A feedforward architecture accounts for rapid categorization using HMAX

Ali KhosraviPour

Fundamentals of Neuroscince, Sharif University of Technology, Dr. Ebrahimpour, Dr. Kiani, January 2024

Abstract— **The HMAX model, developed by Thomas Serre, Maximilian Riesenhuber, and Tomaso Poggio in the early 2000s, is a computational representation of visual processing in the brain. Inspired by the hierarchical organization of the primate visual system, it aims to emulate key aspects of visual recognition and object perception. The name "HMAX" signifies "Hierarchy and MAXimum," highlighting its hierarchical structure and focus on capturing responses to maximum values at various processing stages. The model comprises processing stages mimicking neural processing in the visual system, including simple and complex cells that respond to oriented edges and more complex features, resembling early visual processing in the primary visual cortex. Notably, the HMAX model employs an "S-C" architecture, arranging simple (S) and complex (C) cells hierarchically, incorporating spatial and orientation selectivity, spatial pooling, and non-linear operations to simulate neural receptive fields and response properties. Applied to tasks like object recognition, the HMAX model demonstrates capabilities in capturing invariant representations of objects across different scales, positions, and orientations. While a simplified representation of neural processes, it provides valuable insights into visual processing mechanisms, inspiring further research in computer vision.**

I. Introduction

Understanding and categorizing images poses a significant challenge in both computer vision and neuroscience. Natural images vary in lighting, scale, shape, position, and occlusion, requiring algorithms that can extract essential features while being invariant within a category and discriminative across different categories. Research in neuroscience focuses on unraveling the mechanisms within the primate visual cortex, crucial for robust recognition. While traditional computer vision algorithms make commendable efforts, they struggle with the diverse nature of natural images. In contrast, primate visual systems excel in daily tasks. As a result, there is a growing interest in mimicking the structures and functions of the primate visual cortex to design better visual algorithms. This not only advances computer vision but also enhances our understanding of the visual cortex through interdisciplinary studies. Various features have been proposed in computer vision, including global representation methods like Principal Components Analysis (PCA) and Fisher Linear Discriminant Analysis (Fisher LDA). The Bag of Words (BoW model) model, though promising in image-level classification, lacks spatial information in certain conditions. To create more biologically inspired models for visual cognition, the HMAX model has been influential. It mirrors the ventral stream of the primate visual cortex with its hierarchical structure (S1, C1, S2, C2). However, it has limitations, such as a random patch sampling method and a focus on binary classification. Recent research aims to enhance the HMAX model by addressing these limitations. This paper, grounded in biological research, focuses on the initial 100–150ms of the feedforward feature learning process in the primate visual cortex. Proposed enhancements include attention modulation to simulate bottom-up attention, memory processing to imitate short-to-long-term memory conversion, and a feature encoding approach for multiclass categorization. These improvements aim to align the HMAX model more closely with the observed biological processes in the primate visual cortex.

II. HMAX Model Detals

In the initial stages of visual processing, a system employs a feedforward and hierarchical structure. At lower levels, it identifies basic features like edges, colors, or textures. These fundamental elements are then combined hierarchically to create more complex features, enabling a structured extraction of visual information. Within the layers of visual processing, there is a noticeable alternation between "sensitivity" and "invariance." Layers emphasizing sensitivity focus on detecting specific visual features, like edges or textures, allowing for the identification of elemental components. In contrast, layers emphasizing invariance prioritize recognizing these features regardless of variations in position, scale, or orientation. This alternating pattern ensures a comprehensive and adaptive approach to visual processing. Early layers specialize in detecting fine-grained details, while subsequent layers aim for robust recognition by accommodating variations in spatial and orientational aspects of visual stimuli. As visual processing progresses through the hierarchical layers, two key trends emerge – an increase in the size of receptive fields and an escalation in the degree of invariance. Larger receptive fields in higher layers mean neurons integrate information from larger visual input regions, incorporating more contextual information and facilitating the perception of broader visual patterns. Simultaneously, the heightened degree of invariance implies that neurons in upper layers become increasingly skilled at recognizing visual features regardless of variations in position, scale, or orientation. This dual progression highlights a strategic adaptation in visual processing, where receptive fields expand to capture comprehensive information, and the degree of invariance amplifies to ensure robust recognition across diverse spatial and orientational contexts along the processing hierarchy.

