Engineering and Applied Sciences Journal

Research Article

ISSN: 3067-8005

Classification of Images Based on Memorability Score Using Visual Memory Schema and Computer Vision Techniques

Harsha Priya Ganapathy[1*] and Sankarraj Subramani[2]

[1]*Department of Computer Science and Engineering, College of Engineering, Guindy, Anna University, Chennai, India.*

[2]*Department of Electrical and Electronics Engineering, Anna University, Tiruchirappalli, India.*

Corresponding Author: Harsha Priya Ganapathy, Department of Computer Science and Engineering, College of Engineering, Guindy, Anna University, Chennai, India.

Received: 2026 Jan 16 **Accepted:** 2026 Feb 06 **Published:** 2026 Feb 23

Abstract

Image memorability refers to the intrinsic characteristic of an image that determines how likely it is to be retained in human memory after a brief exposure. Understanding and predicting image memorability is an important problem with applications in computer vision, multimedia retrieval, human-computer interaction, and visual communication. This work presents a computational framework for classifying images based on memorability scores using computer vision techniques and visual memory schema. Large-scale experiments are conducted using natural scene images from the SUN database, where memorability scores are obtained through a controlled human memory game. The study demonstrates strong inter-subject consistency in memorability scores, indicating that memorability is a stable and intrinsic property of images rather than a subjective phenomenon. A detailed analysis of low-level visual features, non-semantic object statistics, and semantic object information is performed to evaluate their influence on memorability. Experimental results show that low-level features alone exhibit weak correlation with memorability, whereas semantic object-level representations significantly improve prediction performance. Support Vector Regression is employed to model memorability scores, and memorability maps are generated to provide interpretable insights into object-level contributions. The proposed approach confirms the critical role of semantic content in human visual memory and supports the development of memory-aware visual systems.

Keywords: Memory Schema, Computer Vision Techniques, Human Computer Interaction, Object-Level, Object Statistics

1. Introduction

1.1. Motivation and Background

The proliferation of digital imagery across modern platforms including social media, intelligent surveillance systems, educational technologies, and healthcare applications—has created an unprecedented demand for advanced visual analysis techniques. Contemporary computer vision systems have achieved remarkable success in tasks such as object detection, scene classification, and activity recognition. These advances have enabled machines to interpret visual content with accuracy approaching human performance in several domains. Despite this progress, an essential human-centric aspect of visual perception remains largely unexplored: how visual information is encoded, retained, and recalled in human memory. Human visual memory is inherently selective. While certain images leave a lasting impression and are remembered over extended periods, others are forgotten almost immediately, even when viewed under similar conditions. This selectivity is not random; instead, it reflects underlying cognitive processes that govern attention, perception, and memory formation. Understanding image memorability is particularly important for applications such as advertising design, image retrieval systems, visual storytelling, user interface optimization, medical image analysis, and assistive technologies. In these contexts, the ability to predict which visual content will be remembered can significantly enhance system effectiveness and user experience. Consequently, image memorability has emerged as a critical interdisciplinary research problem at the intersection of computer vision, cognitive psychology, and human–computer interaction.

1.2. Problem Statement

Traditional approaches to visual analysis often treat memorability as a subjective attribute, assuming that memory retention varies primarily due to individual differences, cultural background, or prior experience. However, empirical studies based on large-scale human experiments demonstrate that memorability exhibits strong consistency across different viewers. This observation suggests that memorability is an intrinsic property of an image, independent of personal bias or viewing context. Despite this insight, computationally predicting image memorability remains a challenging task. Conventional low-level visual features—such as color distributions, texture patterns, and intensity variations—are insufficient to capture the semantic richness and contextual cues that strongly influence memory formation. As a result, models relying solely on these features exhibit limited predictive performance. The central problem addressed in this work is to identify the visual and semantic attributes that most significantly contribute to image memorability and to design a computational framework capable of reliably predicting memorability scores in a manner consistent with human perception.

