# Understanding Image Memorability through Localized Stimuli

Amir Shokri[1], Farzin Yaghmaee[1]*

***Abstract*— In today's digital age, we are bombarded with images from the internet, social media, and online magazines. It is fascinating how we can remember so many of these images and their details. However, not every image is equally memorable; some stay with us more than others. Scientists have explored why this is the case. In our research, we are particularly interested in how images that showcase Iranian life and culture stick in the memories of Iranian adults. To investigate this, we created a new collection called the SemMem dataset, which is full of culturally relevant images. We adapted a memory game from earlier studies to test how memorable these images are. To analyze memorability, we used two deep learning architectures, ResNet 50 and ResNet 101. These architectures helped us estimate which images are likely to be remembered. Our findings confirmed that images connected to Iranian culture are indeed more memorable to Iranians, highlighting the impact of familiar cultural elements on memory retention.**

***Index Terms*— Visual Memory, Memorability, Image Memorability, Recognition Memory, Quantifying image memorability.**

## I. INTRODUCTION

In our daily routines, we encounter numerous images and videos on social media and in newspapers. we watch images when we read the newspaper. We are exposed to many advertising images while driving on highways and streets, where billboards and other media catch our eye. Our memory has an exceptional ability to retain images, making them either memorable or forgettable [1]. Marketing, photography, design, and many other fields leverage image memorability. We have a primary question in image memory research is, " What makes an image memorable?" [2]

As a result, image memorability is a considered qualitative variable, but there is a desire to quantify it to calculate the rank of image memorability. [3] According to previous research, image memorability can be quantified and is related to the features of the image.

Although our attention to images varies, image memorability is initially seen as a descriptive variable. However, previous research demonstrated that memorability rank is a numerical value that can be quantified and calculated. Every research project requires datasets, as they transform fundamental research into exemplary research. Measuring image memorability necessitates comparison with other images since the rank of image memorability is calculated within a series of images.

## II. MATERIAL AND METHODS

### A. Measurement of memorability

The memorability rank in visual memory assessment runs from 0 to 1. In order to gauge how memorable an image is, Isola et al. [4] presented a visual memory game that users could play using Amazon Mechanical Turk crowdsourcing services. In this game, participants see a sequence of pictures for one second at a time, with 1.4 seconds elapsing between each image. The task for participants task is to identify duplicate images within the sequence by pressing the spacebar. After collecting the data from participants, researchers calculate the memorability rank for target images. According to recent research by Lore Goetschalckx et al., the memorability rank is consistent across participants, indicating that both memorable and forgettable images remain consistent across different individuals. By modifying the procedure for visual memory games to show each image for 600 milliseconds and allowing an 800-millisecond gap before the next image, we presented 200 images from the MemCat [5] dataset. This set included 66 target images with one repetition, 44 filler images without repetition, and 12 images to keep participants engaged, each repeated once.

In this method, it is necessary to calculate the $H/N_{resp}$ [4] for each image to assign a memorability rank to the target image. The memorability rank is a numerical value that ranges between zero and one.

### B. The most popular memorability dataset

LaMem and MemCat are two of the largest datasets in image memorability, among several others[6]. When using a dataset for image memorability studies, specific criteria must be considered. These include the number of images within the dataset, the diversity and number of participants involved in the memorability studies, and the extent to which the images are balanced in terms of variety and representation. The LaMem dataset is significant for measuring image memorability more than 60,000 images without bounding boxes or segmentation information. In contrast, MemCat is a category-based dataset equipped with bounding boxes and includes 10,000 images for the measurement of image memorability.

$$\rho = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)} \quad (1)$$

$\rho = Spearman'srank\ correlation$ [7] $coefficient.$
$d_i = diffrence\ between\ the\ two\ ranks$

1- Faculty of Electrical and Computer Engineering (ECE), Semnan University, Semnan, Iran.

Corresponding author: f_yaghmaee@semnan.ac.ir

$of\ each\ observation$
$n = number\ of\ observations$

Spearman's Rank Correlation Coefficient, ρ, is used to measure the strength and direction of the association between two ranked variables in the context of image memorability assessment.

Spearman's Rank Correlation Coefficient provides a crucial statistical link between the consistency of image definitions and memorability. With ρ values of 0.75 for LaMem consistency and 0.78 for image memorability, there is a significant positive correlation within the LaMem dataset. This correlation highlights the trend that images with consistent definitions across different viewers are often more memorable. Such insight e crucial advancing our understanding of image memorability, a field that benefits from diverse datasets including Face Mem, SUN, MASSIVE, FIGRIM, Object Mem, VISCHMEA, and LNSIM.