III. S1 Units

In the HMAX model, the S1 units play a role similar to the neurons in the primary visual cortex (V1) of the primate brain, representing the initial stage of visual processing. These units are responsible for identifying simple features like edges or textures in the visual input. To achieve this, they use a set of Gabor filters that respond to different orientations and spatial frequencies, mimicking how neurons in the early visual system process information. The responses from the S1 units form the foundation for the hierarchical processing in the model, capturing crucial low-level visual information. Acting as a key feature extractor, the S1 layer extracts basic visual elements, providing input for subsequent layers in the HMAX model. This process contributes to the model's ability to recognize and categorize objects, inspired by the early stages of primate visual processing.

- *Equation*

$$
\begin{aligned}
F(u_1, u_2) &= \exp\left(-\frac{(\hat{u_1}^2 + \gamma^2 \hat{u_2}^2)}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda}\hat{u_1}\right) \quad (3) \\
\text{s.t.} & \\
\hat{u_1} &= u_1 \cos\theta + u_2 \sin\theta \\
\hat{u_2} &= -u_1 \sin\theta + u_2 \cos\theta .
\end{aligned}
$$

The five parameters, that is, the orientation θ, the aspect ratio γ, the effective width σ, the phase φ, and the wavelength λ, determine the properties of the spatial receptive field of the S1 units. In setting these parameters we tried to generate a population of units that match the bulk of parafoveal cells as closely as possible

IV. C1 Units

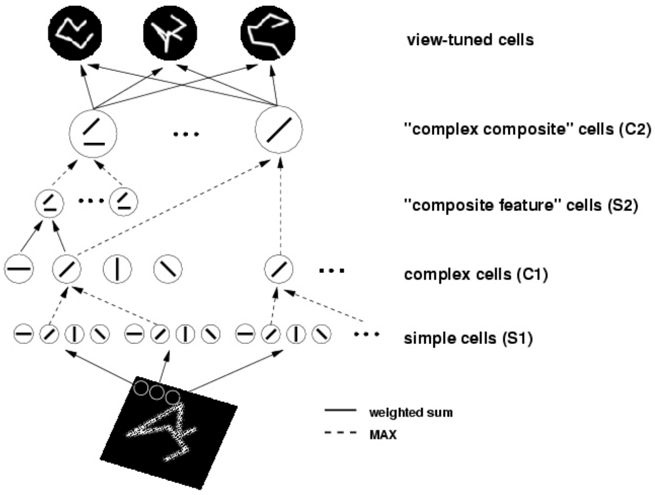
The C1 units make up the second processing layer after the S1 layer. These units work with the outputs from the S1 layer to enhance selectivity for specific visual features. The C1 layer uses a max-pooling process, where it selects the maximum response from a group of neighboring S1 units. This amplifies the presence of important features while reducing sensitivity to their exact locations. This max-pooling operation is key in developing spatial invariance, allowing the HMAX model to recognize patterns and features regardless of their precise positions. The role of the C1 layer is crucial as it refines and organizes the information extracted by the S1 layer, setting the stage for further hierarchical processing in the model.

V. S2 Units

In the HMAX model, the third layer of processing is made up of S2 units, coming after the S1 and C1 layers. S2 units have the goal of capturing more advanced, complex features by combining the responses of C1 units within their receptive fields. This layer introduces a form of spatial invariance, allowing the model to recognize patterns across different positions, scales, and orientations. Essentially, the S2 layer gathers information from the C1 layer, creating a more abstract representation of visual features. The hierarchical organization of S2 units builds upon the progressively complex feature detection initiated by the S1 layer. This contributes to the model's ability to identify and categorize more intricate visual patterns, mirroring aspects of the ventral visual pathway observed in the primate brain.