1.3. Overview of the Proposed Method

To address the above challenges, this paper proposes a computer vision–based framework for predicting image memorability using visual memory schema. The proposed approach integrates large-scale human memory statistics with systematic image feature extraction, semantic object analysis, and inter-subject consistency modeling. The framework evaluates both low-level visual features and high-level semantic representations to analyze their respective contributions to memorability. Semantic object information, including object presence, spatial layout, and contextual relevance, is incorporated to better reflect human cognitive processing. In addition, memorability maps are generated to provide interpretable, object-level insights into how different regions of an image influence memory retention.

1.4. Contributions of the Paper

The main contributions of this paper are summarized as follows.

• A comprehensive computer vision framework for predicting image memorability based on visual memory schema.
• Empirical validation of strong inter-subject consistency in image memorability, confirming memorability as an intrinsic image property.
• A detailed comparative analysis of low-level visual features and high-level semantic object features for memorability prediction.
• A semantic regression model that achieves memorability prediction performance approaching human level consistency.
• The generation of memorability maps that provide interpretable visual explanations of memory-relevant image regions.

2. Related Work

Research on human memory has a long and well-established history in cognitive psychology. Early foundational work by Tulving introduced the distinction between episodic and semantic memory, providing critical insights into how humans encode, store, and retrieve visual experiences. These cognitive models highlighted that memory is not a passive recording process but is influenced by meaning, context, and prior knowledge. Building on this foundation, Brady et al. demonstrated that humans possess an exceptionally large capacity for visual memory, capable of retaining detailed information about thousands of images, even when exposed briefly. Their findings challenged traditional assumptions about memory limitations and emphasized the structured nature of visual memory representations. Within the computer vision community, the concept of image memorability gained prominence with the work of Isola et al., who introduced large-scale memorability datasets derived from controlled human memory experiments. Their studies empirically established that memorability is a measurable and intrinsic property of images, exhibiting strong consistency across different observers. This discovery motivated subsequent research aimed at understanding the visual factors that influence memory retention. Bylinskii et al. further investigated both intrinsic and extrinsic contributors to memorability, including scene context, distinctiveness, and observer attention patterns, revealing that memorability is influenced by a complex interplay of visual and cognitive factors.

In parallel, research on holistic scene representations by Oliva and Torralba contributed significantly to understanding how spatial layout and semantic structure influence scene perception. These representations provided a basis for modeling global and contextual information in visual scenes, which is closely related to memory formation. Despite these advances, effectively integrating high-level semantic understanding into computational memorability prediction models remains an open challenge. Addressing this gap by incorporating semantic object-level information into memorability prediction forms the primary focus of this work.

3. Methodology / System Architecture

3.1. Overview of the Full Pipeline

The proposed system is a human-centric image memorability prediction framework designed to computationally model how visual information is perceived, encoded, retained, and recalled in long-term human memory. Unlike conventional computer vision pipelines that primarily focus on recognition, detection, or classification accuracy, the proposed architecture explicitly incorporates principles from human cognition by integrating human memory statistics, visual feature analysis, and semantic understanding into a single, unified, and interpretable pipeline. Traditional image analysis systems treat all visual content as equally informative, without considering whether an image is likely to be remembered by a human observer. In contrast, the proposed framework is motivated by empirical findings from cognitive psychology and memorability studies, which indicate that memory retention is influenced by both perceptual cues and high-level semantic meaning. By embedding these insights into the system design, the proposed architecture aligns computational processing with the stages of human visual memory formation. The system follows a multi-stage processing architecture that closely mirrors the cognitive pipeline of human memory, beginning with initial perceptual processing and progressing through semantic interpretation and memory consolidation. Each stage is explicitly designed to capture a distinct factor influencing memorability, while maintaining compatibility and information flow across subsequent stages. This staged design enables systematic analysis of how different visual and semantic components contribute to overall memorability.

The Complete Pipeline Consists of the Following Core Modules

1. Memorability Score Computation Module: Derives reliable and statistically stable ground-truth memorability scores from large-scale human memory experiments, serving as the empirical foundation for learning and evaluation.