### *C. Quantifying image memorability in Iranian adult*

In this study, we examine image memorability among Iranian adults. Our goal is to uncover what differentiates image memorability among individuals from diverse regions. In our next phase, we will focus on the memorability on the memorability of local images. Consequently, some images in previous datasets in the field of image memorability lack contextual relevance for Iranians, prompting our choice of this study. For instance, certain cars or animals present in international datasets may not be familiar in our region, whereas local cars and animals are. Contextual significance plays a crucial role in image memory.

### *D. Redesign Visual memory game*

Previously [15] [5], researchers used Amazon Mechanical Turk, but these tools were unavailable to Iranians.

In order to solve this, we used the Python programming language with Django framework to create a novel visual memory game. Our stimuli were published via cloud services and made accessible online. Initially, participants read the rules of our stimuli and signed all agreements to participate in our research, with the option to remove their data at any time (we hard delete cancelled participants from our database). Our participants are volunteers, receiving no rewards. We developed an English-Persian visual memory game and validated participants via email or mobile number to prevent repeat participation (we removed participant email and mobile numbers to prevent repeat participation (participant email and mobile numbers were removed after our research)). We collected user agent information and IP addresses for verification at the research's end, ensuring each participant only played our task once. Our cloud server, running Ubuntu OS with 16GB of RAM, a 50GB SSD hard drive, and a static IP address, hosted the visual memory game, which is secured against XSS and CSRF attacks and uses SSL certificates for data encryption.

Efficient image loading is essential for online visual memory games. To fix this, we included a page loader and disabled the game's start button until all picture data had loaded, allowing players to begin playing at the same time. Given the relatively small size of our dataset and the limited number of participants, we adjusted the stimuli variable for our game. Participants engaged in five blocks, each a game round, viewing a series of images for 600ms, followed by an 800ms pause. Responses were made by pressing the spacebar or clicking the mouse.

Each block was carefully composed to balance engagement and memory task demands, including 30 unique target images central to the task, 22 unique filler images from the MemCat dataset to add variety without repetition, and 9 vigilance images to monitor attention. Target and vigilance images were shown twice within the same block to test recall and recognition, with no filler image repetitions, totalling 100 images per block. Participants were encouraged to minimize false alarms, with the system allowing up to three 'red flags' before preventing progression to subsequent blocks, balancing challenge and feasibility given the study's scale and participant pool.

### *E. Participant of our research*

Our study recruited participants aged between 18 and 56 years; however, 6 participants could not pass any of the blocks in the visual memory game, whereas 243 participants passed at least one block. Of the 249 participants, 142 are male, 76 are female, and 31 have not disclosed their gender. The average age of participants is 28.2. Our research was conducted in April 2022 using an exclusive visual memory game.

## III. RESULTS

### *A. SemMem Dataset*

In our research, we randomly selected 170 images per main category from the MemCat [5] dataset, aiming to choose images representative of subjects unfamiliar to Iranians, such as various foods or beverages. Additionally, we incorporated 30 images related to Iranian life, including Iranian cars, animals, and places into each category. Our dataset is named SemMem. SemMem comprises five categories, each containing 200 images, totaling 1000 images. In Table I, we compare our dataset with MemCat and LaMem [15].

TABLE I
Characteristics and Attributes Comparison of the LaMem, MemCat, and SemMem Datasets. The table highlights the categorical organization, subcategorization, dataset size, and the presence of annotation details for each dataset.

| | LaMem | MemCat | SemMem |
|---|---|---|---|
| Category based | No | Yes | Yes |
| By subcategory | No | Yes | No |
| Number of images | ~ 60,000 | 10,000 | 1,000 |
| Bounding boxes or segmentation data | No | Yes | No |

The SemMem dataset categories include animals, food, landscapes, sports, and vehicles. In the images, there are no subcategories or bounding boxes.

### *B. Image memorability rank calculation*

We use the primary memory rank measurement introduced in LaMem. The hit rate is the number of correct recognitions, and $N_{resp}$ is the number of images displayed to all participants in a visual memory game. The average $N_{resp}$ in this study was

37. Memorability ranks range from 0 to 1.

$$\text{image memorability rank} = (\text{Hit Rate})/N_{resp} \ [15] \quad (2)$$

In our research, the average memorability score was 0.519, with a standard deviation of 0.09. For consistency analysis, we performed an analysis based on averaging over 25 random split-half trials and obtained an average of 0.54 for Spearman's rank correlation. The comparison of the results of our dataset with those of MemCat is presented in Table II.