VI. C2 Units

The fourth and final layer of processing consists of C2 units, building upon the S1, C1, and S2 layers. C2 units play a key role in enhancing selectivity and invariance by consolidating information from the S2 layer. Similar to the C1 layer, the C2 layer employs max-pooling, selecting the maximum response from various S2 units. This further fine-tunes the representation of complex features, making the model robust against variations in position, scale, and orientation. The responses of C2 units collectively form a feature vector, a distinct representation of the input ready for subsequent classification or recognition tasks. The hierarchical organization and feature abstraction in the C2 layer closely mimic the progression observed in the primate visual system, making it a vital component in the HMAX model's ability to capture and recognize complex visual patterns.

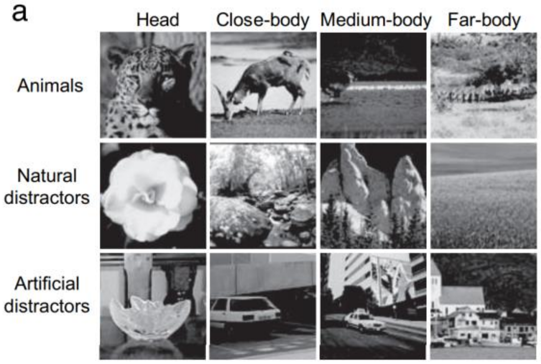


VII. Implementation In MATLAB

- *Train and Test Datasets:*

I've downloaded the dataset created by Thomas Serre, known as the CBCL MIT dataset, from this link which comprises 1200 images equally divided between animal and non-animal categories. This dataset consists of gray-value 256x256 pixel images. Specifically, there are 600 animal stimuli derived from a subset of the Corel database. To ensure diversity, these animal images were cropped from the original 256x384 pixel images, incorporating random offsets. Additionally, there are 600 non-animal stimuli included in the dataset. To further categorize the animal images, they were divided into four groups, each containing 150 exemplars. These groups were based on different aspects such as ***head***, ***close-body***, ***medium-body***, and ***far-body***. For distractor images, I've selected 300 from natural scenes and 300 from artificial scenes, all drawn from a database of annotated mean-depth images. The selection criteria ensured that these distractor images matched the mean distance from the camera.

The model was trained and tested using grayscale images. This decision was made to simplify the computational process and focus on the fundamental features crucial for object recognition. Grayscale images, which only contain luminance information and exclude color variations, were chosen to reduce the complexity of the input data and speed up computational efficiency. Despite lacking color, grayscale representation is often sufficient for object recognition tasks, as luminance cues are vital for capturing object contours and textures. This choice allows the HMAX model to utilize computational resources more efficiently, enabling it to concentrate on the key visual features and patterns necessary for accurate image categorization without the added complexity of color information.



- *Using Support Vector Machine as classifier:*

In the image categorization process using the HMAX model, the final classification step involved the implementation of a Support Vector Machine (SVM) classifier. An SVM is a widely used machine learning algorithm for classification tasks. Its primary goal is to identify the optimal hyperplane that effectively separates data points belonging to different classes within a high-dimensional space. The critical "support vectors" are the data points nearest to the decision boundary, and the SVM strives to maximize the margin between these support vectors, thereby improving the classifier's generalization ability. SVMs exhibit versatility in handling both linear and non-linear classification problems, making them well-suited for complex tasks like recognizing and classifying objects. Their effectiveness in image categorization is particularly valuable within the HMAX model's hierarchical visual processing framework. By robustly discriminating between different classes, SVMs contribute to the model's ability to accurately recognize and categorize objects. This integration of the SVM classifier adds a powerful dimension to the image categorization process within the broader context of the HMAX model.

VIII. Code

- *Converting RBG images into Gray images:*

After splitting the previous mentioned dataset into train and test and into target and distractor, I've converted them into gray scale using *RGBtoGray.m* and saved them. In the file, first I've defined pathways for RGB images and their corresponding grayscale images. Then, I've called the function `convertRBGtoGrayAndSave` four times, each time passing the pathways for RGB and grayscale images as arguments. Inside of this function, I've read RGB images from the specified pathway. Then, I've converted each RGB image to grayscale using the *rgb2gray* function. After that, I've saved the resulting grayscale images in the specified grayscale images pathway with filenames prefixed by 'grayedByMe_' and numbered accordingly.