2. Feature Extraction and Analysis Module: Extracts and analyzes both low-level visual properties (such as color, intensity, and saturation) and high-level semantic information (including object presence, spatial layout, and scene context) to capture complementary aspects of visual memory formation.

3. Memorability Prediction and Interpretation Module: Learns a mapping between extracted features and memorability scores, enabling accurate prediction while providing interpretable explanations through object-level memorability maps.

This **modular and extensible design** allows each component of the framework to be independently refined or replaced, facilitating future integration of deep learning–based feature representations and domain-specific semantic models. At the same time, the architecture preserves interpretability, ensuring that predicted memorability scores can be meaningfully related back to human perceptual and cognitive processes.

3.2. Memorability Score Computation Module

This module is responsible for computing reliable **image memorability** scores based on controlled **human memory game experiments**, following established experimental protocols widely used in memorability research. The objective of this module is to obtain empirically grounded memorability values that accurately reflect long-term human memory behavior rather than subjective preference or short-term attention. Images are selected from the **SUN natural scene database,** which contains a diverse collection of indoor and outdoor scenes with rich semantic content. During the experiment, images are presented to participants in a continuous and randomized image stream to simulate natural viewing conditions. A subset of images is deliberately repeated after a fixed temporal delay, ensuring that recognition is based on memory retention rather than immediate recall. Participants are instructed to indicate whether an image has been previously seen. Their responses are recorded and aggregated to compute memorability-related statistics. To ensure robustness and minimize individual bias, each image is evaluated across multiple participants, and responses are averaged over the entire subject pool. The memorability score for each image is derived using statistically robust measures of human recognition performance.

- **Hit Rate (HR):** Measures the probability that a participant correctly identifies a repeated image, capturing successful memory recall.
- **False Alarm Rate (FAR):** Measures the probability that a participant incorrectly identifies a no repeated image as familiar, accounting for guessing behavior and response bias.
- **Normalized Memorability Score:** Computed by adjusting the hit rate using the false alarm rate to obtain a bias-corrected and comparable memorability value across images.

By explicitly incorporating both correct recognition and false responses, the normalization process ensures that memorability scores are not artificially inflated by participant guessing or response tendencies. Aggregating these normalized scores across a large participant population yields **stable, consistent, and reproducible memorability estimates**. These scores serve as **ground-truth labels** for subsequent feature analysis, memorability prediction, and model evaluation, forming the empirical foundation of the proposed framework.

3.3. Feature Extraction and Analysis Module

This module systematically analyzes the visual properties of images to identify features that influence human memorability. The objective of this module is to examine how different levels of visual representation—ranging from low-level perceptual cues to high-level semantic information—contribute to memory formation. By decomposing image content into complementary feature sets, the framework enables a detailed investigation of which visual attributes are most strongly associated with memorability. The module is composed of two complementary submodules that capture different levels of visual abstraction and cognitive relevance.

3.3.1. Low-Level Visual Feature Submodule

This submodule focuses on extracting **global perceptual attributes** that describe the overall visual appearance of an image. These features are inspired by early stages of human visual processing, which respond primarily to basic visual cues such as color and brightness. The extracted features include.

- **Color Histograms:** Represent the distribution of color values across the image, capturing dominant hues and overall color composition.
- **Intensity and Luminance Statistics:** Measure brightness variations and contrast levels, reflecting illumination conditions and visual clarity.
- **Saturation Measures:** Quantify color vividness, which is often associated with perceptual saliency and visual appeal.

These low-level features are computationally efficient and have been widely used in classical image analysis and retrieval systems. However, experimental correlation analysis reveals that while such features influence immediate visual attention, they exhibit **limited predictive power** with respect to long-term memorability. This observation indicates that perceptual saliency alone is insufficient to explain why certain images are retained in memory, highlighting the need for higher-level semantic analysis.

