F1 scores as evaluation metrics to assess model performance. Accuracy measures the percentage of correctly classified question types, while F1 score provides a balanced measure of precision and recall, particularly useful for imbalanced question distributions.

Results. We compared the question classification performance of our proposed approach with baseline methods on the SalChartQA dataset. Table 1 summarizes the performance of different models in terms of Accuracy and F1 score. The experimental results demonstrate that our proposed method outperforms all baseline models, achieving the highest Accuracy and F1 score. Similarly, VisRecall has suboptimal performance since the model only takes chart images as input, which lacks question-related representations. In contrast, ChartQC effectively integrates multiple modalities, including chart image features, question-driven saliency maps, and statistical metrics extracted from human attention data. By spatially concatenating saliency maps with chart image features, the model could learn richer representations that better align with human cognitive

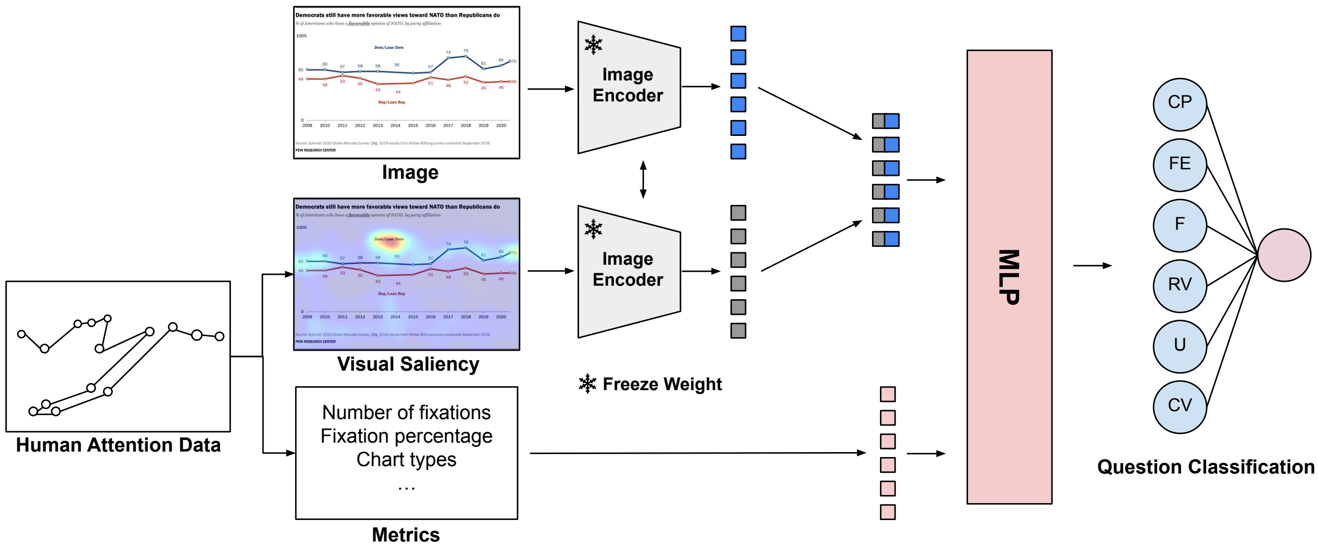

Figure 2: Overview of the model architecture of ChartQC. ChartQC integrates image features extracted from a chart, attention-based features from a visual attention map, and statistical characteristics derived from human attention data. These features are concatenated and processed through a neural network to predict question types.

Table 1: Evaluation of question classification methods on the SalChartQA dataset. The best results are shown in bold.

Model	Accuracy	F1
Random	0.155	0.156
Major	0.172	0.170
SVM	0.425	0.387
VisRecall [Wang et al. 2022]	0.350	0.303
ChartQC (ours)	**0.452**	**0.434**

processes when interacting with information visualizations. The inclusion of statistical features further enhances performance by providing additional context regarding user interaction patterns.

4.3 Ablation Study

To further analyze the contributions of individual components in ChartQC, we conduct an ablation study by evaluating variations of our model on the SalChartQA dataset. Each model isolates specific input modalities to understand their impact on overall performance. Table 2 presents the ablation study results. By taking away visual features and human attention maps, the performance significantly dropped from 0.452 to 0.377. By removing visual features and metrics, the performance dropped to 0.366. The contribution of human attention metrics is also confirmed by removing just the metrics from the full model (0.452 vs. 0.383). In conclusion, the ablation study confirmed the contribution of every input channel of ChartQC – human attention maps, visualizations, and statistical metrics – for effective question type classification on charts.