TABLE II

Statistical Comparison Between the SemMem and MemCat Datasets. The table presents a summary of the key memorability statistics for each dataset, revealing that MemCat, with its larger sample size, exhibits a broader memorability range and higher average scores. The correlation coefficient (ρ) indicates a stronger consistency in memorability features within MemCat compared to SemMem. These insights highlight the potential variability in memorability based on image content and dataset characteristics.

| | MemCat | SemMem |
|---|---|---|
| Number of images | 10,000 | 1,000 |
| Minimum | 0.205882 | 0.269231 |
| Maximum | 1.0 | 0.913043 |
| Standard Deviation | 0.131130 | 0.090983 |
| Average | 0.759451 | 0.5139398 |
| ρ | 0.78 | 0.54 |

Fig. 1. illustrates the distribution of memorability scores across categories within the SemMem dataset, and Fig. 2. presents a comparison between local images and those from the older SemMem dataset.

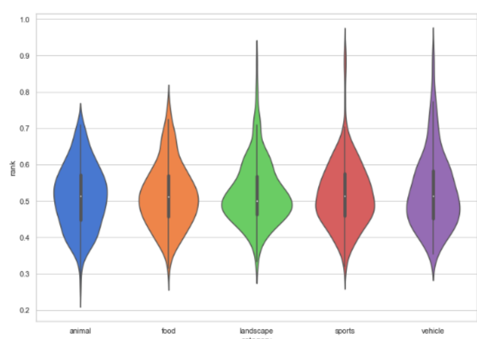


Fig 1: Distribution of Memorability Scores by Category. This figure illustrates the variation and density of memorability scores across different image categories, including animals, food, landscapes, sports, and vehicles. The median and distribution breadth category-specific memorability potential, with sports and vehicles showing slightly higher central tendencies, suggesting these categories may inherently possess more memorable attributes.

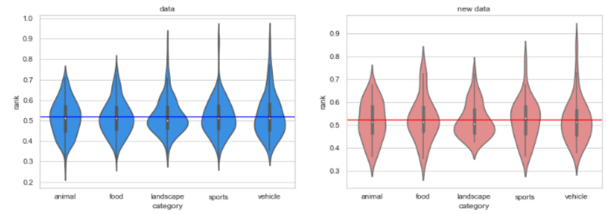


Fig. 2: Comparative Analysis of Image Memorability in the SemMem Dataset contrasts the 'Local Images' subset, selected for their cultural relevance to Iranian heritage and higher memorability, with the 'Old Images' subset, randomly sampled from the MemCat dataset. This comparison highlights the impact of cultural relevance on memorability, particularly evident in non-sports categories.

Fig. 3. displays a random sample of images from the SemMem Dataset, along with their memorability ranks.

Fig 3: Some images from the SemMem dataset with their memorability rank.

## C. *Comparison Memorable and forgettable category*

Our research presents a consolidated comparison of the memorable and forgettable categories within SemMem and MemCat. The comparative analysis includes both new and existing images in SemMem and indicates that local image features can significantly influence memorability. The results show variations in the memorable and forgettable categories when comparing SemMem to MemCat, underscoring the impact of image content on memory retention.

TABLE III

Comparison of Memorable and Forgettable Categories in SemMem and MemCat, Including an Analysis of New Images Versus the Entire Dataset.

| | MemCat | SemMem (total) | SemMem (new images) |
|---|---|---|---|
| Top 20 | Food | Vehicle | Animal |
| Top 50 | Food | Vehicle | Landscape |
| Bottom 20 | Landscape | Animal | Animal |
| Bottom 50 | Landscape | Animal | Animal |
| Memorable category (All) | Food | Vehicle | Food |
| Forgettable category (All) | Landscape | Animal | Animal |

In our analysis of memorability across different categories, Table III presents a clear contrast between the most and least memorable image categories within the MemCat and SemMem datasets. For example, while 'Food' consistently emerges as the most memorable category within the MemCat dataset, 'Vehicle' and 'Animal' categories take precedence in the SemMem dataset. Notably, 'Landscape' appears as the least memorable category across both datasets, suggesting a possible universal trend in image memorability. Furthermore, the new images added to SemMem demonstrate a shift in memorability, with the 'Animal' and 'Landscape' categories becoming more prominent. This indicates that the introduction of culturally relevant images can affect the memorability factors within a dataset.

