Contrast inversion reveals hierarchical asymmetries of contrast processing in biological and artificial vision

Sohrab Najafian * SOHRAB_NAJAFIAN@HMS.HARVARD.EDU
Giordano Ramos-Traslosheros * GIO@HMS.HARVARD.EDU
Akshay V Jagadeesh
Margaret Livingstone
Harvard Medical School, Neurobiology Dept., Boston, USA

Editors: List of editors' names

Abstract

Contrast is a fundamental visual feature, encoded as early as the retina by segregated ON and OFF pathways. While these pathways are largely symmetric, subtle biases exist that shape perception and cortical responses. Here, we extend the study of contrast processing to color images across the hierarchy of the primate visual ventral stream and convolutional neural networks (CNNs). Using Neuropixels recordings from macaque V1 through IT, and contrast inversion we show that in a biological system contrast polarity is weakly encoded in early cortex but becomes stronger downstream, peaking in IT. Surprisingly, CNNs exhibit the opposite trend: contrast polarity is strongly represented in the first layer, lost in intermediate layers, and partially recovered later. Thus, early visual areas in the brain rely on local features symmetric to contrast inversion and this symmetry is broken in high-level visual areas. while the CNNs rely on asymmetric local and high-level features. This divergence reveals a fundamental asymmetry in how biological and artificial systems balance tolerance and sensitivity to strong out of distribution images, such as contrast inversion, as early as the first layer, providing new constraints for improving both neural models and machine vision.

Keywords: contrast polarity; ON/OFF pathways; inferior temporal cortex (IT); convolutional neural networks; ventral visual stream

1. Introduction

Biological contrast processing Humans and non-human primates are strongly sensitive to contrast polarity. For example, face recognition is impaired by inverting the contrast of the image (e.g., black to white and vice versa) Kobatake and Tanaka (1994); Nederhouser et al. (2007). Across many species, contrast information is computed as early as the second visual synapse. In primates, downstream of the photoreceptors, visual contrast signals are split into ON and OFF pathways, which encode light increments and decrements. These pathways remain segregated through the retina and LGN Schiller (1992), and many V1 simple cells respond selectively to one contrast polarity Kremkow et al. (2014). V1 complex cells integrate contrast polarities, producing partial invariance. From V1, information travels down the ventral stream through V2, V4 and inferotemporal cortex (IT), with increasing selectivity. Patches of face selective neurons emerge in IT Tsao et al. (2006). Face-neurons strongly prefer natural vs inverted contrast faces Ohayon et al. (2012); Freiwald and Tsao

* These authors contributed equally

(2010). While contrast polarity is differentially encoded in face neurons, it is not known how contrast information propagates across the ventral stream hierarchy.

Most studies used grayscale images, constraining contrast to ON and OFF pathways and neglecting important color information. Recently, visual cortical neurons have been characterized with color stimuli, showing prior ON/OFF neurons contain color information Li et al. (2022). Therefore, it is critical to study how full contrast is encoded across the ventral stream. ON and OFF contrast pathways are highly symmetric, but have subtle functional asymmetries from retina to early visual cortex, which shows a bias toward dark stimuli St-Amand and Baker (2023); Rahimi-Nasrabadi et al. (2021). The progression from contrast symmetry to eventual asymmetry has remained unexplored with chromatic stimuli.

Artificial contrast processing Artificial neural networks (ANNs), and in particular convolutional neural networks (CNNs) are state-of-the-art models of the visual system. Multiple works have demonstrated broad correspondences between CNN layers and primate ventral stream areas, with early layers mapping onto V1/V2 and later layers onto IT (Yamins et al., 2014; Güçlü and van Gerven, 2015). Filters in early CNN layers have Gabor-like components resembling V1 cells with ON and OFF components, and late layers have category-related information as in IT Zeiler and Fergus (2013). Hendrycks and Dietterich (2019) showed that CNNs do not generalize to contrast transformations. However, it is not known how symmetric ON/OFF and color contrast pathways are in CNNs and to what extent they persist along the network depth.

