Machine Learning-Based Model for Visualization Memorability Prediction

Kevin Choi[1] and Sangyoon Bae[#]

[1]Boston College, USA
[#]Advisor

ABSTRACT

Creating impactful visualizations is essential for effectively conveying complex data insights. However, current visualization techniques often fall short in ensuring that the data remains memorable and easily understandable. This research addresses this problem by investigating the factors that influence the memorability of visualizations, with the purpose of enhancing how key insights are retained and utilized. The main hypothesis is that combining visual content with simple textual metadata can predict and improve the memorability of visualizations. Our study introduces a web-based system powered by deep learning to predict memorability, analyzing both visual elements and textual descriptors, such as recognizable objects. The model assigns a memorability score based on these inputs. Results show that the hybrid model, integrating image and metadata analysis, provides accurate memorability predictions. The system offers designers immediate feedback, enabling them to iteratively refine their visualizations. This data-driven approach supports the creation of more memorable visual content, enhancing information retention and impact.

Introduction

Accurate data visualization is crucial for understanding and communicating complex information. Effective visualizations transform data into formats like charts and infographics, aiding quick comprehension and decision-making. This is particularly vital in fields such as scientific research, healthcare, and business. Despite their importance, designing memorable visualizations is challenging, often relying on manual, heuristic-based methods.

Recent studies have highlighted the significant role of visual elements in enhancing the memorability of data visualizations. For example, it demonstrated that certain visual features, such as color and the presence of human-recognizable objects, can significantly impact how well visualizations are remembered (Borkin, 2013). Similarly, it found that aesthetically appealing infographics are more likely to be recalled by viewers (Harrison, 2015). In addition, Convolutional Neural Networks (CNNs) predict image memorability at a large scale by approximating human memorability scores by analyzing visual features like texture, color, and object recognition (Khosla, 2015). These show the potential of deep learning in predicting and enhancing visual content memorability. In parallel, systems like DEEP-EYE automate data visualization by using machine learning to select and rank visualizations, highlighting the need for tools that efficiently create and evaluate visual content based on empirical data (Luo, 2018).

However, current visualization techniques often fall short in ensuring that data remains memorable and easily understandable. The manual design process is typically heuristic-based, relying on designers' intuition rather than empirical evidence, making it difficult to consistently create impactful visualizations. Additionally, the volume of data and the complexity of visual elements further complicate the design process. This gap between intuitive design and empirical methods indicates a need for better tools and approaches to creating memorable visualizations.

Given these challenges, our study aims to investigate the factors that influence the memorability of visualizations, with the purpose of enhancing how key insights are retained and utilized. We hypothesize that the memorability of data visualizations can be predicted by combining visual content with simple textual descriptions. This

approach leverages deep learning techniques to analyze both visual elements and contextual metadata, providing designers with immediate feedback on the potential impact of their visualizations.

To address this, we developed a web-based system that predicts the memorability of visualizations. Our system analyzes visual content alongside simple textual descriptors, such as the presence of recognizable objects, which act as metadata to inform the model's predictions. By assigning a memorability score based on both visual content and textual metadata, the system offers an objective measure of a visualization's potential impact on memory retention.

Our findings indicate that the hybrid model, which integrates image and metadata analysis, provides accurate memorability predictions. This system empowers designers to experiment, refine their creations iteratively, and adopt a more empirical, data-driven approach to visualization design. In summary, this research not only advances our understanding of visualization memorability but also provides a practical tool for creating more effective visual content.

We utilized the MASSVIS dataset to train our model, which contains a diverse collection of infographic-type images. As shown in figure 1 and figure 2, the dataset includes various visualizations, such as charts and diagrams, that are rich in visual features like color, texture, and recognizable objects. These images provided a robust foundation for training the hybrid model to predict and enhance the memorability of visual content effectively.