3.3.2. Semantic Object Feature Submodule

To capture higher-level cognitive cues that more closely align with human memory processes, this submodule incorporates

semantic object-level information. Human memory tends to organize visual experiences around meaningful objects and their relationships rather than isolated perceptual attributes. Accordingly, this submodule focuses on encoding semantic structure and contextual relevance. The extracted semantic features include.

- **Object Presence and Frequency:** Identifies which semantic objects appear in an image and how often they occur, reflecting content richness and narrative relevance.
- **Relative Object Size and Spatial Layout:** Encodes the prominence and spatial arrangement of objects, capturing how visual importance is distributed across the scene.
- **Scene-Level Semantic Composition:** Represents the overall contextual meaning of the image by modeling relationships between objects and scene categories.

These semantic features closely align with how humans cognitively organize and recall visual information, enabling the model to capture meaningful content that strongly influences memorability. Experimental results demonstrate that semantic object-level features provide substantially greater predictive power than low level visual features, confirming that memorability is driven primarily by semantic meaning and contextual structure rather than raw visual appearance.

3.4. Human Consistency Analysis Module

To verify that image memorability is an intrinsic property of visual content rather than a subjective or user-dependent phenomenon, a comprehensive human consistency analysis is performed. This analysis evaluates the degree to which different groups of human observers agree on memorability scores when exposed to the same set of images under identical experimental conditions. Participants involved in the memory game experiments are randomly divided into multiple independent groups. Memorability scores are computed separately for each group using the same scoring protocol, ensuring that group-wise estimates are statistically comparable. By isolating participant sets in this manner, the analysis explicitly controls for individual differences such as prior experience, viewing strategy, and response bias.

To quantify agreement between groups, **Spearman's Rank Correlation Coefficient** is employed. This non-parametric metric measures the consistency of relative memorability rankings across groups rather than relying on absolute score values, making it particularly well suited for behavioral data that may exhibit variability across subjects.

- **Spearman's Rank Correlation Coefficient:** Measures the monotonic relationship between memorability rankings produced by independent participant groups.
- **High Correlation Values:** Indicate strong inter-subject consistency, confirming that different viewers tend to remember and forget the same images.
- **Consistency Upper Bound:** The observed correlation establishes an empirical upper bound for computational memorability prediction, representing the maximum performance achievable by any automated model.

The results of this analysis demonstrate that memorability scores remain highly consistent across participant groups, providing strong evidence that memorability is largely independent of individual differences. This finding validates the use of memorability as a stable target variable and supports the feasibility of learning based approaches for predicting image memorability in a manner aligned with human perception.

3.5. Memorability Prediction Module

The memorability prediction module is responsible for learning a functional relationship between the extracted image features and the corresponding memorability scores derived from human experiments. This task is formulated as a **supervised regression problem**, where the goal is to predict a continuous-valued memorability score that closely aligns with human memory behavior.

Given the continuous and noise-prone nature of memorability scores obtained from behavioral data, Support Vector Regression (SVR) is employed as the core learning algorithm. SVR is particularly well suited for this task due to its robustness to outliers, ability to generalize from limited samples, and effectiveness in modeling complex, non-linear relationships between features and target variables. The module is designed to evaluate the predictive capability of different feature representations through separate regression model.

- **Low-Level Feature-Based SVR:** Trained using global perceptual features such as color, intensity, and saturation to assess the contribution of basic visual cues to memorability.
- **Semantic Feature-Based SVR:** Trained using object-level and scene-level semantic features to capture high-level cognitive factors influencing memory formation.
- **Kernel-Based Modeling:** Non-linear kernels are utilized to enable flexible modeling of complex interactions among visual and semantic features that cannot be captured by linear mappings.