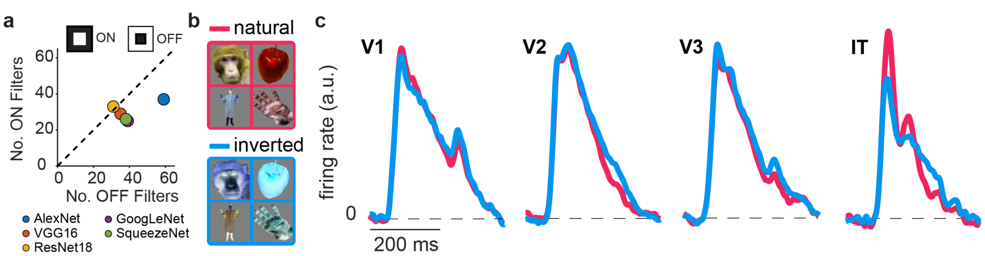

Figure 1: **Contrast polarity encoding. a**. First layer filters in CNNs have an OFF bias. **b**. Examples of positive, natural (top, red) and inverted contrast images (bottom, blue). **c**. Mean population responses per contrast polarity from Monkey T.

Here we investigate contrast processing along the hierarchies of the primate visual ventral stream and convolutional neural networks. We record along the ventral stream with a novel approach, using a single, chronically implanted, high-density electrode Neuropixels probe, and present images of natural and inverted contrasts. For both systems, we hypothesize that early areas which encode local features will be symmetric to contrast inversion, but more global high-level features will break the symmetry. We have two main contributions in this work: **(1) In the macaque visual stream, contrast polarity information increases along the ventral hierarchy from close to chance in V1 to peak decoding in IT. (2) In CNNs, contrast polarity information peaks in early layers, decays in intermediate layers, and re-emerges to intermediate levels in late layers.** Our results highlight a fundamental difference in how biological and artificial systems process fundamental visual features. This divergence provides an opportunity to refine computational models towards greater biological fidelity.