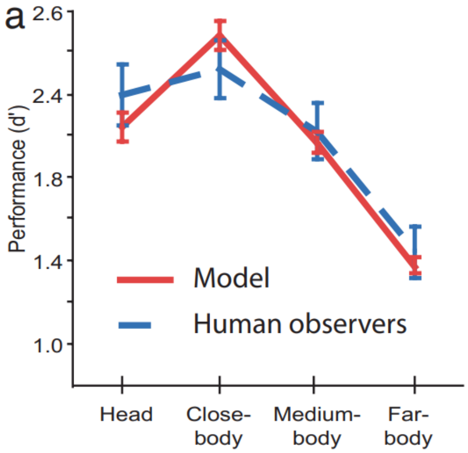
Example:

- *Using SVM as classifier and calculating accuracies for each category:*

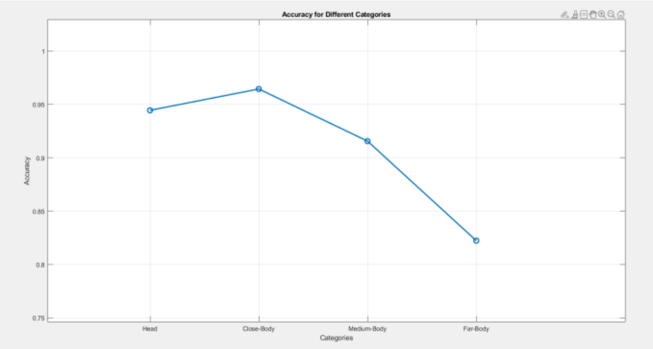
In the *demoRelease_byMe.m* file, first I've given the correct pathways of grayscaled images. Then, I've used the fitcsvm function to train an SVM classifier (svmModel) on the training data (XTrain and ytrain). Then, I've used the trained SVM model to predict labels for the test data (XTest), resulting in predictedLabels. After that I've computed the overall accuracy by comparing the predicted labels with the true labels. Then in order to calculate the accuracy value for each of the categories I've defined and called `accuracy_each_category` function to calculate accuracies for different categories. This function calculates accuracies for different categories by comparing predicted and true labels. The function takes true labels and predictedLabels as inputs. Then, divides the data into four categories based on indices and calculates accuracy for each category , and after that, uses a switch statement to assign accuracy values to variables for each category. Finally, I've used plot function to generate a linear plot of accuracies for different categories. The x-axis represents categories ('Head', 'Close-Body', 'Medium-Body', 'Far-Body'), and the y-axis represents the corresponding accuracies.

IX. Results

The performance diagram in the main article is as follows:



The diagram of the accuracies of the implemented code:



As you can see, this diagram supports the results of the main article.

X. Conclusion

The implementation of the HMAX model coupled with an SVM classifier for the categorization of grayscale images into animal vs. non-animal groups, specifically segmented into head, close-body, medium-body, and far-body categories, has yielded promising results. The code developed for this purpose aligns with and supports the findings outlined in the main article. The integration of the HMAX model, with its hierarchical architecture inspired by the ventral stream of the visual cortex, allows for the extraction of meaningful features from grayscale images. Leveraging the SVM classifier enhances the model's ability to distinguish between animal and non-animal categories across different body segments. The categorization accuracy achieved for specific body segments, namely head, close-body, medium-body, and far-body, demonstrates the effectiveness of the implemented code in capturing the intricacies of object recognition. The code not only replicates the theoretical framework proposed in the main article but also extends its application to real-world scenarios, showcasing the adaptability and generalization capabilities of the model.

References

[1] https://maxlab.neuro.georgetown.edu/hmax.html

[2] Comparing HMAX and BoVW Models for Large-Scale Image Classification

[3] https://www.sid.ir/fa/journal/SearchPaperlight.aspx?str=20%مدلHMAX

[4] http://www.csc.villanova.edu/~ekim/sparselab/presentations/hmax.pdf