The Representational Organization of Static and Dynamic Visual Features in the Human Cortex

Hamed Karimi (karimike@bc.edu)
Boston College, Boston, MA 02467, USA

Jeff Wang (jw162@rice.edu)
Boston College, Boston, MA 02467, USA

Stefano Anzellotti (stefano.anzellotti@bc.edu)
Boston College, Boston, MA 02467, USA

Abstract

Visual information consists of static and dynamic properties. How is their representation organized in the visual system? Static information has been associated with ventral temporal regions and dynamic information with lateral and dorsal regions. However, investigating the representation of static and dynamic information is complicated by the correlation between static and dynamic features of continuous visual input. Recent work addressed this challenge by using point-light displays and kinematograms to isolate motion information, but such stimuli might not capture the rich dynamics contained in realistic videos. Here, we separated static and dynamic features in realistic videos using two-stream deep convolutional neural networks and used them in conjunction with fMRI and representational similarity analysis to investigate the representation of static and dynamic information in the visual system. First, consistent with recent findings, we showed that both static and dynamic features are encoded across all visual streams. Second, we found that brain streams represent shared as well as unique static and dynamic visual information.

Keywords: action recognition; visual pathways; static and dynamic visual features; deep neural networks;

Introduction

The visual system is organized into distinct streams (Ungerleider, Mishkin, et al., 1982; Pitcher & Ungerleider, 2021); the ventral stream has been proposed to encode *static* object identity (Issa, Cadieu, & DiCarlo, 2018; Logothetis, Pauls, & Poggio, 1995; Grill-Spector & Weiner, 2014), while *dynamic* information has been associated with the lateral and dorsal streams (Ganel & Goodale, 2003; Culham et al., 2003; Pitcher, Duchaine, & Walsh, 2014). Nonetheless, this evidence does not rule out the possibility that ventral regions might also encode some dynamic information. To what extent does the dissociation between visual streams correspond to differences in their representation of static vs. dynamic visual features?

Ventral regions show responses to static objects that can be explained using image statistics (Rose, Johnson, Wang, & Ponce, 2021; Doshi & Konkle, 2023). Yet, object identity information, generally associated with the ventral stream, can be recognized even in the absence of informative static features: participants can categorize objects from structured movement depicted by point-light displays (Mather & West, 1993; Vuong, Friedman, & Read, 2012). Similar stimuli can also support action recognition (Alaerts, Nackaerts, Meyns, Swinnen, & Wenderoth, 2011; Dittrich, Troscianko, Lea, & Morgan, 1996). Together, this evidence suggests that dynamic information plays a role in tasks traditionally associated with the ventral as well as the lateral and dorsal streams. Others have isolated motion signals (Robert, Ungerleider, & Vaziri-Pashkam, 2023) as well as motion direction (Ramezanpour, Ilic, Wildes, & Kar, 2024) in ventral stream responses, and parallel work identified static shape information in dorsal stream regions (Freud, Culham, Plaut, & Behrmann, 2017).

We hypothesize that all streams encode both static and dynamic visual features, with differences in content that depend on each stream's functional role. In this study, we first present a systematic investigation of the unique contributions of static and dynamic visual features, using different deep neural networks (DNNs) trained either with or without supervision. We compared the deep networks' internal representation with human fMRI responses to naturalistic videos (Forrest Gump movie) to quantify how accurately each model can account for the neural activity across regions in different visual streams. Based on a probabilistic atlas of brain regions (Wang, Mruczek, Arcaro, & Kastner, 2015), visual streams were subdivided into ventral, dorsal, lateral, and parietal.

Our results show that all streams represent both static and dynamic features (even after controlling for features of the other type). We also find that different pairs of streams represent both shared and unique static and dynamic information, indicating that representational content is shaped by the streams' unique functional roles.

Methods

Stimuli and Neural Data

BOLD fMRI responses (3×3×3 mm) to the movie 'Forrest Gump' were obtained from the *studyforrest* audiovisual dataset (Hanke et al., 2016) (`http://studyforrest.org`).

Hidden two-stream Convolutional Neural Network

We used the models in (Zhu, Lan, Newsam, & Hauptmann, 2019) and trained three DNNs separately, to encode static and dynamic features: A supervised (sup) static net that predicts action labels from a single frame, an early unsupervised (unsup) dynamic net that reconstructs a future frame by inferring the optic flow from preceding frames by minimizing an unsupervised learning objective, and finally a late supervised (sup) dynamic net which predicts action labels from optic flows extracted by the early (unsup) dynamic net. The representational similarity was computed with Pearson correlation.