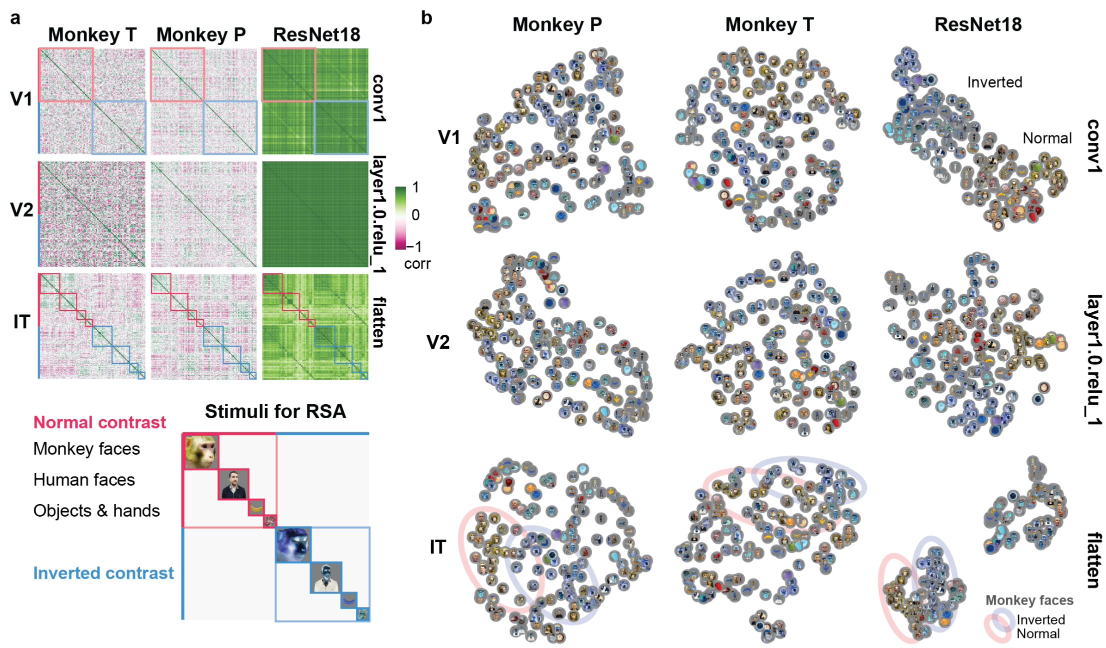

Figure 2: **Representational similarity along the visual hierarchy. a.** RSA (correlation) of ventral stream in two monkeys and ResNet18 layers, stimuli organized below. **b.** UMAP dimensionality reduction from data in **a** (zoom to see). ResNet18 conv1 organizes by color. Last layers cluster categories (monkeys encircled).

2. Results

Contrast asymmetry in CNN filters and visual cortex Contrast asymmetries have not been extensively characterized in CNNs. Analysis of pretrained networks revealed that filters in the first convolutional layer exhibit a pronounced *OFF bias*, responding more strongly to dark features than to light ones (Fig. 1a), consistent with observations in early visual cortex Kremkow et al. (2016). To assess overall symmetry of the encoding of contrast in colored natural images, we compared neuronal responses to natural images and their contrast-inverted counterparts (Fig. 1b). Using Neuropixels probes, we recorded from neurons across multiple cortical areas within a single penetration along the superior temporal sulcus. Population-averaged peristimulus time histograms (PSTHs) showed that early visual areas (V1/V2) responded with nearly identical strength to positive- and negative-contrast images, indicating relative symmetry at early stages (Fig. 1c). In contrast, differences between positive and negative contrast images were more visible in inferotemporal cortex (IT). Therefore, the mean population response of cortical neurons is symmetric to contrast inversion.

Contrast inversion reveals an asymmetry between primate visual cortex and CNNs To better understand the representation of contrast and its inversion along the depth of brains and CNNs, we performed representational similarity analysis, visualization and decoding. To quantify representational similarity, we used the pairwise correlation distance of the response vectors to the image set from biological neurons within a cortical area or artificial units within a network layer (Fig. 2). The representation of the inverted contrast and categories emerged gradually along the ventral stream, with no apparent structure in V1. In CNNs, such as ResNet18, a contrast polarity division started at the first layer (polarity blocks). To better visualize this discrepancy, we projected the image responses into 2D using UMAP (Fig. 2b). In ResNet18, contrast polarity segregated from the first layer. This segregation was lost at intermediate layers and emerged later in a category-dependent way. In the ventral stream, embeddings of natural and inverted images overlapped across contrasts and categories in V1 (low-level feature symmetry), and separated more clearly by category and contrast towards IT, reflecting stronger polarity encoding at higher levels. To further quantify contrast-polarity information, we used a linear support vector machine (SVM) trained for binary classification using 5-fold stratified cross-validation (chance = 50%). In macaque cortex, accuracy increased along the ventral stream from near chance in V1/V2 and peaking in IT. In CNNs, the trend reversed: accuracy was highest in the first layer, lowest in intermediate layers, and intermediate in deeper layers, yielding a U-shaped profile with depth (Fig. 3). Therefore, contrast polarity was encoded asymmetrically in biological and artificial vision already from the first processing stage, even though their categorical representations were similar for normal contrast images.

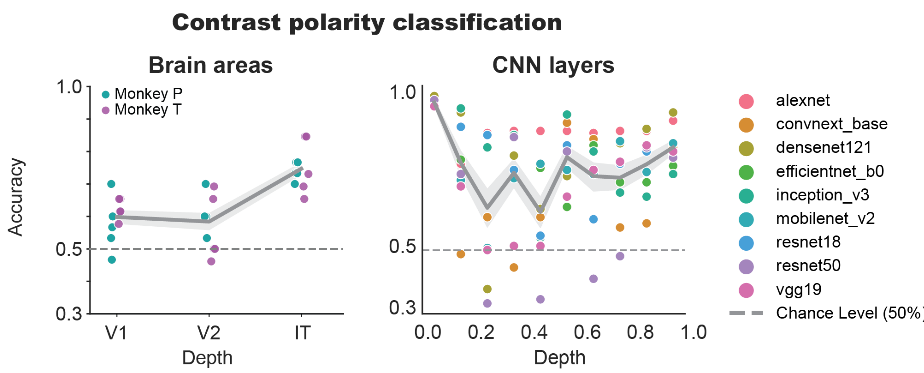

Figure 3: **Contrast polarity hierarchies diverge in brains and CNNs.** Binary SVM accuracy by 5-fold CV; in brains, each dot is a fold from two animals; in CNNs, each dot is the layer mean across folds. Layer depth is normalized to 0-1 range.

