

Modeling Visual Memorability Assessment with Autoencoders Reveals Characteristics of Memorable Images

Authors : Elham Bagheri, Yalda Mohsenzadeh

Abstract : Image memorability refers to the phenomenon where certain images are more likely to be remembered by humans than others. It is a quantifiable and intrinsic attribute of an image. Understanding how visual perception and memory interact is important in both cognitive science and artificial intelligence. It reveals the complex processes that support human cognition and helps to improve machine learning algorithms by mimicking the brain's efficient data processing and storage mechanisms. To explore the computational underpinnings of image memorability, this study examines the relationship between an image's reconstruction error, distinctiveness in latent space, and its memorability score. A trained autoencoder is used to replicate human-like memorability assessment inspired by the visual memory game employed in memorability estimations. This study leverages a VGG-based autoencoder that is pre-trained on the vast ImageNet dataset, enabling it to recognize patterns and features that are common to a wide and diverse range of images. An empirical analysis is conducted using the MemCat dataset, which includes 10,000 images from five broad categories: animals, sports, food, landscapes, and vehicles, along with their corresponding memorability scores. The memorability score assigned to each image represents the probability of that image being remembered by participants after a single exposure. The autoencoder is finetuned for one epoch with a batch size of one, attempting to create a scenario similar to human memorability experiments where memorability is quantified by the likelihood of an image being remembered after being seen only once. The reconstruction error, which is quantified as the difference between the original and reconstructed images, serves as a measure of how well the autoencoder has learned to represent the data. The reconstruction error of each image, the error reduction, and its distinctiveness in latent space are calculated and correlated with the memorability score. Distinctiveness is measured as the Euclidean distance between each image's latent representation and its nearest neighbor within the autoencoder's latent space. Different structural and perceptual loss functions are considered to quantify the reconstruction error. The results indicate that there is a strong correlation between the reconstruction error and the distinctiveness of images and their memorability scores. This suggests that images with more unique distinct features that challenge the autoencoder's compressive capacities are inherently more memorable. There is also a negative correlation between the reduction in reconstruction error compared to the autoencoder pre-trained on ImageNet, which suggests that highly memorable images are harder to reconstruct, probably due to having features that are more difficult to learn by the autoencoder. These insights suggest a new pathway for evaluating image memorability, which could potentially impact industries reliant on visual content and mark a step forward in merging the fields of artificial intelligence and cognitive science. The current research opens avenues for utilizing neural representations as instruments for understanding and predicting visual memory.

Keywords : autoencoder, computational vision, image memorability, image reconstruction, memory retention, reconstruction error, visual perception

Conference Title : ICCV 2024 : International Conference on Computational Vision

Conference Location : Venice, Italy

Conference Dates : June 20-21, 2024