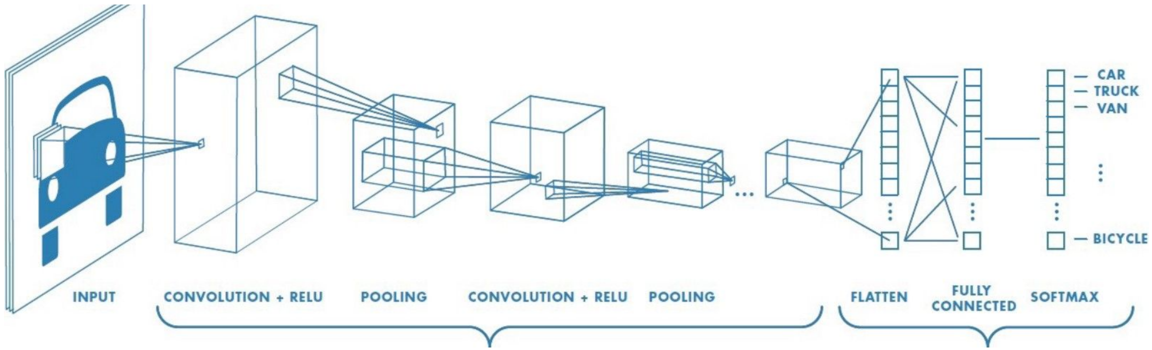
Object Recognition and Neural Network Connections

By: Nicholas Van Nostrand

COSC 521 - Computational Cognitive Neuroscience

Motivation

- Object recognition has been a growing area of interest in the area of artificial intelligence
- Convolutional neural networks and deep neural networks have been used to achieve the current best performing models



Problem Statement

- Object recognition in artificial intelligence still doesn't match human performance
- Current models can only handle basic inputs with little real world variations (movement, illumination, etc.)
- Solution: make artificial neural networks like the human brain to match human performance

Outline

- Issues with object variation
- Differences between an artificial neural network and the brain
- Overview of perception in the human brain
- Methods of human brain implementation with results
- Summary

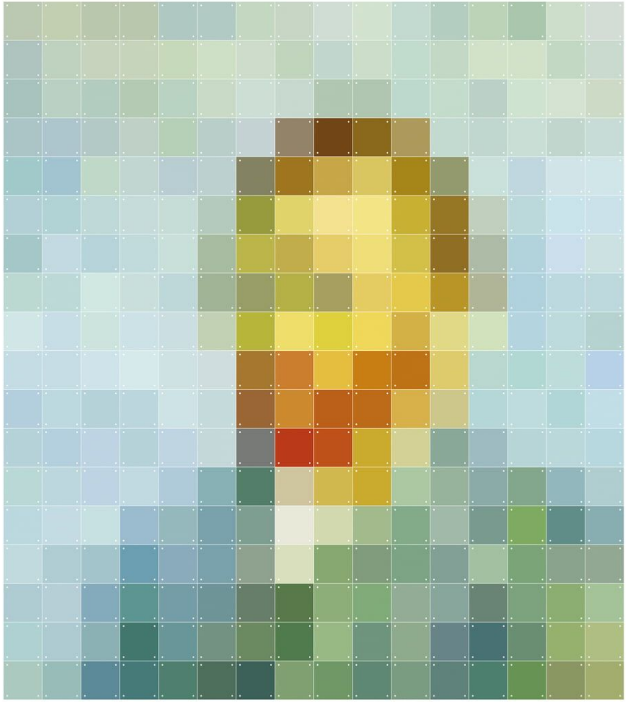
Object Variation

- Types of object variation: position, size, pose, lighting, occlusion, color
- Example: Different breeds of cats and their poses