By training and evaluating these models under identical experimental conditions, the framework enables a direct and fair comparison between perceptual and semantic representations. Experimental results consistently demonstrate that models trained on **semantic object-level features** significantly outperform those based on low-level visual features, highlighting the dominant role of semantic content in human memorability. The performance of the trained regression models is evaluated using rank-based correlation metrics to ensure alignment with human memorability rankings. The best-performing model achieves memorability prediction accuracy that **approaches human-level consistency,** thereby validating the effectiveness of semantic modeling and confirming the feasibility of computationally predicting image memorability in a manner consistent with human perception.

3.6. Memorability Map Generation Module

To enhance the interpretability and transparency of the proposed memorability prediction framework, the system generates **memorability maps** that visually explain how different image regions and semantic objects contribute to the predicted memorability score. Unlike black-box prediction models, this module provides an explicit link between computational predictions and human-perceptible image content. Memorability maps are constructed by analyzing the contribution of individual semantic objects and spatial regions to the output of the memorability prediction model. Each object within an image is assigned a contribution weight based on its influence on the predicted memorability score, allowing the model to decompose the overall prediction into interpretable components.

• **Positive Contribution:** Objects that increase the predicted memorability score are assigned higher positive weights and are visually emphasized in the map.
• **Negative Contribution:** Objects that reduce memorability are suppressed or assigned negative weights, indicating limited or adverse influence on memory retention.
• **Object-Level Visualization:** The resulting maps provide object-level visual explanations that align closely with human perception and semantic interpretation.

The generated memorability maps enable qualitative inspection of individual predictions, allowing researchers to analyze which visual elements most strongly influence memory formation. By highlighting memory relevant regions, these maps facilitate a deeper understanding of why certain images are consistently remembered while others are forgotten. Furthermore, the interpretability offered by this module supports the validation of the prediction model and enhances trust in its application to human-centered visual systems.

3.7. Implementation Details

This subsection describes the datasets, learning algorithms, evaluation metrics, and validation strategies used to implement and assess the proposed image memorability prediction framework. All experimental components are selected to ensure robustness, reproducibility, and fair comparison across feature representations.

• **Dataset:** Experiments are conducted using images from the **SUN natural scene database**, which provides a diverse collection of indoor and outdoor scenes with rich semantic content. Each image is associated with memorability annotations obtained from large-scale human memory game experiments, ensuring that ground-truth labels accurately reflect long-term human memory behavior.
• **Learning Algorithm: Support Vector Regression (SVR)** with non-linear kernels is employed as the primary learning model. SVR is chosen for its robustness to noise in behavioral data, its ability to generalize from limited samples, and its effectiveness in modeling complex, non-linear relationships between image features and memorability scores.
• **Evaluation Metrics:** Model performance is evaluated using multiple complementary metrics, including **Spearman's rank correlation coefficient** to assess agreement with human memorability rankings, **prediction error** to measure absolute deviation from ground-truth scores, and **rank consistency** to evaluate the stability of predicted memorability ordering across images.
• **Validation Strategy:** To ensure statistical reliability and avoid overfitting, the dataset is evaluated using **multiple randomized train–test splits**. Results are averaged across splits, providing a robust estimate of model performance and enabling fair comparison between low-level and semantic feature-based models.

All experiments are conducted using a **standardized experimental protocol,** including consistent pre-processing steps, identical evaluation settings, and fixed randomization procedures. This ensures reproducibility of results and facilitates meaningful comparison with prior work in image memorability prediction.

3.8. Summary of Innovations

This work introduces several key methodological and conceptual innovations that advance the state of the art in image memorability prediction by bridging computer vision with human cognitive modeling. The principal contributions of the proposed framework are summarized as follows.