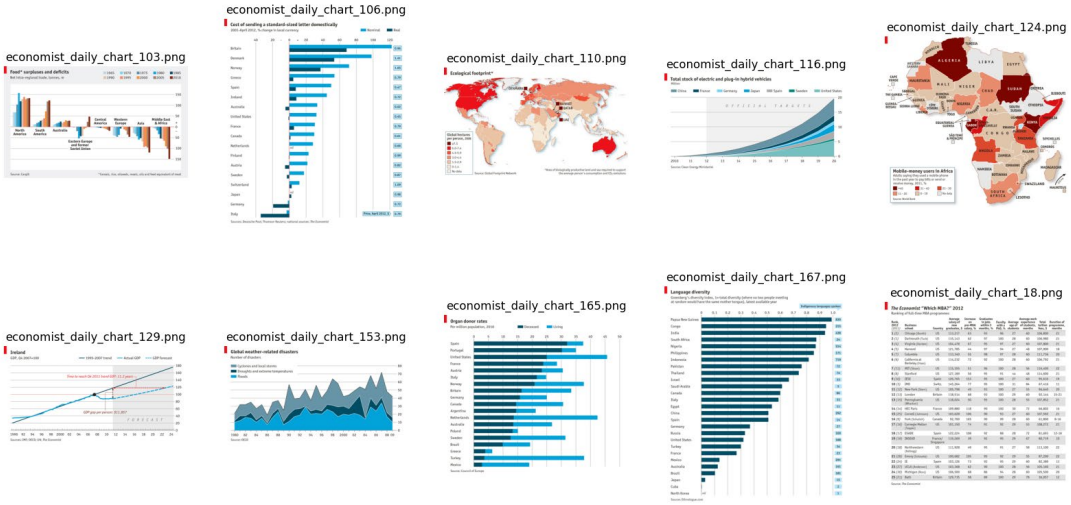

Figure 1. A news media visualization from the dataset, such as those sourced from "Economist Daily." These visualizations typically include moderate to high visual density with a focus on clarity and straightforwardness, aiding quick comprehension.

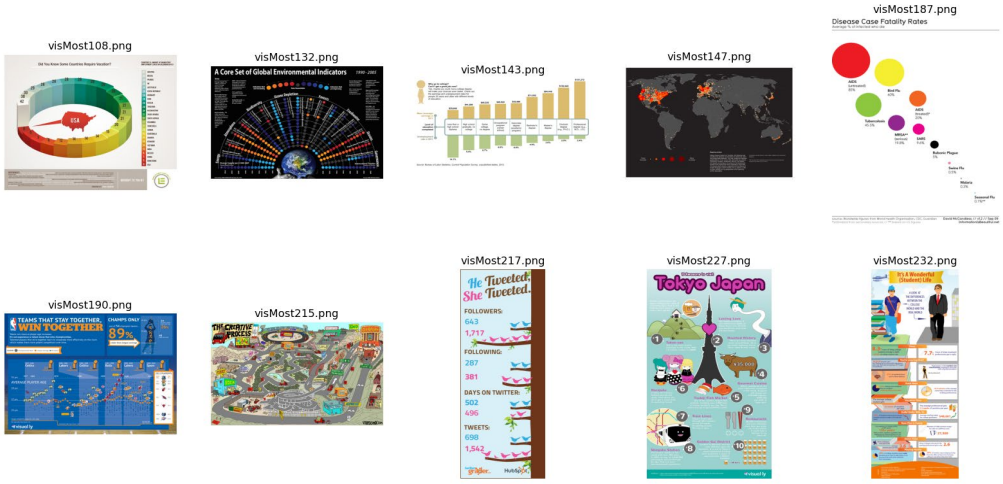

Figure 2. Example of an infographic from the dataset, showing a high number of distinct colors and complex visual density. These visualizations often feature human-recognizable objects and depictions, which significantly enhance their memorability.

Materials and Methods

Overview

The methods employed in this study systematically analyze and predict the memorability of data visualizations using both metadata and image data. Our approach included extensive metadata analysis, image data analysis using pre-trained Convolutional Neural Networks (CNNs), and the development of a hybrid model combining both types of analysis.

To quantify the memorability of visualizations, we utilized the d-prime (d') value derived from Signal Detection Theory. This metric effectively measures how well a memorable visualization can be distinguished from less memorable ones. Higher d-prime values indicate better discrimination and, consequently, higher memorability. D-prime is calculated by converting hit and false alarm rates into probabilities and then into z-scores to compute the d-prime as follows.

$$d' = z_H - z_F$$

This approach ensures accurate and reliable memorability measurements (Stanislaw, 1999).