Object Variation

Artificial Perception

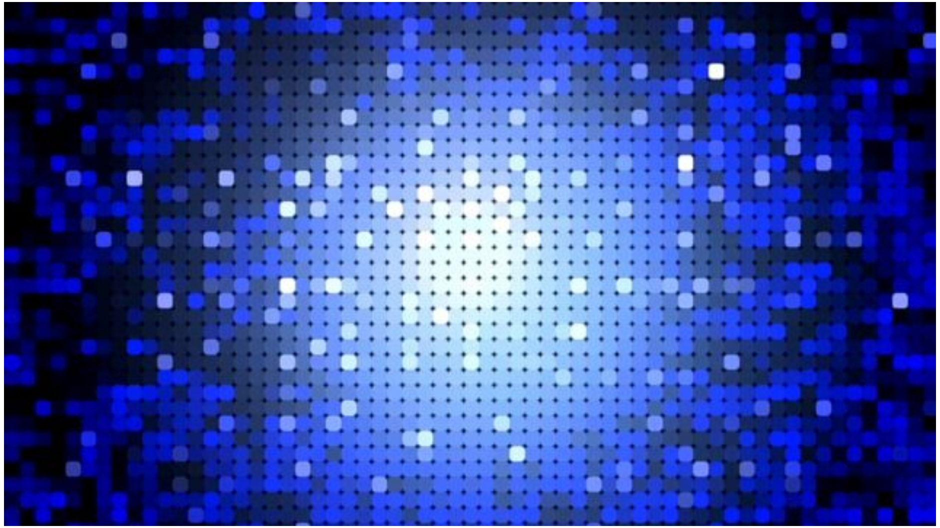
- Computers see objects in pictures as pixels, which isn't what humans do
- By default, computers can't distinguish specific traits of an object
- There are some methods for extracting features (ex: edge detection), but computers still lack human accuracy



Object Variation

Background

- Computers have issues working with image with backgrounds
- Pixels of an object can be interlaced with background pixels, lowering accuracy



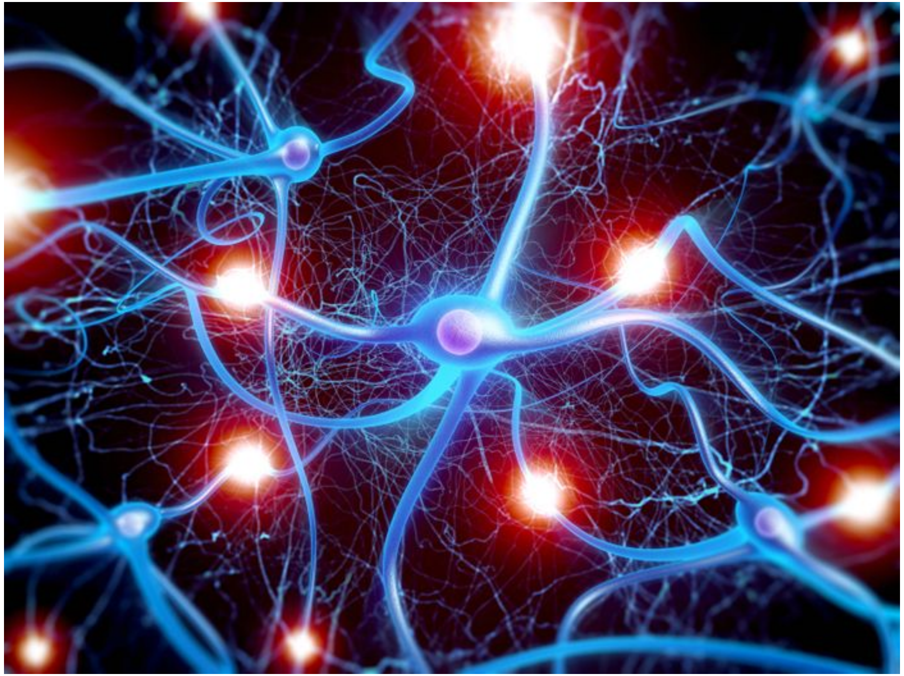
Object Variation
Movement

- Real-time movement also messes with pixels
- Current artificial neural networks usually convert videos to stationary images (with background subtraction)
- Movement blurs the image and interlaces pixels, leading to false identification

Differences

Network Sizes

- Artificial neural networks have fixed number of perceptrons (artificial neurons)
- Commonly up to 1,000 perceptrons
- Human brain has between 86 billion and 100 trillion neurons
- Number of human brain neurons can change
- So computers don't have as many to work with
- Limits how much a computer can process at any given time