Table 2: Ablation study for ChartQC. The best results are shown in bold.

Metrics	Saliency	Image	Accuracy	F1
✓	×	×	0.377	0.321
×	✓	×	0.366	0.328
×	✓	✓	0.383	0.382
✓	✓	✓	**0.452**	**0.434**

4.4 Qualitative Evaluation

Figure 3 (top) illustrates cases where the model successfully predicts the correct question type. In the left example, visual attention was mainly focused on the labels of the depicted line graphs. The model recognizes this (e.g. using the saliency distribution metrics) and correctly predicts the 'filter labels' category. In the middle example, multiple labels in the pie chart, especially their corresponding percentages, were at the center of attention. This indicates a value-based task, e.g., calculating a sum, which was predicted as the C label accordingly. In the right example, most of the fixations were focused on exactly one value, which suggests a simple task, for example, just finding and answering a data value.

However, as shown in Figure 3 (bottom), there are instances where the model misclassifies the question type. These errors often occur when the visual attention map is ambiguous or when the question requires a more complex contextual understanding beyond visual features alone. By analyzing these failure cases, we can identify areas for improvement, such as refining feature representations or incorporating additional contextual information to enhance model performance. Some of the incorrect classifications can be attributed to the labels in the dataset: The left example shows an

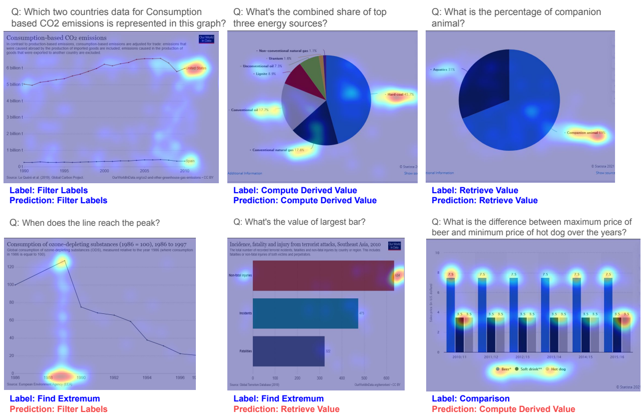

Figure 3: Examples of questions that the ChartQC model correctly (top row) and incorrectly classified (bottom row).

extreme value task in which the label must be identified. In this case (and in many others) more than one task label is suitable for the question, which can lead to problems, especially with single-label classification. This shows the need for further research into clearer labels and multi-label classification in intention prediction. In the middle example, the viewer is asked to read a value from the largest bar. Almost all fixations are directly placed on the target value in the heatmap, similar to the previous RV examples. We can therefore see that the visual behavior of already sorted bar charts, in particular, makes it difficult for the model to identify the FE label. In the right example, the model predicted a comparison even though it was a calculation task. Here, different labels fit the question again. Also, the step-by-step difference calculation of neighboring values looks similar to a comparison task. The specific behavior needed for one question type can therefore lead to potential patterns of another question type being imitated.

5 DISCUSSION

ChartQC Performance. ChartQC only achieves 0.452 in accuracy, which can be attributed to the inherent difficulty of predicting user intent from attention data. As discussed in [Wang et al. 2024a], attention data is often noisy, making extracting clear patterns for accurate classification challenging. The variability in human visual attention further complicates the prediction task, leading to suboptimal model performance. Still, the ablation study (Table 2) highlights that while individual components contribute to model performance, their optimal integration is crucial. The results suggest that metrics alone provide a weak signal, but when combined with saliency and visual features, they enhance predictive power. Additionally, the consistent improvement observed across models incorporating human attention maps indicates their strong relevance in capturing user attention patterns. Future work may explore alternative fusion strategies to enhance performance, such as adaptive weighting mechanisms for different modalities.