Mean Squared Error (MSE)

Mean Squared Error (MSE) is a commonly used metric for evaluating the performance of regression models. It measures the average squared difference between the predicted values and the actual values. In the context of this study, MSE quantifies the accuracy of our model's predictions regarding the memorability scores of visualizations. A lower MSE indicates that the model's predictions are closer to the actual memorability scores, implying better model performance.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where n is the number of data points, y_i represents the actual value, and $\hat{y}_i$ represents the predicted value. Lower MSE indicates better model performance.

Data Splitting and Preprocessing

To ensure objective evaluation, the dataset was split into training (80%) and testing sets (20%). This split helps prevent overfitting and ensures that models can generalize to new data. The splitting was done using the 'train_test_split' function from the 'sklearn' library as shown below

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Metadata Analysis

The goal was to identify key attributes significantly influencing memorability, thereby reducing the number of required metadata inputs while maintaining high prediction accuracy. Various regression models were employed to understand the relationships between metadata features and memorability. Regression models used include Linear Regression, Ridge Regression, Lasso Regression, Support Vector Regression, Polynomial Regression, K-Nearest Neighbors, Decision Tree, Random Forest, Gradient Boosting, Bayesian Ridge Regression, and Gaussian Process Regression. Each model's performance was evaluated using MSE to determine the most effective features and their impact on memorability. As shown in figure 3, the correlation matrix highlights the relationships between key attributes and the d-prime value, showing that attributes like the presence of human-recognizable objects and the number of distinct colors significantly influence memorability.

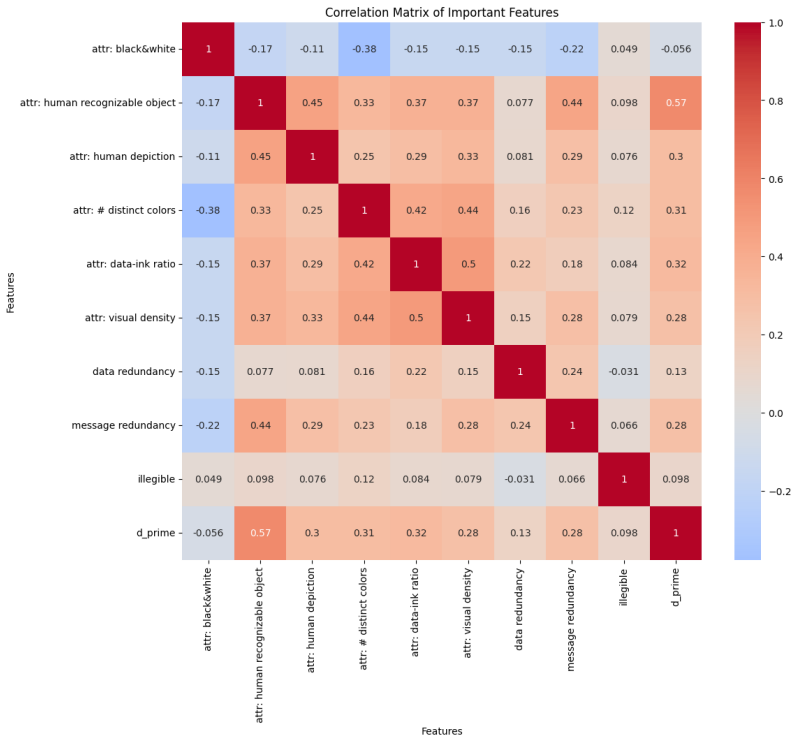

Figure 3. The correlation matrix of important features highlights the relationships between key attributes and the d-prime value. Bar graph showing correlation coefficients for various features such as the presence of human-recognizable objects and the number of distinct colors. Positive correlations indicate a strong influence on memorability, while negative correlations suggest a reduction in memorability.