3. Discussion

Our novel approach used a single Neuropixels probe to record neurons along the macaque ventral visual stream. Using contrast inversion, we found that although CNNs show a strong hierarchical correspondence to ventral stream areas, important differences appear immediately: in the brain, symmetry breaking to contrast inversion increases from low- to high-level areas, whereas CNNs show the opposite pattern. We are currently training networks and collecting more electrophysiology data to dissect the mechanisms of this asymmetry. Because contrast inversion is out-of-distribution for both systems, these results raise important questions about network training and visual development. However, as early visual areas are more symmetric to contrast inversions, we hypothesize that raising a monkey under inverted contrast experience will modify high level areas while leaving low level areas unchanged. Thus, fundamental gaps remain between artificial and biological vision.

Acknowledgments

Funding: Alice and Joseph Brooks Fellowship (S.N., G.R.T.) and Hearst Fellowship (G.R.T.)

References

Winrich A. Freiwald and Doris Y. Tsao. Functional Compartmentalization and Viewpoint Generalization Within the Macaque Face-Processing System. *Science*, 330(6005):845–851, November 2010. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1194908. URL https://www.science.org/doi/10.1126/science.1194908.

Umut Güçlü and Marcel A. J. van Gerven. Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 35(27):10005–10014, July 2015. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.5023-14.2015.

Dan Hendrycks and Thomas Dietterich. Benchmarking Neural Network Robustness to Common Corruptions and Perturbations, March 2019. URL http://arxiv.org/abs/1903.12261. arXiv:1903.12261 [cs].

E. Kobatake and K. Tanaka. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, 71(3):856–867, March 1994. ISSN 0022-3077, 1522-1598. doi: 10.1152/jn.1994.71.3.856. URL https://www.physiology.org/doi/10.1152/jn.1994.71.3.856.

Jens Kremkow, Jianzhong Jin, Stanley J. Komban, Yushi Wang, Reza Lashgari, Xiaobing Li, Michael Jansen, Qasim Zaidi, and Jose-Manuel Alonso. Neuronal nonlinearity explains greater visual spatial resolution for darks than lights. *Proceedings of the National Academy of Sciences*, 111(8):3170–3175, February 2014. doi: 10.1073/pnas.1310442111. URL https://www.pnas.org/doi/abs/10.1073/pnas.1310442111. Publisher: Proceedings of the National Academy of Sciences.

Jens Kremkow, Jianzhong Jin, Yushi Wang, and Jose M. Alonso. Principles underlying sensory map topography in primary visual cortex. *Nature*, 533(7601):52–57, 2016. URL https://www.nature.com/articles/nature17936. Publisher: Nature Publishing Group UK London.

Peichao Li, Anupam K. Garg, Li A. Zhang, Mohammad S. Rashid, and Edward M. Callaway. Cone opponent functional domains in primary visual cortex combine signals for color appearance mechanisms. *Nature Communications*, 13(1):6344, October 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-34020-2. URL https://www.nature.com/articles/s41467-022-34020-2. Publisher: Nature Publishing Group.

Marissa Nederhouser, Xiaomin Yue, Michael C. Mangini, and Irving Biederman. The deleterious effect of contrast reversal on recognition is unique to faces, not objects. *Vision Research*, 47(16):2134–2142, July 2007. ISSN 0042-6989. doi: 10.1016/j.visres.2007.04.007. URL https://www.sciencedirect.com/science/article/pii/S0042698907001757.