Limitations and Future Work. This work used mouse clicks as a proxy for human attention data. However, mouse clicks do not fully capture natural gaze behavior, which may limit the accuracy of attention-based modeling. Using gaze-tracking data could provide a more reliable representation of human attention patterns. Our current model does not explicitly consider task-related features such as label contents or whole-chart descriptions, which provide crucial context for question answering. Additionally, the image encoder was not explicitly trained to process human attention maps. Fine-tuning the image encoder on datasets incorporating gaze information, while also integrating task-related features, could enhance the model's ability to utilize both contextual and attention-based signals effectively. Another limitation is that some of the questions in the dataset contain multiple tasks (e.g. finding maximum values and performing calculations on them). While this can lead to performance problems with single-label classification (as in this study), there is great potential for future research and the application of multi-label classification.

Privacy and Ethics Statement. Demonstrating the feasibility and simplicity of learning user intent (question types) from human attention data is critical for raising awareness within the community about the potential privacy and ethical risks associated with human attention data. By highlighting these vulnerabilities, we emphasize the need for strong safeguards to prevent unintended leaks of sensitive information, encourage responsible development, and drive the creation of privacy-preserving solutions. Without a clear grasp of these risks, researchers and developers may unintentionally neglect ethical considerations, leaving systems vulnerable to exploitation and compromising user data privacy.

6 CONCLUSION

In this work, we explored the use of human attention data to predict question types in chart question-answering tasks. We proposed ChartQC , a novel model combining chart images and visual attention maps, utilizing spatial correspondence through feature concatenation to enhance classification performance. Our experiments demonstrated the feasibility of classifying question types using this approach, providing a promising direction for further research in human-computer interaction and designing adaptive systems for information visualization.

ACKNOWLEDGMENTS

T. Nishiyasu was supported by JST Adopting Sustainable Partnerships for Innovative Research Ecosystem (ASPIRE), Grant Number JPMJAP2303, and SPRING Grant Number JPMJSP2108. Y. Wang was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 251654672 – TRR 161. A. Bulling was funded by the European Research Council (ERC) under grant agreement 801708.

REFERENCES

Seoyoung Ahn, Conor Kelton, Aruna Balasubramanian, and Greg Zelinsky. 2020. Towards Predicting Reading Comprehension From Gaze Behavior. In *Proceedings of the ACM International Symposium on Eye Tracking Research and Applications (ETRA)*. 1–5.

Michael Barz, Sven Stauden, and Daniel Sonntag. 2020. Visual search target inference in natural interaction settings with machine learning. In *Proc. ACM Symposium on Eye Tracking Research and Applications (ETRA)*. 1–8.

Ali Borji, Andreas Lennartz, and Marc Pomplun. 2015. What do eyes reveal about the mind?: Algorithmic inference of search targets from fixations. *Neurocomputing* 149 (2015), 788–799.

Michelle A Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. 2015. Beyond memorability: Visualization recognition and recall. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 22, 1 (2015), 519–528.

Neil Bruce and John Tsotsos. 2005. Saliency based on information maximization. *Advances in neural information processing systems* 18 (2005), 1–8.

Andreas Bulling and Daniel Roggen. 2011. Recognition of visual memory recall processes using eye movement analysis. In *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp)*. 455–464.

Andreas Bulling and Thorsten O Zander. 2014. Cognition-aware computing. *Proceedings of the International Conference on Pervasive Computing (Pervasive)* 13, 3 (2014), 80–83.

François Chollet. 2017. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 1251–1258.

Alireza Fathi, Yin Li, and James M Rehg. 2012. Learning to recognize daily actions using gaze. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part I 12*. Springer, 314–327.

Steven R Gomez, Radu Jianu, Ryan Cabeen, Hua Guo, and David H Laidlaw. 2016. Fauxvea: Crowdsourcing gaze location estimates for visualization analysis tasks. *IEEE Transactions on Visualization and Computer Graphics* 23, 2 (2016), 1042–1055.

Zhiming Hu, Andreas Bulling, Sheng Li, and Guoping Wang. 2021. EHTask: Recognizing User Tasks from Eye and Head Movements in Immersive Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 29, 4 (2021), 1992–2004.

Michael Xuelin Huang, Jiajia Li, Grace Ngai, Hong Va Leong, and Andreas Bulling. 2019. Moment-to-Moment Detection of Internal Thought during Video Viewing from Eye Vergence Behavior. In *Proceedings of ACM Multimedia (MM)*. 1–9.