Image Data Preparation and Analysis

Initially, various pre-trained CNN architectures were tested on the image data to extract complex visual features. Images were standardized to a common size and scale. All images were resized to 224x224 pixels and normalized to a [0, 1] range.

Consistency in Input Data

Initially, various pre-trained CNN architectures were tested on the image data to extract complex visual features. Images were standardized to a common size and scale. All images were resized to 224x224 pixels and normalized to a [0, 1] range. The sub-function for the pre-process image is programmed as follows.

```
def load_and_preprocess_image(img_path):
    img = image.load_img(img_path, target_size=(224, 224))
    img_array = image.img_to_array(img)
    return img_array / 255.0
```

Data Augmentation

Data augmentation techniques, including rotation, width shift, height shift, and horizontal flip, were applied to enhance the diversity and robustness of the training data. The algorithm of data generation is shown as follows.

```
data_generator = ImageDataGenerator(rotation_range=20, width_shift_range=0.2, height_shift_range=0.2, horizontal_flip=True)
augmented_data = data_generator.flow(X_train, y_train, batch_size=8)
```

Pre-Trained CNN Models Used and Model Comparison

Several pre-trained CNN models were employed in this study, including VGG16, ResNet-50, EfficientNet, VGG19, Xception, MobileNet, and DenseNet. Among the tested models, ResNet-50 emerged as the best pre-trained image model, achieving the lowest MSE.

Hybrid Model Development

Given the application's nature, where users are likely to upload images without detailed metadata, it was crucial to integrate both approaches. A hybrid model was developed that combines the strengths of both metadata and image analysis. The hybrid model integrates a pre-trained ResNet-50 CNN for image analysis with a Multi-Layer Perceptron (MLP) for metadata analysis as depicted in figure 4. A Test MSE of the proposed combined model shows a superior performance compared to that of ResNet-50.

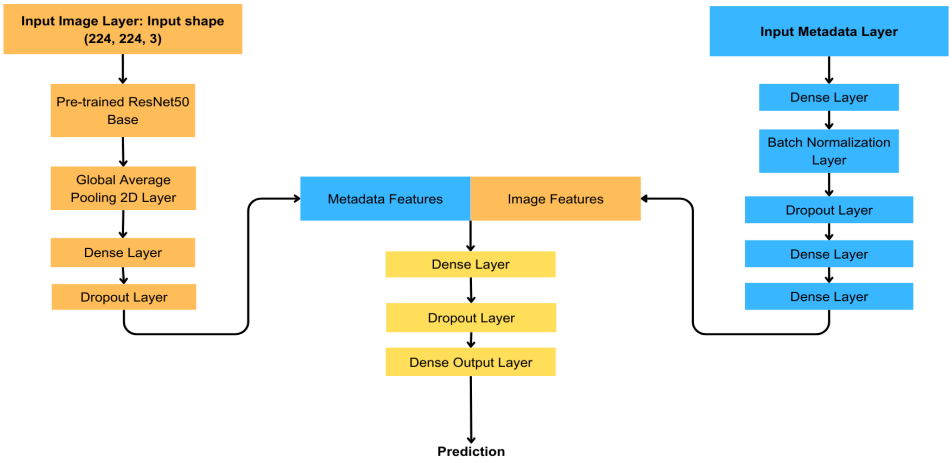

Figure 4. The architecture of the Hybrid Model illustrates the combined image processing using ResNet-50 and metadata processing using a multi-layer perceptron (MLP), demonstrating the integrated approach for predicting memorability.

Results

The experiments aimed to evaluate the performance of various models in predicting the memorability of visualizations using metrics such as Mean Squared Error (MSE) and training loss. Our approach involved three main analyses: assessing regression models using metadata features, adapting pre-trained Convolutional Neural Networks (CNNs) to our image data, and developing a hybrid model that combines both metadata and image data. To assess the impact of metadata on the memorability of visualizations, we evaluated several regression models using key metadata features such as the presence of human-recognizable objects, the number of distinct colors, black-and-white status, and human depiction. These features were chosen because previous research has shown they significantly influence how well visualizations are remembered. The performance of each model was measured using MSE, which quantifies how close the predicted values are to the actual memorability scores. The Linear Regression model achieved the lowest MSE of 0.229, indicating it provided the most accurate predictions of memorability based on metadata alone. In table 1, Ridge Regression and Bayesian Ridge Regression closely followed with MSE values of 0.230 and 0.236, respectively.

