Using Artificial Neural Networks to Understand Fluency in the Perception of Paintings

Yi Lin (yi.lin@kuleuven.be)
KU Leuven, Belgium

Johan Wagemans (johan.wagemans@kuleuven.be)
KU Leuven, Belgium

Hans Op de Beeck (hans.opdebeeck@kuleuven.be)
KU Leuven, Belgium

Abstract

Fluency—the ease with which an image is processed—is closely linked to aesthetic appraisal. However, existing objective measures of fluency fail to accurately capture subjective ratings. Recently, deep convolutional neural networks (DCNNs) have been proposed as a tool for measuring integration, a concept linked with fluency, using natural scenes as stimuli. Yet, the direct link between fluency and integration remains untested, and it is unclear whether these findings generalize to art perception. In this work, we investigate (1) whether the integration measure via DCNNs effectively captures fluency, potentially outperforming existing methods in the context of art perception, and (2) if DCNNs provide a superior measure of fluency, what specific mechanisms they reveal. Our findings indicate that the DCNN-based integration measure captures subjective fluency well and significantly outperforms other objective fluency measurements. Additionally, we observed that the peak correlation between DCNN-derived integration and various visual characteristics—intended to quantify different aspects of fluency—occurs at different DCNN layers. This suggests that fluency may be a multi-level process, integrating distinct visual characteristics at various processing stages. In summary, a DCNN-based measure of integration provides valuable insights into the concept of fluency.

Keywords: Visual processing; Fluency; Aesthetic Perception

Introduction

Aesthetic experience is a dynamic process shaped by interactions between image properties (e.g., color, symmetry) and personal factors (e.g., experience). This phenomenon has long fascinated researchers across disciplines, yet key questions remain. One central aspect is fluency—the ease with which an image is processed.

Fluency arises from the interaction between the image and the perceiver and includes perceptual fluency (ease of processing sensory features) and conceptual fluency (ease of understanding meaning) (Reber et al., 2004). It is linked to aesthetic appraisal (Graf & Landwehr, 2015; Pelowski et al., 2017; Nara & Kaiser, 2024), yet its underlying mechanism remains unclear.

While certain features, such as complexity, may be linked to fluency, calculating these features independently does not fully capture how humans process fluency. A promising approach to studying fluency involves using biologically inspired vision models to develop objective measures that simulate fluency ratings. Once validated, these measures can provide deeper insights into the mechanisms underlying fluency perception.

Recently, Nara and Kaiser (2024) introduced visual integration, proposing that when image elements are combined into a meaningful whole rather than perceived as separate components, the visual system simplifies representations, leading to more fluent, efficient processing and an enhanced feeling of beauty.

They quantified integration by measuring the sign-inverted correlation between a DCNN layer's activation pattern for the whole image and the average activation of its two halves. A lower correlation (before inversion) indicated higher integration and fluency.

However, Nara and Kaiser (2024) only reported the correlation between integration and beauty in natural scene images, and did not measure fluency. This highlights two key points: (1) While fluency is conceptually linked to the DCNN-based integration measure, there is no direct evidence confirming that this measure reliably captures fluency, nor is it clear whether the proposed measurements can be applied to painting perception. (2) It also remains unclear how the measure operates across different DCNN layers.

Here, we address these unknowns by testing the method of Nara and Kaiser (2024) on a dataset of Western art paintings for which human fluency ratings are available (https://osf.io/cufnj/; Lin, Op de Beeck, & Wagemans, 2024). The focus of the study includes (1) comparing subjective fluency with DCNN-based integration and (2) investigating the potential fluency processing mechanisms of the DCNN.

Results & Conclusions

1. Capturing Fluency with DCNN Integration Measures

As a first step, we evaluated how accurately the DCNN-based measurement captures fluency. To achieve this, we used the art image set, Leuven Orthogonalized Art Dataset (LOAD), an open-source database (https://osf.io/cufnj/; Lin et al., 2024). LOAD includes 343 paintings spanning nine major art styles or movements with diverse content, along with various behavioral ratings, including fluency. Each painting was rated by 50 or 51 participants.

Following Nara and Kaiser's method, we input the 343 paintings into VGG16, pretrained on ImageNet, and extracted the activation patterns from each layer. Notably, integration is computed as the sign inverse of the correlation between the activation of full images and the average activation of their two halves. During computation, the two halves of an image can be created using different division sizes. Specifically, the size of the individual mask for each half can vary, including divisions into 2 × 2, 4 × 4, 8 × 8, 16 × 16, or 32 × 32 identical squares (top legend in Fig. 1). Changing the size can enhance the depth at which spatial granularity influences fluency processing.

We found a strong correlation[1] between the DCNN-based integration measure and subjective fluency (Fig. 1), with the highest correlations reaching r=0.6. Moreover, finer spatial scales (e.g., 32 × 32) exhibited a stronger correlation with fluency. To evaluate how well the DCNN-derived integration measure compares to other fluency-related visual characteristics obtained through objective algorithmic measurements, we used the R package *imagefluency* (Mayer & Landwehr, 2018) to compute five additional visual characteristics associated with fluency: visual simplicity, visual symmetry (measured along both the horizontal and vertical axes), visual contrast, and visual self-similarity (i.e., the extent to which zooming in and out of an image reveals the same repeating visual pattern). We found relatively low correlations between these fluency-related visual characteristics and subjective fluency ratings: visual simplicity (r=.37, p<.001), vertical symmetry (r=.34, p<.001), horizontal symmetry (r=.26, p<.001), self-similarity (r=.13, p=.023), and contrast (r=.00, p=0.95).

In summary, these analyses suggest that Nara and Kaiser's approach may serve as a more effective tool for measuring fluency than other available metrics.