Shay Ohayon, Winrich A. Freiwald, and Doris Y. Tsao. What makes a cell face selective? The importance of contrast. *Neuron*, 74(3):567–581, 2012. URL https://www.cell.com/neuron/fulltext/S0896-6273(12)00329-7?script=true&code=cell-site. Publisher: Elsevier.

Hamed Rahimi-Nasrabadi, Jianzhong Jin, Reece Mazade, Carmen Pons, Sohrab Najafian, and Jose-Manuel Alonso. Image luminance changes contrast sensitivity in visual cortex. *Cell reports*, 34(5), 2021. URL https://www.cell.com/cell-reports/fulltext/S2211-1247(21)00005-X. Publisher: Elsevier.

Peter H. Schiller. The ON and OFF channels of the visual system. *Trends in Neurosciences*, 15(3):86–92, January 1992. ISSN 0166-2236. doi: 10.1016/0166-2236(92)90017-3. URL https://www.sciencedirect.com/science/article/pii/0166223692900173.

David St-Amand and Curtis L. Baker. Model-Based Approach Shows ON Pathway Afferents Elicit a Transient Decrease of V1 Responses. *Journal of Neuroscience*, 43(11):1920–1932, March 2023. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.1220-22.2023. URL https://www.jneurosci.org/content/43/11/1920. Publisher: Society for Neuroscience Section: Research Articles.

Doris Y. Tsao, Winrich A. Freiwald, Roger B. H. Tootell, and Margaret S. Livingstone. A cortical region consisting entirely of face-selective cells. *Science (New York, N.Y.)*, 311 (5761):670–674, February 2006. ISSN 1095-9203. doi: 10.1126/science.1119983.

Daniel L. K. Yamins, Ha Hong, Charles F. Cadieu, Ethan A. Solomon, Darren Seibert, and James J. DiCarlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 111(23):8619–8624, June 2014. ISSN 1091-6490. doi: 10.1073/pnas.1403112111.

Matthew D. Zeiler and Rob Fergus. Visualizing and Understanding Convolutional Networks, November 2013. URL http://arxiv.org/abs/1311.2901. arXiv:1311.2901 [cs].

Appendix A. Appendix. Methods

Electrophysiological data collection We recorded multi-unit activity from two rhesus macaques (Monkeys P and T) using Neuropixels (NHP) probes inserted along the superior temporal sulcus. The penetrations traversed ventral-stream areas, sampling—in Monkey P: V1 (59 units), V2 (73), and posterior IT (pIT; 121); and in Monkey T: V1 (22), V2 (17), V3 (22), central IT (cIT; 56), and anterior IT (aIT; 23). "IT", in the paper, refers to the concatenated dataset across IT subregions. Animals passively fixated a central point while images were presented at the center of the recorded population's receptive field for 183 milliseconds. Each image was repeated multiple times, and recording sites were included in the analysis based on split-half reliability across repetitions. For all analyses, we computed mean firing rate from 90 to 300 milliseconds from image onset to account for response latency.

Stimulus set We used 75 natural images paired with their contrast-inverted versions (150 total) spanning multiple categories (monkey and human faces, bodies, hands, objects).

Contrast representation analysis To quantify the information about contrast polarity, we used linear support vector machines to classify natural contrast vs inverted contrast images. We performed 5-fold stratified cross validation over our balanced dataset of 150 images, 75 images in their natural and inverted contrasts. In addition to two monkey dataset from V1, V2 and IT, we used several pretrained CNNs from torchvision ResNet18, ResNet50, AlexNet, VGG19, ConvNeXT_Base, MobileNet_V2. For CNNs we sampled 10 layers, uniformly spaced between the first and last, to obtain a normalized depth from 0 to 1 to plot in the same axis.

Contrast polarity sensitivity in CNNs To investigate contrast polarity sensitivity in convolutional neural networks, we analyzed the first convolutional layer (conv1) of five widely used pretrained models: AlexNet, VGG16, ResNet18, GoogLeNet, and SqueezeNet. Filters were classified as ON- or OFF-dominated based on the sign of their summed spatial weights—positive sums indicated ON dominance, and negative sums indicated OFF dominance.