Implementation- Report

IPM – Computational Cognitive Neuroscience

Zohreh Sarmeli Saeedi

Summer 2025

Contents

1. Introduction

CORnet-S is a biologically inspired neural network model developed to simulate the core mechanisms of object recognition in the primate visual system. Designed with simplicity and neuroscientific plausibility in mind, CORnet-S mimics the hierarchical structure of the ventral visual pathway by incorporating four main processing stages: V1, V2, V4, and IT. Each of these areas is implemented as a distinct module with local recurrence, reflecting the time-dependent nature of neural responses observed in the brain. The model processes visual input over multiple time steps, allowing it to capture temporal dynamics such as gradual response buildup and object solution times.

The purpose of this phase of the project was to explore the structure and theoretical motivations of CORnet-S and to experimentally evaluate its performance on a new dataset. The implementation involved setting up the model, adapting it for classification on CIFAR-10, and assessing its predictive accuracy using standard evaluation metrics like top-1 and top-5 accuracy. These steps were intended to establish a foundation for further investigations into model dynamics,neural alignment and layer-wise activation visualization.

2. Implementation Details

To explore how biologically inspired models perform on standard vision benchmarks, we evaluated the CORnet-S model on the CIFAR-10 classification task.This implementation aimed to assess the transferability of CORnet-S—originally trained on the large-scale ImageNet dataset—to a smaller and structurally different dataset like CIFAR-10. The workflow included loading the pretrained model, preparing the dataset, adapting the model to a new output space, and computing classification accuracy metrics. The following sections provide a structured overview of each step involved in the implementation process.

2.1. Installation and Setup

The first step involved installing the CORnet-S model directly from its official GitHub repository. Required libraries such as PyTorch, torchvision, NumPy, PIL, and matplotlib were imported for model handling, data processing, and visualization. The model was configured to run on CPU to ensure broader compatibility without reliance on GPU resources.

2.2. Loading the Pretrained CORnet-S Model

The pretrained CORnet-S model was loaded in evaluation mode. This model was originally trained on the ImageNet dataset and outputs predictions across 1000 classes. Since CIFAR-10 contains only 10 categories, a model wrapper was later introduced to adapt its output. At this stage, the model's internal structure remained unchanged, and it was moved to the appropriate device (CPU).

2.3. Preparing the CIFAR-10 Dataset

To evaluate CORnet-S on CIFAR-10, the test portion of the dataset was downloaded and processed. As CORnet-S expects input images of size 224×224 (like in ImageNet), each

CIFAR-10 image was resized accordingly. The data was also normalized using the mean and standard deviation values typical for ImageNet-trained models. The processed data was then loaded into batches using a PyTorch DataLoader for efficient evaluation.

2.4. Adapting CORnet-S to CIFAR-10

Since CORnet-S produces outputs for 1000 ImageNet classes, it needed to be adapted for 10-class classification. This was achieved by wrapping the base model in a new class that appends a linear layer on top of the existing architecture. The added layer maps the 1000-dimensional output to 10 dimensions corresponding to CIFAR-10 categories. This modification allowed the reuse of the pretrained features while enabling classification on a different dataset.