### D. Predicting image memorability with ResNet architecture

We utilized convolutional neural network (CNN) models, ResNet-50 and ResNet-101, adapted to predict continuous memorability scores of images, diverging from their common use in classification. The choice of ResNet is due to its ability to address the vanishing gradient problem, maintaining efficacy in deeper network architectures. The Adam optimization algorithm facilitated efficient computation and memory management, essential for our dataset's scale.

To tailor the model for regression, we eliminated the activation functions from the final layer, traditionally used for classification tasks. This adjustment enables the network to output continuous scores, in line with our aim to emulate the human cognitive evaluation of image memorability. We measured prediction accuracy using the mean square error loss function.

The training was conducted on a MacBook Pro M1 2021 with 16GB unified memory, utilizing 80% of the MemCat dataset for training and the remaining 20% for validation. The training involved 25 epochs and a batch size of 128, with all input images resized to 224x224 pixels. Fig. 4. depicts the training and validation loss, providing insight into the learning trajectory. Initially, the networks were pre-trained on ImageNet [17] and subsequently trained on MemCat using Pytorch.

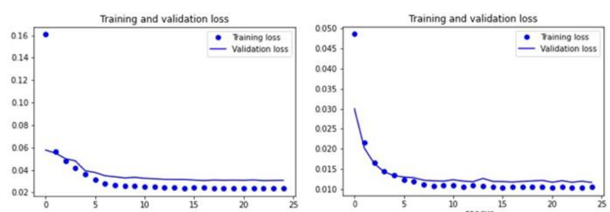


Fig. 4: Training and validation loss in ResNet 50 (left) and ResNet 101 (right)

Upon completing the training, the models were applied to the SemMem dataset, which comprises local Iranian images, to predict memorability scores. We then compared the predicted ranks with the actual memorability ranks to evaluate the models' prediction accuracy. Table IV presents this comparative analysis, showcasing the prediction error and training time for ResNet-50 and ResNet-101. The results demonstrate that ResNet-50 not only provided more accurate predictions but also required a shorter training duration, highlighting its efficiency in predicting image memorability.

TABLE IV

Comparison of Training Time and Prediction Error between ResNet 50 and ResNet 101.

| | ResNet 50 | ResNet 101 |
|---|---|---|
| Training time | 331 min, 53 sec | 449 min, 6 sec |
| Prediction error (MSE) | 0.080998 | 0.082507 |

## IV. DISCUSSION

In our study titled "Understanding Image Memorability through Localized Stimuli," we aimed to discover what distinguishes image memorability among people from different areas by studying it in Iranian adults. We found that some images in previous datasets in the image memorability field do not hold contextual significance for Iranians, minds, and we chose to focus on this demographic to address this issue. We stated that contextual meaning impacts image memory. We provided transparency and openness in our methodology, task development, changes in stimuli variables, participant selection and demographics, dataset description, and image memorability rank calculation. We also included details on our exclusive visual memory game, developed based on the Django-based Python programming language and validated participants by email or mobile number to prevent repeat participation.

We collected data from 249 participants aged between 18 and 56, with an average of 28.2, using our SemMem dataset, which includes 1000 images from five categories, each with 200 images, and contains Iranian-specific images. We calculated the image memorability rank based on the primary memory rank measurement introduced in LaMem and reported the average memorability score, standard deviation, and correlation analysis results.

We compared the SemMem dataset with MemCat and LaMem to identify differences in image categories, the number of images, and the presence of additional information like bounding boxes or segmentation. Notably, as reflected in Table III, we observed shifts in memorability within categories. The 'Vehicle' and 'Animal' categories became more memorable in the SemMem dataset compared to MemCat, where 'Food' was consistently the most memorable. Conversely, 'Landscape' was identified as less memorable across both datasets, indicating a possible universal perception of memorability. These observations suggest that the introduction of culturally relevant images can significantly influence memorability, highlighting the importance of dataset diversity in memory retention studies.