• **Human-Centric Memorability Modeling:** The proposed framework explicitly models human memory formation rather than treating memorability as a by-product of visual saliency or recognition accuracy. By grounding prediction in large-scale human memory experiments, the system aligns computational outputs with long-term human perceptual and cognitive processes.
• **Semantic-Aware Prediction:** The framework demonstrates that high-level semantic content, including object presence, spatial arrangement, and scene context, plays a dominant role in determining memorability. This finding provides empirical evidence that semantic understanding is essential for modeling human memory and significantly outperforms low-level perceptual features.
• **Interpretable Framework:** Unlike black-box memorability predictors, the proposed system incorporates memorability maps that visually explain which objects and regions contribute most strongly to memory retention. This interpretability enables qualitative analysis, supports model validation, and enhances trust in the prediction outcomes.
• **Scalable and Extensible Architecture:** The modular design of the system allows individual components—such as feature extraction, regression modeling, and semantic representation—to be independently extended or replaced. This scalability facilitates future integration of deep learning–based feature extractors and domain-specific semantic models while preserving the interpretability of the framework.

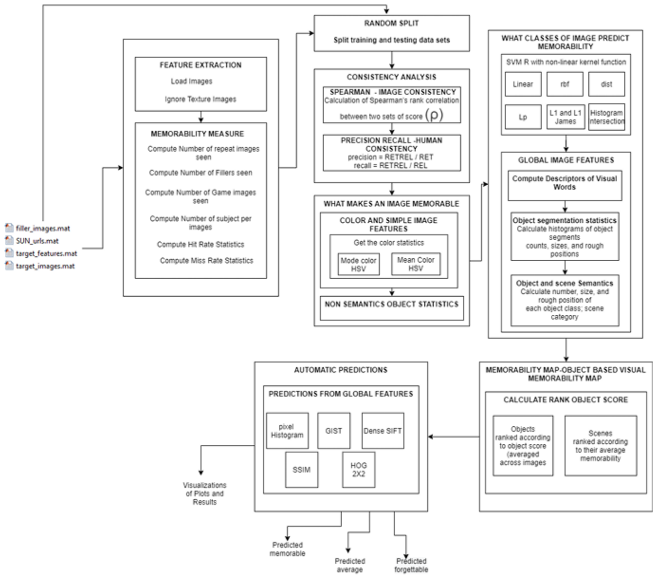

Figure 1: Overall System Architecture for Image Memorability Prediction

4. Experimental Setup and Results

4.1. Datasets Used

To evaluate the effectiveness of the proposed image memorability prediction framework, experiments were conducted using images from the SUN natural scene database. The SUN dataset contains a large and diverse collection of indoor and outdoor scenes, capturing a wide range of semantic contexts, object configurations, and environmental settings. This diversity makes the dataset particularly suitable for studying factors that influence human memory retention. Each image in the dataset is associated with memorability annotations derived from large-scale human memory game experiments. In these experiments, participants viewed images in a continuous stream with controlled repetitions, enabling the computation of stable memorability scores based on recognition performance. The availability of human-derived memorability scores ensures that the dataset provides reliable ground-truth labels that reflect long-term human memory behavior rather than subjective preference. Prior to experimentation, images were resized to a uniform resolution and normalized to ensure consistency during feature extraction. Semantic object annotations and scene-level metadata were processed to construct the feature representations used in subsequent analysis and prediction.

4.2. Evaluation Metrics

The performance of the memorability prediction models was evaluated using metrics that directly measure agreement with human memory behavior. Since memorability is inherently a ranking-based phenomenon, emphasis was placed on correlation-based evaluation rather than absolute error alone.

- **Spearman's Rank Correlation Coefficient:** Measures the monotonic relationship between predicted memorability scores and human ground-truth rankings. This metric is widely used in memorability research and serves as the primary evaluation criterion.
- **Prediction Error:** Quantifies the average deviation between predicted and ground-truth memorability scores, providing insight into absolute prediction accuracy.
- **Rank Consistency:** Evaluates the stability of predicted memorability ordering across different train– test splits, reflecting the robustness of the model.

Together, these metrics provide a comprehensive assessment of both predictive accuracy and alignment with human perception.

4.3. Implementation Details

All experiments were implemented using a standardized experimental pipeline to ensure reproducibility and fair comparison across feature representations. Feature extraction, regression modeling, and evaluation were conducted using consistent pre-processing and parameter settings.