Results

All streams represent static and dynamic features

We measured the similarity between representations in each deep neural network's layer with each brain stream. Figure 1a shows that static features correlated with the ventral stream. They also correlated with dorsal, lateral, and parietal streams. Dynamic features correlated with dorsal, lateral, and parietal brain streams as well. Critically, dynamic features also correlated with the ventral stream, suggesting that ventral regions do not encode exclusively static information (all p-values significant at Bonferroni-corrected thresholds of 0.001).

Static and dynamic features might covary (e.g., cars look in a certain way and also tend to move in a certain way). Therefore, to test whether brain streams represent uncorrelated static and dynamic features, we measured the semi-partial correlation between a DNN's features and a brain stream, while controlling for the other two DNNs.

Figure 1b shows that the uncorrelated (sup) static and the early (unsup) dynamic models' features correlate with all the visual streams (significant at Bonferroni-corrected p-values of 0.001). The similarity between the late (sup) dynamic model and brain streams can be accounted for by the (sup) static and early (unsup) dynamic features, showing a drop in the correlation between the late (sup) static model and the neural responses in all of the brain streams.

This indicates that the correlation of visual streams with either static or dynamic visual features cannot be fully accounted for by the covariation between dynamic features and static features.

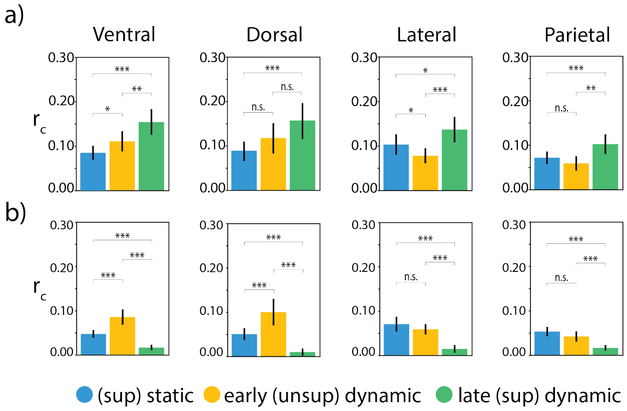

Figure 1: Similarities between models' features and fMRI responses in brain streams. Each bar is the sum of the semi-partial Pearson correlation between neural activity and a DNN's layers, a) while controlling for all the model's previous layers' features, and b) while additionally controlling for the other two DNNs (t-tests were conducted at Bonferroni-corrected threshold for 0.05, 0.01, 0.001).

Streams represent unique as well as shared static and dynamic information

We expect different brain streams to perform different functions and thus to encode different information. While all streams represent both static and dynamic features they might not encode the same static and dynamic information. To investigate this, we measured how well the uncorrelated visual features account for each stream's responses while additionally controlling for the neural responses in the other streams (Figure 2).

As expected, the ventral stream represents unique as well as shared static features with the dorsal stream. We also found that lateral and parietal brain streams share dynamic visual features with both the ventral and the dorsal streams while having their own unique static features. Critically, we observed that after controlling for the ventral stream in the dorsal stream, the correlation with the (sup) static model persists ($p < 0.001$, Bonferroni-corrected), indicating that the dorsal brain stream encodes static features that are not captured by the ventral stream. Additionally, the ventral stream remained significantly correlated with the early (unsup) dynamic model, even after regressing out the dorsal stream. This means that the ventral stream represents dynamic features that are not captured by the dorsal stream.

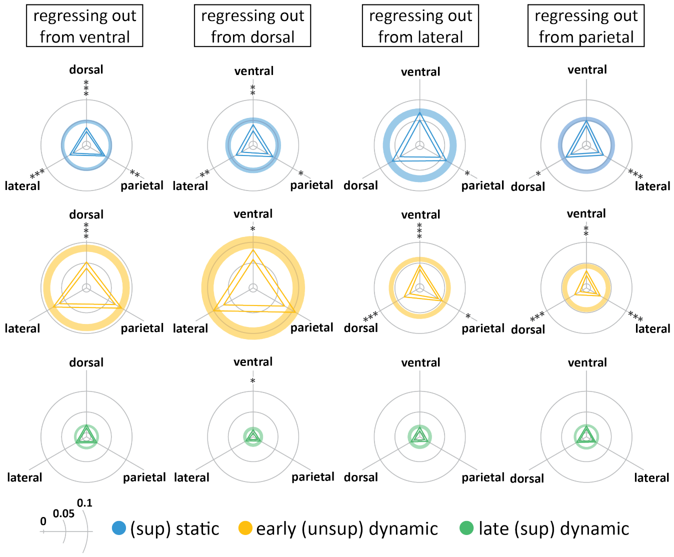

Figure 2: Shared vs. unique representation of static and dynamic across pairs of brain streams. The colored ring displays the representational similarity between a visual stream (noted in the column title) and a DNN, controlling for the other DNNs. The triangles display the representational similarity after additionally controlling for one other stream (noted on the outside of the circle). (*=.05, **=0.01, ***=0.001, Bonferroni corrected)

Discussion

We investigated systematically the representation of static and dynamic information in different visual streams, finding that 1) all streams encode both static and dynamic information, 2) distinct streams encode shared as well as unique static and dynamic features.