The improved memorability of culturally relevant, or 'local', images in the SemMem dataset is a significant finding of our research. Images that have cultural importance for the spectator elicit a deeper level of cognitive and emotional engagement, which can be the reason behind this phenomena. Images infused with local characteristics, such as historical sites, customary clothing, or activities unique to a culture, typically elicit a greater association memory response because they are more relevant and recognizable to the subject's experiences and cultural background. This aligns with the encoding specificity principle, which posits that memory is more effectively encoded and recalled when the contexts of encoding and retrieval are congruent. In the case of Iranian adults, local imagery resonates more profoundly, leading to higher memorability scores. This suggests that the contextual and cultural relevance of images plays a critical role in their memorability, underscoring the need to consider cultural diversity in the study of image memorability. Our findings highlight the significance of incorporating culturally specific stimuli in memorability research, demonstrating that the memorability of an image is not only a function of its intrinsic features but also of its cultural and contextual relevance to the observer.

Furthermore, we trained a CNN model using ResNet 50 and ResNet 101 to predict the memorability rank, and we utilized the learned model to forecast the SemMem pictures dataset. In summary, our study concentrated on the memorability of images among adult Iranians and presented our methods and findings with transparency and openness. We explained the rationale behind our chosen demographic and the inclusion of images unique to Iran in our dataset. Additionally, we included information about our unique visual memory game and how participants were verified. We also compared our dataset and results to previous studies and used ResNet architecture to predict image memorability.

## V. TRANSPARENCY AND OPENNESS

Methodology and Task Development: We clearly describe the development of our visual memory game, including the programming language used, stimuli publication, and accessibility. We explained the steps we took to ensure participant validation, data protection, and prevention of repeat participation. Additionally, we provide details about the server setup and security measures we implemented.

Changes in Stimuli Variables: We decided to modify the stimuli variables due to our relatively small dataset and the limited number of participants. We described the specifics of our visual memory game, including the duration of the image display, participant reactions, and the composition of different image categories.

Participant Selection and Demographics: We outlined the criteria for participant selection, including the age range and gender distribution. We provided the number of participants who passed our visual memory game blocks, as well as the average participant age.

Dataset Description: We introduced our SemMem dataset and its composition, including the random selection of images from the MemCat dataset and the addition of Iranian-specific images. We compared our dataset to MemCat and LaMem, highlighting the differences in categories, the number of images, and the availability of bounding box or segmentation data.

Image Memorability Rank Calculation: We described the calculation of image memorability rank based on the primary memory rank measurement introduced in LaMem. We provided the formula we used and reported the average memorability score, standard deviation, and correlation analysis results.

We collected our data in April 2022.

## VI. DATA AVAILABILITY

The data and code for this article can be found here: https://github.com/amirshnll/image-meorability-based-on-local-images content.

## VII. CONCLUSION

A challenging area of research is making images memorable. For several years, camera quality has improved, and photography has become more critical. Previously, photographers were not satisfied with the quality of cameras and data storage for taking pictures, but a few years ago, semantic subjects became very important. Technology has made it easier for educational content to incorporate attractive visual elements. Image memorability is an exciting field due to the many challenges associated with the concept of imagememorability.

In this study, we observed that local images significantly impact memorability, suggesting that the concept of memorability can vary among individuals. Our results indicate that while memorability is influenced by image features, it also reflects the viewer's cultural and personal context. By analyzing new data and performing memorability rankings in our SemMem dataset, we achieved a correlation coefficient (ρ) of 0.54. Typically, memorable images contain a salient object or area, often centrally located. This research underscores the multifaceted nature of image memorability, rooted in both the inherent qualities of the images and the diverse perceptions of viewers.

In future studies, we aim to delve deeper into the realm of image memorability, extending our research to include more specialized datasets, such as those focusing on local facial images. The collection of such datasets poses unique challenges, particularly concerning the need for explicit permissions due to privacy and ethical considerations.

Furthermore, we recommend enhancing the methodology of our Visual Memory Game by conducting it under controlled environmental conditions for all participants, employing eye-tracking technology and EEG monitoring to mitigate the effects of auditory distractions and variable lighting conditions. One of the primary hurdles we encountered in this study was the recruitment of participants. While our current research spanned a broad age range of 18 to 56 years, future investigations could benefit from focusing on narrower age cohorts to discern any age-specific trends in image memorability.

The potential for expanding the scope of memorability research is vast. Beyond the realm of static images, there lies an opportunity to explore the memorability of videos [20], a relatively untapped area that could offer new insights into how dynamic content is processed and remembered by different demographics.

In addition to diversifying the types of content studied, future research should also consider the impact of the data collection environment. While our study utilized online tools for data gathering, offering participants the option to partake in the visual memory task in a laboratory setting could standardize test conditions such as ambient brightness and screen quality.

Laboratory settings could also facilitate the observation of participants' facial expressions and the use of eye-tracking devices [18], providing richer data on the cognitive processes involved in memorability.

