Evaluation of Neural Networks with Regions of the Brain

Infant Vision Project: Part 3

Computational Visual Perception WS 2024/25

Authors: Raj Hemant Panchal, Sohan Joseph Saldanha

Supervisor: Prof. Dr. Bernhard Egger

1. Introduction

This project explores infant vision and its connection to adult visual systems using computational approaches. In Part 3, we evaluate trained artificial neural networks using Representational Dissimilarity Matrices (RDMs) and compare them to different regions of interest (ROIs) in the brain responsible for vision, utilizing the Neural Encoding Dataset (NED).

2. Methodology Implemented

2.1. Layer-Wise Developmental Curriculum Learning

We implemented a *Layer-wise Developmental Curriculum Learning* approach, where the model's layers were gradually unfrozen in stages corresponding to developmental age milestones. At each stage, age-specific data loaders provided training data that reflected perceptual changes during early development. This progressive learning strategy allowed the model to acquire visual representations in a biological way. For the backbone architecture, we used ResNet18, incorporating transformations that simulated infant vision, such as reduced visual acuity and color adjustments. These modifications ensured that the training process better reflected early visual experiences.

2.2. Trained Network Variants

To evaluate the impact of developmental transformations, we trained four different networks:

1. **Acuity Network**: Simulates the lower visual acuity development observed in infants.
2. **Color Network**: Models the gradual development of color perception development in infants.
3. **Combined Network**: Integrates both acuity and color transformations to simulate a more comprehensive developmental process.
4. **No Curriculum Network**: Serves as a baseline model, trained without any developmental transformations.

3. Analyzing Feature Representations Using RDMs

Representational Dissimilarity Matrices (RDMs) measure differences in feature representations across models or network layers by computing pairwise correlation distances. In this study, we constructed RDMs for four trained networks to examine variations in feature representations across key layers ('conv1', 'layer 1', 'layer 3', 'layer 4'). Each RDM was compared to the No Curriculum baseline to evaluate the impact of developmental transformations.

To generate RDMs, we used 100 images from the TinyImageNet validation set, extracting feature maps from key layers. For clearer analysis, these images were transformed into four subsets: *No Transform, Blur, Color, and All Transformations*. This division provided valuable insights into how different transformations on images influenced feature representations across networks in each RDM.

4. Comparing Neural Networks with Brain Data

We evaluated our networks using the Neural Encoding Dataset (NED), which includes pre-generated fMRI responses from the Natural Scenes Dataset (NSD) across multiple brain regions of interest (ROIs) in different subjects. To assess the alignment between artificial and biological representations, we compared the representational dissimilarity matrices (RDMs) of our models with brain RDMs using *Spearman and Pearson correlation coefficients*.

5. Conclusion

The trend observed in the RDM of all networks shows increasing dissimilarity from the early to the later layers (Fig. 2.1, Fig. 2.2, Fig. 2.3). The intermediate layers of the combined network exhibit greater dissimilarity in the RDM compared to the other two networks. Additionally, applying blur and color transformations to the images resulted in lower dissimilarity in the RDM compared to both the "all transforms" and "no transforms" conditions.

Brain alignment across all networks was evaluated using the Spearman correlation coefficient. The results indicate that the color and no-curriculum networks exhibit higher correlation, whereas the curriculum network shows the least alignment. In contrast, using Pearson correlation, the no-curriculum network demonstrates the least alignment, while the acuity network exhibits the highest alignment with different brain regions.

6. Results

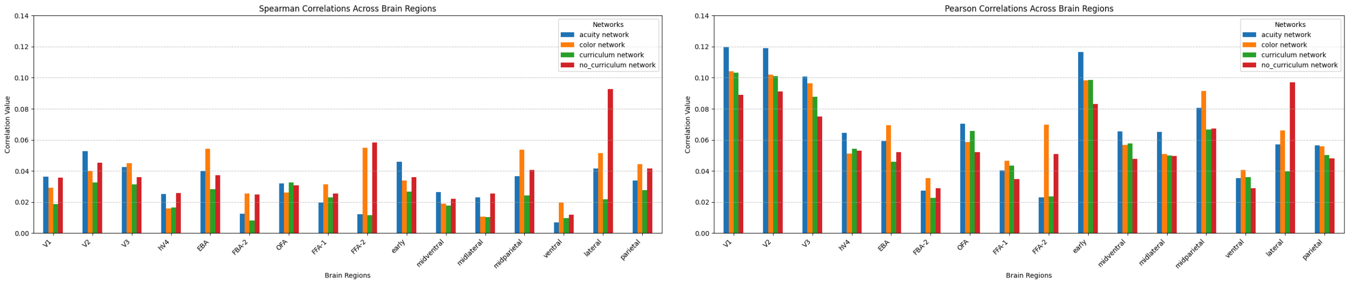

Fig 1: The graphs show the correlation between the four networks with different regions of interest (ROIs) in the brain. The graph on the *left* shows the correlations with the **Spearman correlation coefficient** and the *right graph* shows with **Pearson correlation coefficient**. The **four networks** include, *Acuity* (Blue), *Color* (Yellow), *Combined Curriculum* (Green), and *no curriculum* (Red). The **regions of interest (ROIs)** in the brain include *Early retinotopic visual regions* (V1, V2, V3, hV4), *Body-selective regions* (EBA, FBA-2), *Face-selective regions* (OFA, FFA-1, FFA-2), *Anatomical streams* (early, midventral, midlateral, mid parietal, ventral, lateral, parietal).