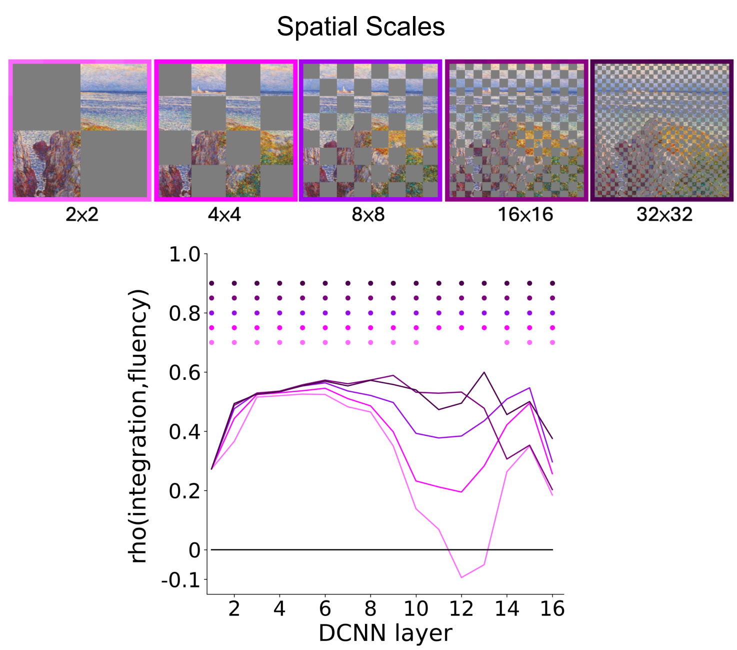

Fig. 1: Integration-subjective fluency correlations

[1] FDR correction was applied to all analyses.

2. Fluency Processing Through the Lens of DCNNs: What Each Layer Reveals?

While DCNNs can serve as a tool for measuring fluency in a faster and more objective manner, the question remains how this measure of fluency evolves across the different DCNN layers, given that all layers can show correlations with subjective fluency.

We examined the correlations between integration values and the four visual characteristics mentioned earlier, which exhibited significant correlations with subjective fluency (Fig. 2). Strong correlations for simplicity were found in early to intermediate layers. Symmetry showed similar trends. Self-similarity, involving both high- and low-resolution processing, had its strongest correlations in intermediate to late-intermediate layers. The results suggest that different aspects of fluency are quantified at distinct DCNN layers, with processing complexity increasing at deeper layers. Moreover, they imply that human fluency processing involves integrating multiple visual features at various stages. In short, the DCNN-based approach with its multiple layers provides a unique tool to capture all these different stages of processing.

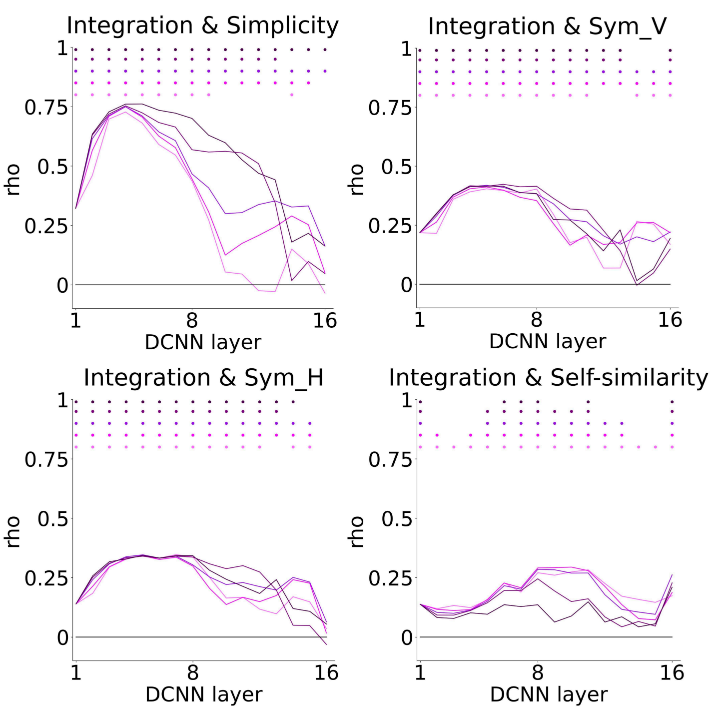

Fig. 2: Integration-visual property correlations. Sym_V denotes symmetry along the vertical axis, while Sym_H denotes symmetry along the horizontal axis. The colors denote spatial scales, as referenced in Fig. 1.

References

Graf, L. K., & Landwehr, J. R. (2015). A dual-process perspective on fluency-based aesthetics: The pleasure-interest model of aesthetic liking. *Personality and Social Psychology Review, 19*(4), 395–410.

Lin, Y., Op de Beeck, H., & Wagemans, J. (in press). Leuven Orthogonalized Art Dataset (LOAD): A multidimensional art image set for aesthetic appreciation research. *Psychology of Aesthetics, Creativity, and the Arts*.

Mayer, S., & Landwehr, J. R. (2018). Quantifying visual aesthetics based on processing fluency theory: Four algorithmic measures for antecedents of aesthetic preferences. *Psychology of Aesthetics, Creativity, and the Arts, 12*(4), 399.

Nara, S., & Kaiser, D. (2024). Integrative processing in artificial and biological vision predicts the perceived beauty of natural images. *Science Advances, 10*(9), eadi9294.

Pelowski, M., Markey, P. S., Forster, M., Gerger, G., & Leder, H. (2017). Move me, astonish me… delight my eyes and brain: The Vienna Integrated Model of top-down and bottom-up processes in art perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates. *Physics of Life Reviews, 21*, 80–125.

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review, 8*(4), 364–382.