These results show that different visual streams cannot be distinguished based on the presence or absence of static or dynamic information, but rather that they differ in terms of the kinds of static and dynamic information they encode. The relationship between each stream's functional role and the kinds of features it represents will need to be elucidated in future studies.

Acknowledgments

We would like to thank the researchers in the StudyForrest project for curating the dataset.

References

Alaerts, K., Nackaerts, E., Meyns, P., Swinnen, S. P., & Wenderoth, N. (2011). Action and emotion recognition from point light displays: an investigation of gender differences. *PloS one*, *6*(6), e20989.

Culham, J. C., Danckert, S. L., Souza, J. F. D., Gati, J. S., Menon, R. S., & Goodale, M. A. (2003). Visually guided grasping produces fmri activation in dorsal but not ventral stream brain areas. *Experimental brain research*, *153*, 180–189.

Dittrich, W. H., Troscianko, T., Lea, S. E., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, *25*(6), 727–738.

Doshi, F. R., & Konkle, T. (2023). Cortical topographic motifs emerge in a self-organized map of object space. *Science Advances*, *9*(25), eade8187.

Freud, E., Culham, J. C., Plaut, D. C., & Behrmann, M. (2017). The large-scale organization of shape processing in the ventral and dorsal pathways. *elife*, *6*, e27576.

Ganel, T., & Goodale, M. A. (2003). Visual control of action but not perception requires analytical processing of object shape. *Nature*, *426*(6967), 664–667.

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*(8), 536–548.

Hanke, M., Adelhöfer, N., Kottke, D., Iacovella, V., Sengupta, A., Kaule, F. R., ... Stadler, J. (2016). A studyforrest extension, simultaneous fmri and eye gaze recordings during prolonged natural stimulation. *Scientific data*, *3*(1), 1–15.

Issa, E. B., Cadieu, C. F., & DiCarlo, J. J. (2018). Neural dynamics at successive stages of the ventral visual stream are consistent with hierarchical error signals. *Elife*, *7*, e42870.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current biology*, *5*(5), 552–563.

Mather, G., & West, S. (1993). Recognition of animal locomotion from dynamic point-light displays. *Perception*, *22*(7), 759–766.

Pitcher, D., Duchaine, B., & Walsh, V. (2014). Combined tms and fmri reveal dissociable cortical pathways for dynamic and static face perception. *Current Biology*, *24*(17), 2066–2070.

Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends in Cognitive Sciences*, *25*(2), 100–110.

Ramezanpour, H., Ilic, F., Wildes, R., & Kar, K. (2024). Object motion representation in the macaque ventral stream–a gateway to understanding the brain's intuitive physics engine. *bioRxiv*, 2024–02.

Robert, S., Ungerleider, L. G., & Vaziri-Pashkam, M. (2023). Disentangling object category representations driven by dynamic and static visual input. *Journal of Neuroscience*, *43*(4), 621–634.

Rose, O., Johnson, J., Wang, B., & Ponce, C. R. (2021). Visual prototypes in the ventral stream are attuned to complexity and gaze behavior. *Nature communications*, *12*(1), 6723.

Ungerleider, L. G., Mishkin, M., et al. (1982). Two cortical visual systems. analysis of visual behavior. *Ingle DJ, Goodale MA, Mansfield RJW*.

Vuong, Q. C., Friedman, A., & Read, J. C. (2012). The relative weight of shape and non-rigid motion cues in object perception: A model of the parameters underlying dynamic object discrimination. *Journal of Vision*, *12*(3), 16–16.

Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cerebral cortex*, *25*(10), 3911–3931.

Zhu, Y., Lan, Z., Newsam, S., & Hauptmann, A. (2019). Hidden two-stream convolutional networks for action recognition. In *Computer vision–accv 2018: 14th asian conference on computer vision, perth, australia, december 2–6, 2018, revised selected papers, part iii 14* (pp. 363–378).