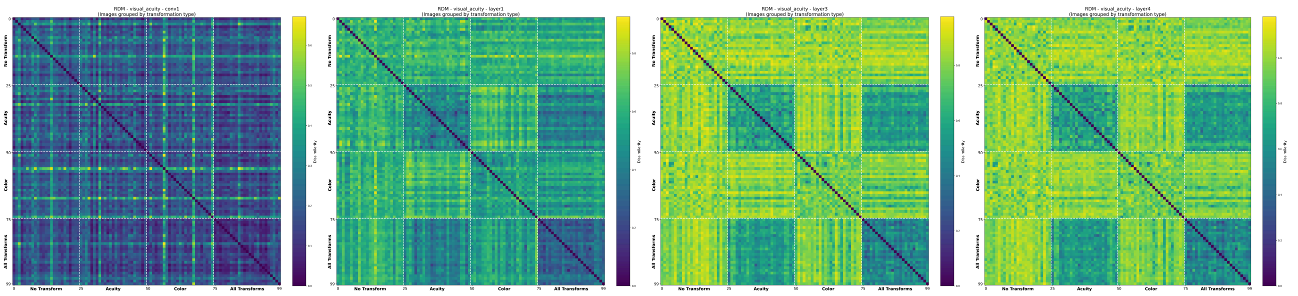

Fig 2.1: RDMs for acuity network, in different layers *conv1* (left-most), *layer 1* (left-center), *layer 3* (right-center), *layer 4* (right-most).

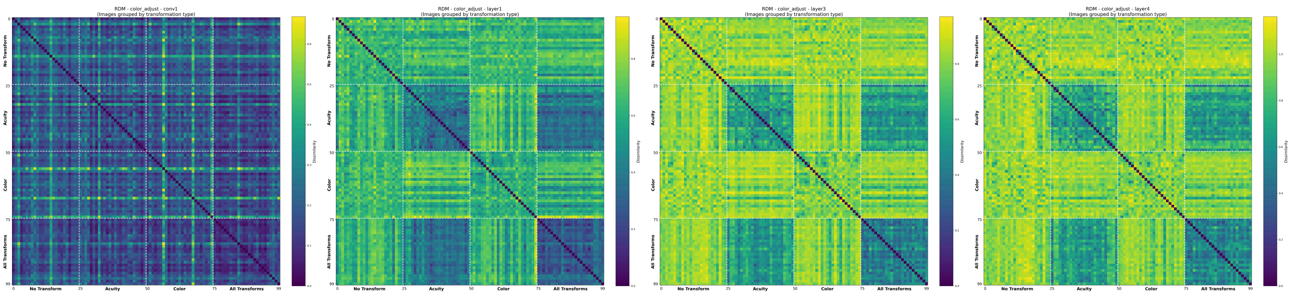

Fig 2.2: RDMs for Color network, in different layers *conv1* (left-most), *layer 1* (left-center), *layer 3* (right-center), *layer 4* (right-most).

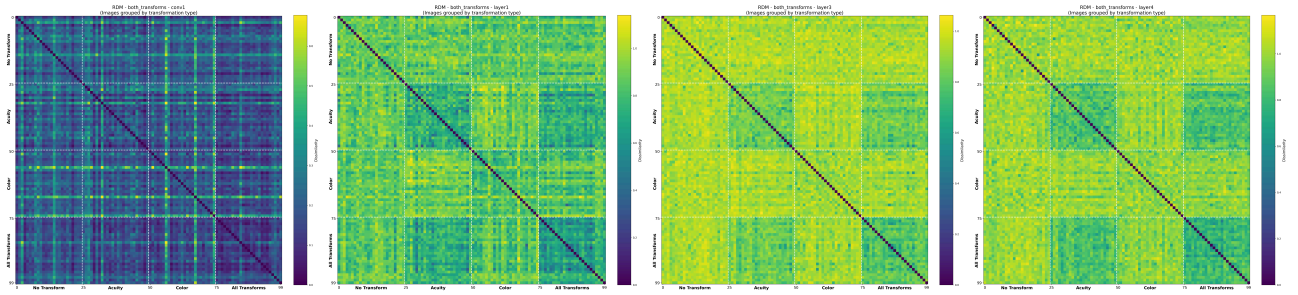

Fig 2.3: RDMs for combined network, in different layers *conv1* (left-most), *layer 1* (left-center), *layer 3* (right-center), *layer 4* (right-most).

7. Reference

1. Cichy, R., Khosla, A., Pantazis, D. et al. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence
2. NED Toolbox, NED manual and TinyImaget dataset
3. Shreya Kapoor: Computation Visual Perception - Comparing CNNs to the brain