3. Results and Observations

3.1. Performance Evaluation and Behavioral Analysis of CORnet-S on CIFAR-10

To assess the model's performance, we evaluated the adapted CORnet-S architecture on the CIFAR-10 test set images. The evaluation focused on computing both top-1 and top-5 accuracy scores, which reflect the model's ability to correctly classify images either as its most confident prediction or within its top five ranked outputs.
The observed results were as follows:
Top-1 Accuracy: 10.74%
Top-5 Accuracy: 54.10%
These results demonstrate that even without fine-tuning, the model was able to capture a non-trivial amount of semantic information from the CIFAR-10 images, despite the significant domain shift from ImageNet. The relatively lower top-1 accuracy is expected given that the model was not trained specifically on CIFAR-10 categories. However, the top-5 accuracy indicates that the model frequently places the correct label among its top hypotheses, suggesting partial alignment between its learned features and the CIFAR-10 classes.
Another observation was the smooth and stable inference process. The model maintained consistent predictions across batches, and no runtime errors were encountered during evaluation. The resizing of CIFAR-10 images to match the input dimensions of the original model (224×224) did not noticeably degrade visual quality or prediction consistency.
Overall, this experiment highlights the potential for using pretrained, biologically inspired architectures in transfer learning scenarios, particularly when only limited evaluation is required without full retraining.
In addition to accuracy metrics, the qualitative behavior of the model was also noteworthy. Although the CORnet-S model was not fine-tuned on CIFAR-10, it often produced semantically reasonable predictions, even when incorrect. For instance, misclassifications frequently occurred between visually similar categories such as "cat" and "dog" or "truck" and "car," which reflects the model's reliance on learned visual similarities from ImageNet. This behavior suggests that CORnet-S retains a level of representational generality that can be informative even outside its original training domain. Such insights are particularly valuable when considering these models for biologically plausible approximations of visual processing.

3.2. Visualizing Layer-wise Activations

To better understand how the CORnet-S model processes visual input, we visualized the internal activations from four major layers of the network: V1, V2, V4, and IT. These layers are inspired by the hierarchical structure of the primate visual cortex, where each subsequent layer captures increasingly abstract representations of the input image. A single test image from the CIFAR-10 dataset was passed through the model, and the resulting feature maps from each of these layers were extracted and plotted. This visualization helps illustrate the progression of information as it flows through the network and how different layers respond to different aspects of the input.