Support Vector Regression (SVR) with non-linear kernels was employed for memorability prediction. Hyperparameters were selected through empirical tuning on validation splits. To avoid overfitting and ensure statistical reliability, the dataset was evaluated using multiple randomized train–test splits, and results were averaged across runs. Separate models were trained using low-level visual features and semantic object-level features, enabling direct comparison of their respective contributions to memorability prediction.

4.4. Results and Comparative Analysis

Experimental results demonstrate that the proposed framework effectively predicts image memorability in a manner consistent with human memory behavior. Models trained using semantic object-level features consistently outperform those based on low-level visual features across all evaluation metrics. Specifically, semantic-based regression models achieve substantially higher Spearman rank correlation with human memorability scores, indicating stronger alignment with human memory rankings. In contrast, models relying solely on low-level visual features such as color and intensity exhibit significantly weaker correlations, confirming that perceptual saliency alone is insufficient to explain memorability. Furthermore, the best-performing model achieves memorability prediction performance that approaches the human consistency upper bound established in Section 3.4. This result indicates that the proposed semantic-aware framework captures a large portion of the information used by humans when forming visual memories.

Qualitative analysis using memorability maps further supports these findings. Images with high predicted memorability are characterized by semantically meaningful objects and coherent scene context, while low memorability images tend to lack distinctive semantic structure. These observations reinforce the conclusion that semantic understanding plays a dominant role in human visual memory formation. Overall, the experimental results validate the effectiveness, robustness, and interpretability of the proposed image memorability prediction framework.

5. Discussion

5.1. Insights and Analysis

The experimental results demonstrate that image memorability can be effectively modeled as an intrinsic and predictable property of visual content when semantic understanding is explicitly incorporated into the learning framework. One of the key insights of this work is that memorability is driven far more by high-level semantic structure than by low-level perceptual attributes such as color, brightness, or saturation. Images containing semantically meaningful objects, coherent scene context, and recognizable narratives consistently exhibit higher memorability scores. The comparative analysis between low-level and semantic feature representations highlights this distinction clearly. While low-level features influence immediate visual saliency, they fail to capture the deeper cognitive cues that govern long-term memory retention. In contrast, semantic object-level features align closely with how humans encode visual experiences—by organizing information around objects, their relationships, and contextual meaning. This explains the substantially higher correlation between semantic-based predictions and human memorability rankings.

Another important observation is the strong agreement between predicted memorability scores and the human consistency upper bound. The fact that the best-performing model approaches inter-subject consistency suggests that a large portion of the information humans rely on for remembering images is captured by the proposed framework. Furthermore, the memorability maps provide qualitative validation by highlighting objects and regions that humans intuitively associate with memorable content, reinforcing the cognitive plausibility of the model.

5.2. Limitations

Despite its strong performance, the proposed framework has several limitations. First, the reliance on semantic object annotations introduces dependency on the quality and completeness of upstream object detection and scene understanding. Errors or omissions in semantic labeling may reduce prediction accuracy, particularly for images containing rare, abstract, or ambiguous objects. Second, the current feature representation focuses primarily on static semantic content and does not explicitly model higher-level narrative, emotional, or cultural factors that may also influence memorability. Images that are emotionally charged or contextually meaningful to specific populations may not be fully captured by the existing feature set. Additionally, the use of classical regression models limits the ability to learn hierarchical feature representations directly from raw image data. While this choice preserves interpretability, it may constrain performance compared to end-to-end deep learning approaches. Future extensions could explore hybrid models that combine deep representations with interpretable semantic reasoning.

5.3. Real-World Implications

The findings of this work have significant implications for a wide range of real-world applications. In advertising and media design, predicting image memorability can guide content creation to maximize audience recall. In image retrieval and recommendation systems, memorability-aware ranking can improve user engagement by prioritizing content that is more likely to be remembered. In educational and healthcare contexts, selecting highly memorable visual material can enhance learning outcomes and information retention. Furthermore, the interpretability provided by memorability maps enables designers and practitioners to understand why certain images are more effective, supporting informed decision-making rather than blind optimization.