Lastly, our study's age demographic was quite broad, encompassing individuals aged 18 to 56. Future research might examine the effects of age on image memorability more closely by studying more homogenous age groups. This could reveal important insights into how memorability changes with age and identify age-specific factors that influence memorability.

By addressing these areas, future research can build upon our findings to develop a more nuanced understanding of image memorability, accounting for the intricate interplay between content characteristics, individual differences, and environmental factors.

## VIII. ACKNOWLEDGMENT

The authors acknowledge Semnan University. Also, the authors would like to thank the editor and reviewers for their constructive comments.

## IX. REFERENCES

[1] Isola, P., Parikh, D., Torralba, A., Oliva, A, "Understanding the intrinsic memorability of images," Advances in neural information processing systems, p. 2429–2437, 2011.
[2] Phillip Isola, Jianxiong Xiao, Antonio Torralba, Aude Oliva, "What makes an image memorable?," in CVPR 2011, 2011.
[3] Nicole C. Rust, Vahid Mehrpour, "Understanding Image Memorability," Trends in Cognitive Sciences, p. 12, 2020.
[4] Isola, P., Xiao, J., Parikh, D., Torralba, A., Oliva, A., "What makes a photograph memorable?," in Pattern Analysis and Machine Intelligence, 2014.
[5] Lore Goetschalckx and Johan Wagemans, "MemCat: A new category-based image set quantified on memorability," PeerJ, 2019.
[6] Zoya Bylinskii, Lore Goetschalckx, Anelise Newman, Aude Oliva, "Memorability: An image-computable measure of information utility," 2021.
[7] C. Spearman, "Correlation calculated from faulty data," British Journal of Psychology, p. 271–295, 1910.
[8] Bainbridge, W.A., Isola, P., Oliva, A, "The intrinsic memorability of face photographs," Journal of Experimental Psychology, 2013.
[9] Xiao, J., Hays, J., Ehinger, K.A., Oliva, A., Torralba, A., "Sun database: Large-scale scene recognition from abbey to zoo," in computer society conference on computer vision and pattern recognition, 2010.
[10] Borkin, M.A., Vo, A.A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., Pfister, H, "WhattWhat makes a visualization memorable?," in Visualization and Computer Graphics, 2013.
[11] Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., Oliva, "Intrinsic and extrinsic effects on image memorability," Vision research, p. 165–178, 2015.
[12] Dubey, R., Peterson, J., Khosla, A., Yang, M.H., Ghanem, B, "What makes an object memorable" in the pieceeie international conference on computer vision, 2015.
[13] Akagunduz, E., Bors, A., Evans, K, "Defining image memorability using the visual memory schema," in pattern analysis and machine intelligence, 2019.
[14] Lu, J., Xu, M., Yang, R., Wang, Z, "Understanding and predicting the memorability of outdoor natural scenes," p. 4927–494, 2020.
[15] Khosla, A., Raju, A.S., Torralba, A., Oliva, "Understanding and predicting image memorability at a large scale," in Proceedings of the IEEE International Conference on Computer, 2015.
[16] He, K., Zhang, X., Ren, S., Sun, J, "Deep residual learning for image recognition," in The IEEE ConferenceConference on Computer Vision and pattern recognition, 2015.
[17] Deng, J., Dong, W., Socher, R., Li, L., Kai Li, Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in IEEE Conference on Computer Vision and Pattern Recognition, 2009.
[18] Võ, M.L.H., Bylinskii, Z., Oliva, A, "Image memorability in the eye of the beholder: Tracking the decay of visual scene representations," bioRxiv, 2017.
[19] Perera, S., Tal, A., Zelnik-Manor, L., "Is image memorability prediction solved?," in Computer Vision and Pattern Recognition Workshops, 2019.
[20] Cohendet, R., Demarty, C., Duong, N.Q.K., Martin, E, "VideoMem: Constructing, Analyzing, Predicting Short-Term and Long-Term Video Memorability," in international conference on computer vision, 2019.
[21] Cohendet, R., Yadati, K., Duong, N.Q., Demarty, C.H, "Annotating, Understanding, and Predicting Long-Term Video Memorability," in International Conference on Multimedia Retrieval, 2018.
[22] Newman, A., Fosco, C., Casser, V., Lee, A., Barry, Mcnamara, Oliva, A, "Multimodal morability: Modeling effects of semantics and decay on video memorability," in ECCV, 2020.