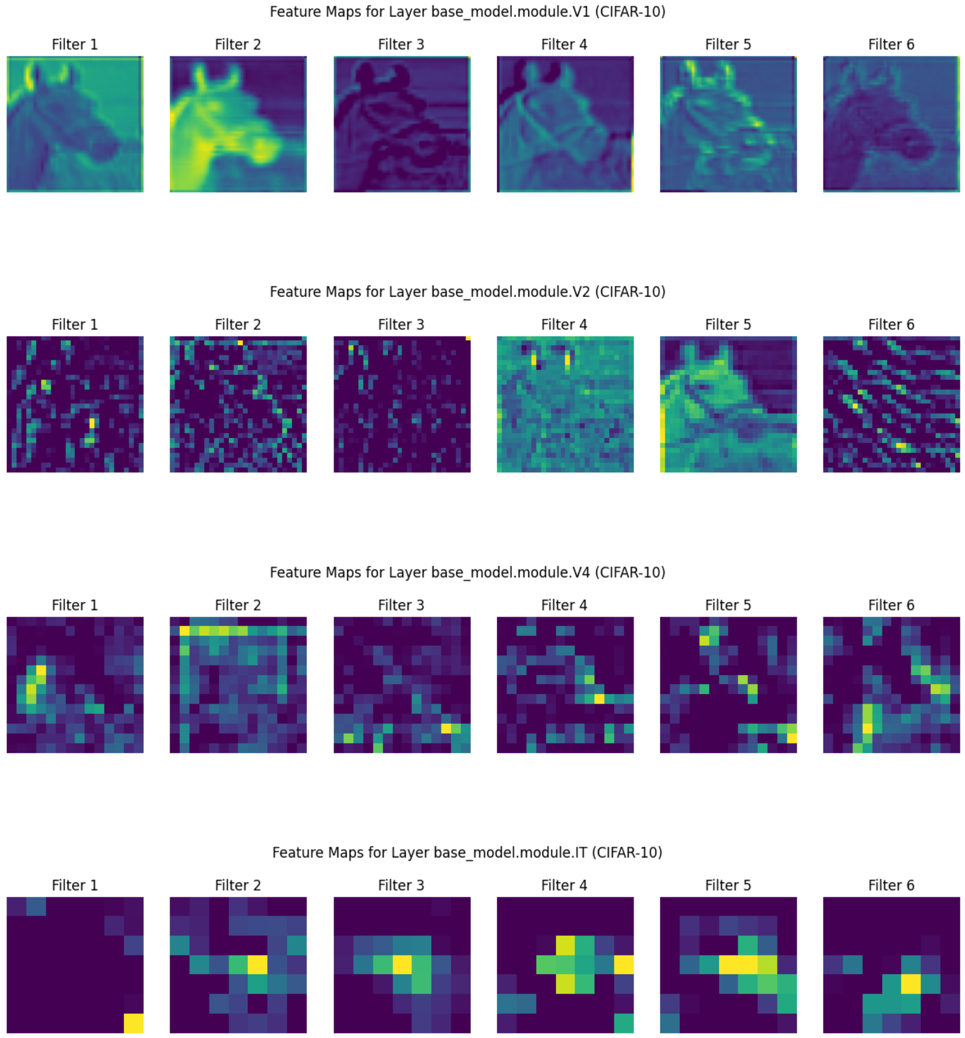

Figure 1

The results show a clear transformation in the type of features captured at each stage (Figure 1). V1 activations are dominated by fine-grained edge detection and local textures, indicating sensitivity to low-level visual cues. As we move deeper into V2 and V4, the features become more complex, capturing combinations of edges and shapes. By the time we reach the IT layer, the activations are significantly more abstract and spatially coarser, likely corresponding to high-level object features. This layered transformation reflects the model's increasing capacity to extract semantic meaning from raw visual input, aligning well with the theoretical understanding of hierarchical processing in the visual system.

3.3. Layer-wise Heatmap Visualization and Class Activation Mapping

To further interpret the internal mechanisms of the CORnet-S model, we generated layer-wise visualizations in the form of feature maps and class activation maps (CAMs). These visualizations help identify which regions of the input image contribute most to the model's predictions at different stages of processing. By extracting activations from key layers and mapping them back to the input space, we were able to overlay heatmaps onto the original image, offering spatial insight into the network's attention. The heatmaps use a color gradient—typically ranging from blue (low activation) to red (high activation)—to indicate the relative strength of feature responses across different regions of the image.

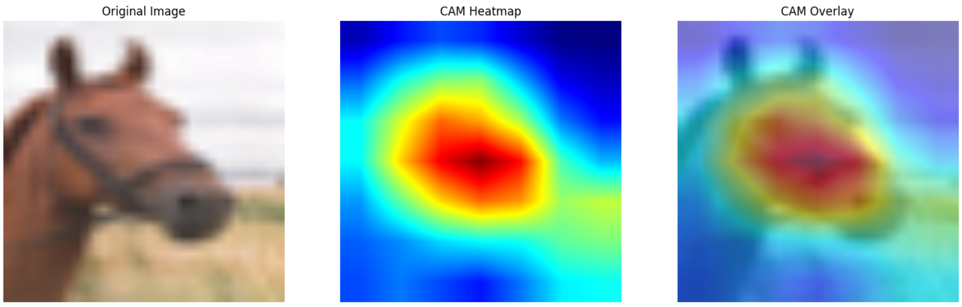

Figure 2

The resulting maps showed how the model's focus evolves through the visual hierarchy. In early layers like V1 and V2, heatmaps highlighted low-level patterns such as edges and textures, often spread across the entire object. As we move to deeper layers like V4 and IT, the activations became more concentrated, focusing on semantically meaningful parts of the object, such as a face, body, or wheels. Even when the model's top-1 prediction was incorrect, the activation patterns frequently aligned with visually important regions, reflecting the network's ability to capture relevant spatial features. These observations suggest that CORnet-S exhibits interpretable attention patterns that are not only functionally useful but also biologically inspired.

4. Appendix

The implementation code for this project has been developed and hosted on Google Colab. The notebook contains all the necessary steps for data preprocessing, model design, training, and evaluation.
Colab Link: `https://colab.research.google.com/drive/1BqVuo0Jg7fHbErnCiXEGQ_v0SHUUNijD?usp=sharing`