Differential Behavioral Manifestations in Emotional versus Neutral Scene Perception within Convolutional Neural Networks

Chun-Hui Li (iamscli.tw@gmail.com)
Department of Psychology, National Taiwan University
Taipei, 106319 Taiwan

Szu-Chi Chen (b11207089@ntu.edu.tw)
Department of Psychology, National Taiwan University
Taipei, 106319 Taiwan

Bo-Cheng Kuo (bckuo@ntu.edu.tw)
Department of Psychology, National Taiwan University
Taipei, 106319 Taiwan

Abstract:

Numerous studies have highlighted the impact of emotional information on visual perception. However, the extent to which emotional information is encoded and processed in visual representations when viewing scenes remains unclear. Here, we conducted representational similarity and variance partitioning analyses to explore the visual representations of scenes containing emotional and neutral information in convolutional neural networks. Our results indicated an increasing similarity between emotion and VGG-16 RDMs, starting from the third convolutional layer, with higher similarities observed for negative images compared to neutral ones. Moreover, variance partitioning results showed that emotion model exhibited an increasing trend in explained variance from shallow to deep layers, whereas color model revealed a converse pattern of decreasing variance. Importantly, emotion model displayed significantly higher unique explanatory power for VGG-16 RDMs when comparing negative images to neutral ones beginning at the fourth layer, suggesting emotional enhancement within visual representations. Overall, our findings demonstrated the hierarchical integration of emotional information within the visual representation underlying scene perception.

Keywords: scene perception; emotion; convolutional neural networks; representational similarity analysis

Introduction

Emotional information from the external environment has been shown to influence the processing of visual perception (Zadra & Clore, 2011). Neuroimaging studies have demonstrated that emotional information affects not only neural responses in the limbic areas but also the visual cortex (Kuo et al., 2018; Saarimäki et al., 2016; Sambuco et al., 2020). However, little is known about the extent to which emotional information is encoded and processed in visual representations when viewing scenes.

This study aims to investigate scene perception, specifically focusing on emotional and neutral information, using convolutional neural networks (CNNs). Employing representational similarity analysis (RSA), we integrated the behavioral performance of emotional ratings with the CNNs to access the layer-wise visual information (Cichy et al., 2014; Dobs et al., 2023). Moreover, we utilized variance partitioning analysis (VPA) to explore the unique contributions of emotional and low-level features to the total variance of visual representations within CNNs (Groen et al., 2018; Jozwik et al., 2023). Our findings provide novel evidence on how visual representation underlying scene perception hierarchically integrates emotional information.

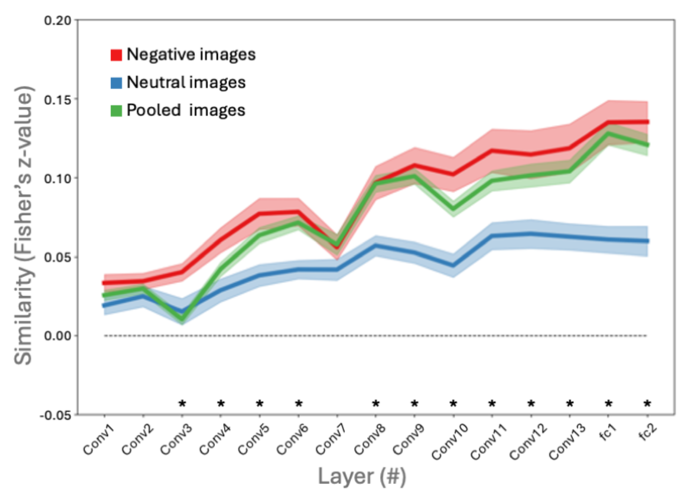

Figure 1: Layer-wise similarities between the RDMs of emotional ratings and VGG-16 model. The asterisks on the bottom of the figure indicate significant emotional enhancement on representational similarities.

Method

Stimuli consisted of the natural images including 192 negative and 192 neutral scenes. Participants (N = 35) were recruited to rate the emotional scores of 384 scene stimuli based on a two-dimensional valence-arousal space. Next, we employed the VGG-16 model in PyTorch, which had been trained on the ImageNet dataset for object recognition (Simonyan & Zisserman, 2014), to derive the layer-wise visual representations for each scene image.

Representational Similarity Analysis

For behavioral and VGG-16 model measurements, pairwise dissimilarities across all combinations of scenes were computed via Euclidean distance and one minus Pearson's correlation coefficient respectively, yielding representational dissimilarity matrices (RDMs). Spearman correlation coefficients were then calculated between the RDM for each participant's ratings and the RDM of each layer of the VGG-16 model, offering a hierarchical profile of emotional information within the visual processing hierarchy. Furthermore, we divided the 384 x 384 pooled RDM into two 192 x 192 RDMs for negative and neutral scenes, to examine the effect of emotional information on representational similarity.

Variance Partitioning Analysis

To examine the contributions of emotional and low-level features, we further adopted color and gist models to extract visual features for each scene image (El-Gayar et al., 2013; Oliva & Torralba, 2001). Pairwise dissimilarities across all scenes were used to construct RDMs for both color and gist models. The RDM for emotion model was identical with the RDM for each participant's ratings. Initially, we simultaneously fitted the RDMs of emotion, color, and gist models to predict VGG-16 RDM for each layer (full GLM), as well as fitted two of the three models individually (reduced GLMs). The unique variances explained by each model were then estimated by contrasting the variance explained by the full GLM with that explained by the reduced GLM. The procedure was carried out separately for each RDM of negative, neutral and pooled scene images.