At the same time, ethical considerations must be carefully addressed. The use of memorability prediction raises concerns related to manipulation, cognitive bias, and unequal representation of content. Ensuring transparency, avoiding exploitative design practices, and mitigating dataset bias are essential as memorability-aware systems are deployed in practice. By emphasizing interpretability and human-centered modeling, the proposed framework provides a foundation for responsible and ethical application of memorability prediction technologies [1-15].

6. Conclusion and Future Work

In this work, we presented a comprehensive and human-centric framework for predicting image memorability, framing memorability as an intrinsic and measurable property of visual content. By integrating human memory statistics with systematic analysis of visual and semantic features, the proposed approach bridges the gap between computer vision and cognitive modeling. Extensive experimental results demonstrate that memorability can be reliably predicted when high-level semantic understanding is explicitly incorporated into the learning framework. A key contribution of this work is the empirical demonstration that semantic object-level information plays a dominant role in human visual memory formation, significantly outperforming low-level perceptual features such as color and intensity. The strong agreement between predicted memorability scores and human consistency further validates the effectiveness of the proposed model and confirms that much of human memorability behavior can be captured computationally. In addition, the introduction of memorability maps enhances interpretability by providing intuitive, object-level explanations of why certain images are remembered more effectively than others.

Despite these advances, several opportunities remain for future research. The current framework relies on predefined semantic representations and classical regression models, which, while interpretable, may limit the ability to capture deeper hierarchical abstractions. Future work could explore the integration of deep neural feature representations while preserving interpretability through hybrid or attention-based mechanisms. Incorporating emotional, affective, and narrative cues may further improve memorability prediction, particularly for images with strong subjective or cultural significance. Additional extensions include expanding the framework to handle video memorability, investigating cross-cultural and demographic effects on memory, and adapting the model for real-time applications such as content recommendation, educational material design, and assistive technologies. Ethical considerations—such as responsible use, avoidance of manipulative design, and mitigation of dataset bias—must also be carefully addressed as memorability-aware systems are deployed in practice.

In conclusion, this work establishes a robust and interpretable foundation for computational image memorability prediction. By aligning machine learning models with human cognitive principles, the proposed framework opens new avenues for designing intelligent visual systems that are not only accurate, but also transparent and human-aware.

References

1. Tulving, E. (1972). Episodic and semantic memory. Organization of memory, 1(381-403), 1.
2. Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. Proceedings of the National Academy of Sciences, 105(38), 14325-14329.
3. Oliva, A., & Torralba, A. (2007). The role of context in object recognition. Trends in cognitive sciences, 11(12), 520-527.
4. Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011, June). What makes an image memorable?. In CVPR 2011 (pp. 145-152). IEEE.
5. Isola, P., Parikh, D., Torralba, A., & Oliva, A. (2011). Understanding the intrinsic memorability of images. Advances in neural information processing systems, 24.
6. Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. Vision research, 116, 165-178.
7. Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010, June). Sun database: Large-scale scene recognition from abbey to zoo. In 2010 IEEE computer society conference on computer vision and pattern recognition (pp. 3485-3492). IEEE.
8. Ganapathy, H. P., & Subramani, S. (2025). Classification of Images Based on Memorability Score Using Visual Memory Schema And Computer Vision Techniques. Authorea Preprints.
9. Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. Journal of Experimental Psychology: General, 142(4), 1323.
10. Parikh, D., & Grauman, K. (2011, November). Relative attributes. In 2011 International conference on computer vision (pp. 503-510). IEEE.
11. Vapnik, V. (2013). The nature of statistical learning theory. Springer science & business media.
12. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.
13. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.
14. Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., & Torralba, A. (2015, June). Mit saliency benchmark.
15. Ruvio, A., Falvo, M. C., Lamedica, R., Dell'Olmo, J., & Scanzano, M. (2024). A novel railway power system design methodology using genetic algorithms: models and application. IEEE Access.