Weidong Huang. 2007. Using eye tracking to investigate graph layout effects. In *2007 6th International Asia-Pacific Symposium on Visualization*. IEEE, 97–100.

Sébastien Lallé, Cristina Conati, and Giuseppe Carenini. 2016. Predicting Confusion in Information Visualization from Eye Tracking and Interaction Data.. In *IJCAI*. 2529–2535.

Firas Lethaus, Martin RK Baumann, Frank Köster, and Karsten Lemmer. 2013. A comparison of selected simple supervised learning algorithms to predict driver intent based on gaze data. *Neurocomputing* 121 (2013), 108–130.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*. IEEE, 10012–10022.

Ahmed Masry, Xuan Long Do, Jia Qing Tan, Shafiq Joty, and Enamul Hoque. 2022. ChartQA: A Benchmark for Question Answering about Charts with Visual and Logical Reasoning. In *Findings of the Association for Computational Linguistics: ACL 2022*, Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). ACL, Dublin, Ireland, 2263–2279. https://aclanthology.org/2022.findings-acl.177

Laura E Matzen, Michael J Haass, Kristin M Divis, Zhiyuan Wang, and Andrew T Wilson. 2017. Data visualization saliency model: A tool for evaluating abstract data visualizations. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 24, 1 (2017), 563–573.

Takumi Nishiyasu and Yoichi Sato. 2024. Gaze Scanpath Transformer: Predicting Visual Search Target by Spatiotemporal Semantic Modeling of Gaze Scanpath. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 625–635.

Bastian Pfleging, Drea K Fekety, Albrecht Schmidt, and Andrew L Kun. 2016. A model relating pupil diameter to mental workload and lighting conditions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI)*. 5776–5788.

Patrik Polatsek, Manuela Waldner, Ivan Viola, Peter Kapec, and Wanda Benesova. 2018. Exploring visual attention and saliency modeling for task-based visual analysis. *Computers & Graphics* 72 (2018), 26–38.

Hosnieh Sattar, Mario Fritz, and Andreas Bulling. 2020. Deep Gaze Pooling: Inferring and Visually Decoding Search Intents From Human Gaze Fixations. *Neurocomputing* 387 (2020), 369–382.

Hosnieh Sattar, Sabine Muller, Mario Fritz, and Andreas Bulling. 2015. Prediction of search targets from fixations in open-world settings. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 981–990.

Danqing Shi, Yao Wang, Yunpeng Bai, Andreas Bulling, and Antti Oulasvirta. 2025. Chartist: Task-driven Eye Movement Control for Chart Reading. In *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. 1–14.

Sven Stauden, Michael Barz, and Daniel Sonntag. 2018. Visual search target inference using bag of deep visual words. In *Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz)*. Springer, 297–304.

Edward R Tufte. 1985. The visual display of quantitative information. *The Journal for Healthcare Quality (JHQ)* 7, 3 (1985), 15.

Xi Wang, Andreas Ley, Sebastian Koch, David Lindlbauer, James Hays, Kenneth Holmqvist, and Marc Alexa. 2019. The mental image revealed by gaze tracking. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI)*. 1–12.

Yao Wang, Mihai Bâce, and Andreas Bulling. 2023. Scanpath Prediction on Information Visualisations. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 30, 7 (2023), 3902–3914.

Yao Wang, Yue Jiang, Zhiming Hu, Constantin Ruhdorfer, Mihai Bâce, and Andreas Bulling. 2024a. VisRecall++: Analysing and predicting visualisation recallability from gaze behaviour. *Proceedings of the ACM on Human–Computer Interaction* 8, ETRA, Article 339 (2024), 18 pages.

Yao Wang, Chuhan Jiao, Mihai Bâce, and Andreas Bulling. 2022. VisRecall: Quantifying information visualisation recallability via question answering. *IEEE Transactions on Visualization and Computer Graphics* 28, 12 (2022), 4995–5005.

Yao Wang, Weitian Wang, Abdullah Abdelhafez, Mayar Elfares, Zhiming Hu, Mihai Bâce, and Andreas Bulling. 2024b. SalChartQA: Question-driven saliency on information visualisations. In *CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. Article 763, 14 pages.