Statistical Inference

In the results of RSA and VPA, statistical tests at the group level were conducted to test the layer-wise difference between negative versus neutral scenes using dependent samples t-tests. Multiple comparisons were corrected using false discovery rate of 0.05.

Results

By integrating the VGG-16 and emotion rating data via RSA, we first accessed the degree to which emotional information was processed within visual representations across layers of VGG-16. The similarity results of negative, neutral and pooled images are illustrated in Figure 1. We observed a gradual rise in similarities in pooled images starting from the third layer. Moreover, significantly higher similarities were found in negative images compared to neutral images, beginning from the third layer and persisting up to the final layer. The results suggested that emotional information is represented within the visual processing of VGG-16.

Importantly, we examined the contribution of emotion and other low-level visual features to visual representations based on the unique variance explained by each RDM of emotion, color and gist models using VPA. These results are illustrated in Figure 2. Variance partitioning results (Fig. 2A) showed that color model outperformed both gist and emotion model in explaining visual representation in the first

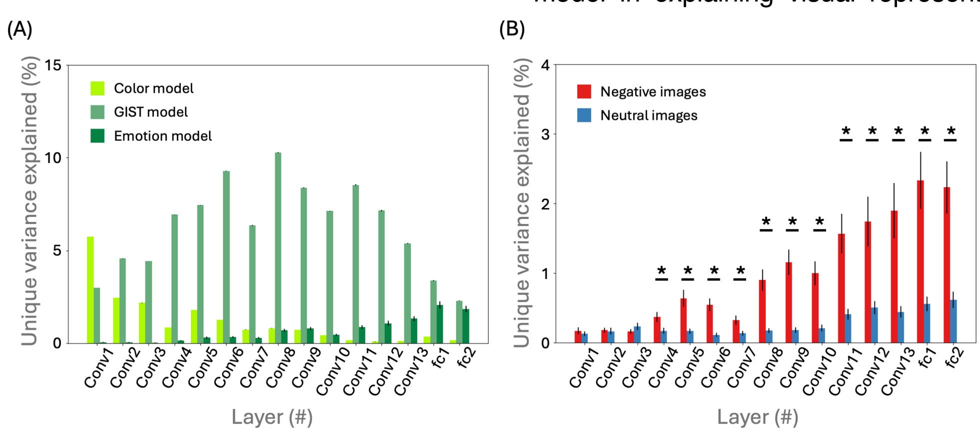

Figure 2: (A) Unique explained variance for color, gist and emotion models in pooled images. (B) Unique explained variance for emotion model in negative and neutral images. The asterisks indicate significant emotional enhancement on representational similarities.

layer of the VGG-16. By the eighth layer, the performance of gist model peaked and subsequently declined gradually across subsequent layers. Importantly, compared to neutral images, emotion model exhibited significantly higher unique explanatory power for VGG-16 RDMs in negative images, beginning at the fourth layer (Fig. 2B). Together, these results indicated emotional enhancement within visual representations of VGG-16.

Conclusion

Our findings of RSA and VPA illustrate the hierarchical processing of emotional information within the layer-wise visual representation of VGG-16. This study provides novel evidence supporting a hierarchical organization of emotional information within the coarse-to-fine visual processing pathway.

Acknowledgments

This work was supported by grants from National Science and Technology Council, Taiwan (110-2410-H-002-129-MY3 and 113-2811-H-002-006).

References

Cichy, R., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462.

Dobs, K., Yuan, J., Martinez, J., & Kanwisher, N. (2023). Behavioral signatures of face perception emerge in deep neural networks optimized for face recognition. *Proceedings of the National Academy of Sciences*, *120*(32), e2220642120.

El-Gayar, M. M., Soliman, H., & others. (2013). A comparative study of image low level feature extraction algorithms. *Egyptian Informatics Journal*, *14*(2), 175–181.

Groen, I. I. A., Greene, M. R., Baldassano, C., Li, F.-F., Beck, D. M., & Baker, C. I. (2018). Distinct contributions of functional and deep neural network features to representational similarity of scenes in human brain and behavior. *Elife*, *7*, e32962.

Jozwik, K. M., Kietzmann, T. C., Cichy, R. M., Kriegeskorte, N., & Mur, M. (2023). Deep neural networks and visuo-semantic models explain complementary components of human ventral-stream representational dynamics. *Journal of Neuroscience*, *43*(10), 1731–1741.

Kuo, B.-C., Lin, S.-H., & Yeh, Y.-Y. (2018). Functional interplay of top-down attention with affective codes during visual short-term memory maintenance. *Cortex*, *103*, 55–70. https://doi.org/10.1016/j.cortex.2018.02.003

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*, 145–175.

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I. P., Lampinen, J., Vuilleumier, P., Hari, R., Sams, M., & Nummenmaa, L. (2016). Discrete neural signatures of basic emotions. *Cerebral Cortex*, *26*(6), 2563–2573.

Sambuco, N., Bradley, M. M., Herring, D. R., & Lang, P. J. (2020). Common circuit or paradigm shift? The functional brain in emotional scene perception and emotional imagery. *Psychophysiology*, *57*(4), e13522.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*.

Zadra, J. R., & Clore, G. L. (2011). Emotion and perception: The role of affective information. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(6), 676–685.