Table 1. Performance of regression models on metadata, shows the mean squared error (MSE) values for each regression model tested. Linear regression achieved the lowest MSE, indicating the most accurate predictions of memorability based on metadata alone, followed by Ridge regression and Bayesian Ridge regression.

Model	Mean Squared Error (MSE)
Linear Regression	0.229
Ridge regression	0.230
Bayesian Ridge Regression	0.236
Support Vector Regression	0.269
Polynomial Regression	0.280

K-Nearest Neighbors	0.296
Gradient Boosting	0.324
Random Forest	0.331
Lasso Regression	0.375
Decision Tree	0.391
Gaussian Process	0.391

To assess the impact of image data on the memorability of visualizations, we evaluated several pre-trained Convolutional Neural Network (CNN) models. These models, originally trained on large datasets like ImageNet, were adapted to our specific task by further training on the MASSVIS dataset. CNNs are particularly effective for image analysis because they can automatically learn and identify complex visual features. Among the models tested, ResNet-50 achieved the lowest MSE of 0.407, indicating its superior ability to extract and leverage complex visual features for predicting memorability as summarized in table 2.

Table 2. Performance of pre-trained CNN models on image data, lists the mean squared error (MSE) values for each pre-trained CNN model tested. ResNet-50 emerged as the best pre-trained image model with the lowest MSE, demonstrating superior ability to extract and leverage complex visual features for predicting memorability.

Model	Mean Squared Error(MSE)
ResNet-50	0.407
VGG16	0.421
EfficientNet	0.419
VGG19	0.467
Xception	0.448
MobileNet	0.495
DenseNet	0.530

We developed a hybrid model by combining Convolutional Neural Networks (CNNs) with select feature data. This hybrid model, integrating both image and feature data, achieved a Test MSE of 0.3446, marking a significant performance improvement. The reason for combining these two approaches is that while CNNs are excellent at analyzing visual content, metadata can provide additional context that enhances prediction accuracy. Switching from a standard CNN to a ResNet-50 architecture for image processing further reduced the MSE to 0.2580 as shown in table 3. The adjustment to ResNet-50 not only maintained performance but also improved it, demonstrating the upgraded model's enhanced efficiency.

Table 3. Performance and training time of hybrid models. This table compares the mean squared error (MSE) and training time of the hybrid models combining CNN and MLP, with the ResNet-50 architecture showing superior performance in terms of both accuracy and efficiency.

Model	MSE	Training time(s)
CNN + MLP	0.3446	78.21
ResNet-50 + MLP	0.2580	64.47

The developed web application allows users to upload visualizations and receive immediate feedback on their memorability scores. This application is user-friendly and designed to help users understand how memorable their visualizations are. It shows the interface where users can upload an image and select corresponding metadata features such as the presence of human-recognizable objects and whether the image is in black and white. Once the user clicks the 'Predict' button, the application calculates and displays the predicted d-prime value, providing immediate feedback on the memorability of the uploaded visualization as shown in figure 5.

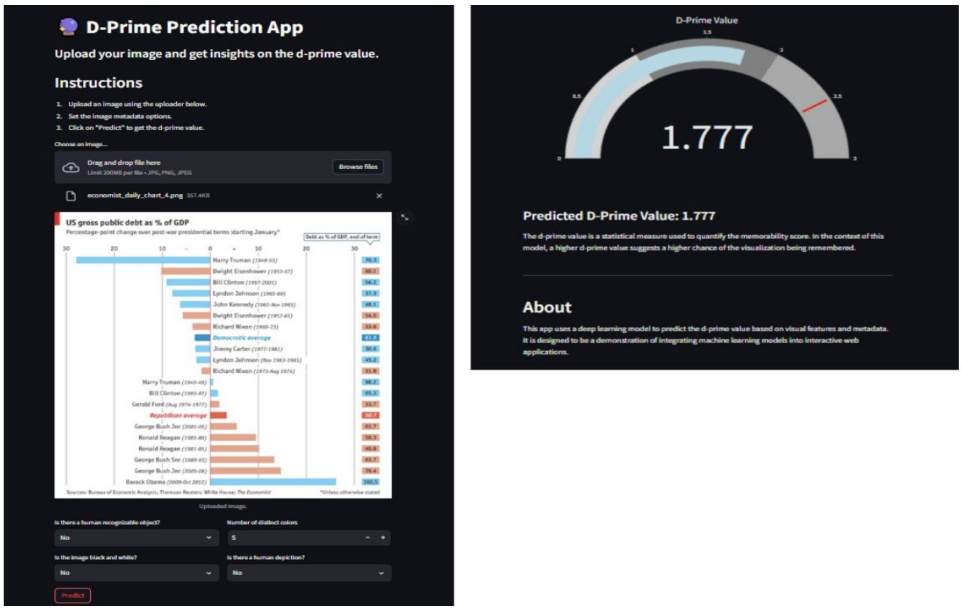

Figure 5. Screenshot of the D-Prime Prediction App web interface. The interface allows users to upload an image and select corresponding metadata features such as the presence of human-recognizable objects and whether the image is in black and white. Once the user clicks the 'Predict' button, the application calculates and displays the predicted d-prime value, providing immediate feedback on the memorability of the uploaded visualization.

Discussion

Reflecting on our research, it is evident that incorporating textual metadata significantly enhances the prediction of visualization memorability. Employing a hybrid model that combines visual and textual features has proven to be an effective strategy, balancing ease of use and predictive accuracy. Our model's strong performance on the MASSVIS dataset highlights the importance of metadata, such as human-recognizable objects and color diversity, in determining

memorability. However, our study is constrained by the scope and resolution of the dataset and metadata, indicating a need for expanding our research to include more diverse visual domains to fully realize the model's potential.

An important aspect for future work is the automation of metadata generation. Vision-language models (Vision-LLMs) present a promising avenue for this task. These models can automatically analyze and interpret visual content, creating descriptive metadata with minimal human intervention. Exploring free tools like CoPilot can offer valuable insights into the performance of Vision-LLMs with various visualization examples. By leveraging Vision-LLMs, we can significantly improve the efficiency and scalability of our approach, enabling broader applications and deeper insights without the need for extensive manual metadata annotation.

Conclusion

In conclusion, the research led to the establishing of a web application grounded in a machine-learning framework specifically designed to assess the memorability of data visualizations. The integration of ResNet-50 for image analysis alongside pivotal textual metadata has been instrumental in refining the prediction accuracy of memorability scores. This synthesis of complex visual information with essential textual descriptions marks a significant progression, offering designers a data-informed mechanism for enhancing the enduring impact of their visual work. The insights and outcomes of this study propose a shift in visualization design, moving towards an empirically supported amalgamation of creative intuition and computational analysis, which could initiate a transformative approach to data presentation.

Acknowledgments

We would like to thank Namwook Kim for his invaluable guidance and support throughout this research project.

References

Borkin, Michelle A., et al. (2013). "What Makes a Visualization Memorable?" IEEE Transactions on Visualization and Computer Graphics, 19(12). https://doi.org/10.1109/TVCG.2013.234.

Harrison, Lane, et al. (2015). "Infographic Aesthetics: Designing for the First Impression." Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. https://doi.org/10.1145/2702123.2702545.

Khosla, Aditya, et al. (2015). "Understanding and Predicting Image Memorability at a Large Scale." Proceedings of the IEEE International Conference on Computer Vision. https://doi.org/10.1109/ICCV.2015.275.

Luo, Yuzhong, et al. (2018). "Deepeye: Towards Automatic Data Visualization." 2018 IEEE 34th International Conference on Data Engineering (ICDE). https://doi.org/10.1109/ICDE.2018.00020.

Stanislaw, Harold, and Natasha Todorov. (1999). "Calculation of Signal Detection Theory Measures." Behavior Research Methods, Instruments, & Computers, 31(1). https://doi.org/10.3758/BF03207704.