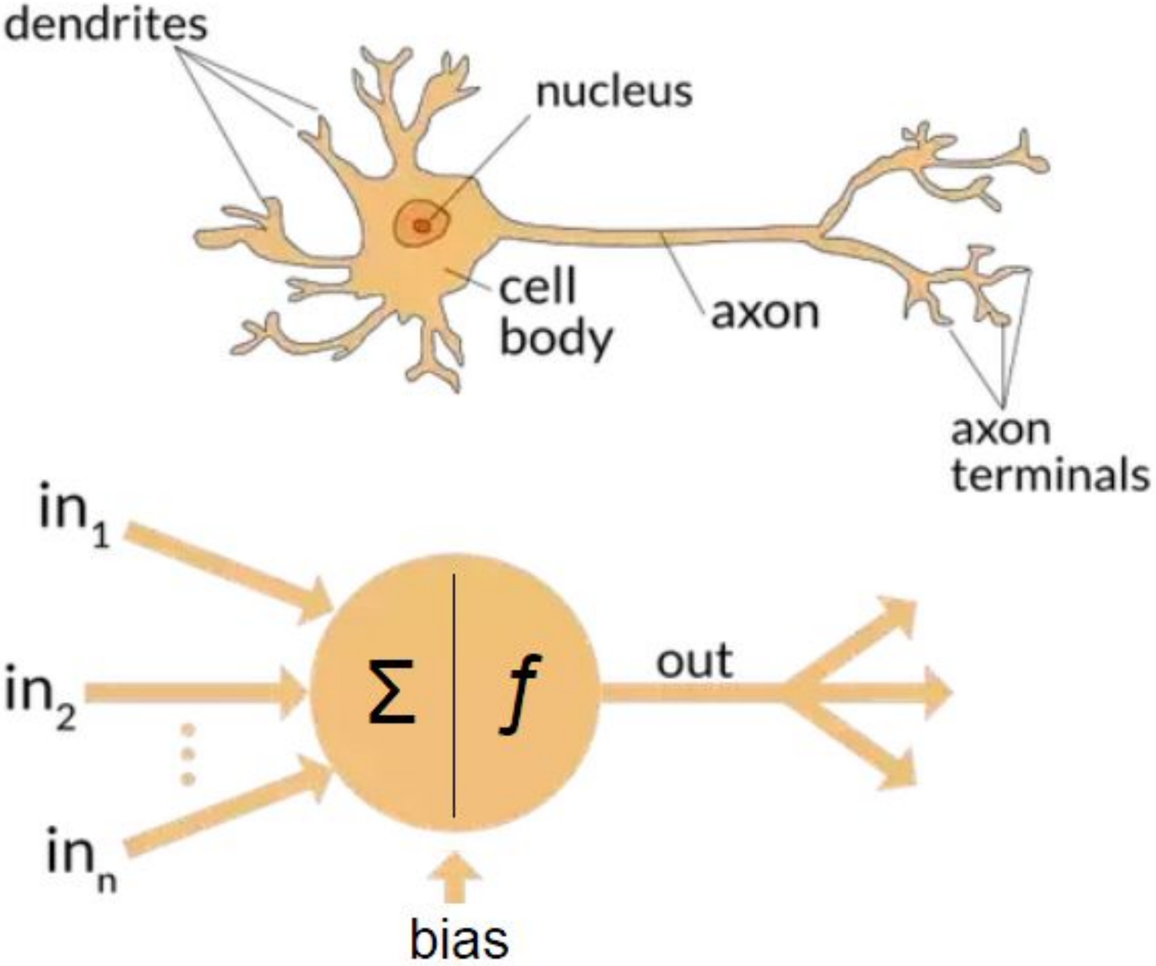
Differences

Speed

- Brain neurons can travel between 0.61 and 119 ms, though this can vary based on physical/mental state
- Perceptrons fixed at set floating point value for their synaptic weights
- Perceptrons only affected by GPU processing speed
- The human brain is faster and, coupled with more neurons, allows it to do more in less time

Differences
Topology

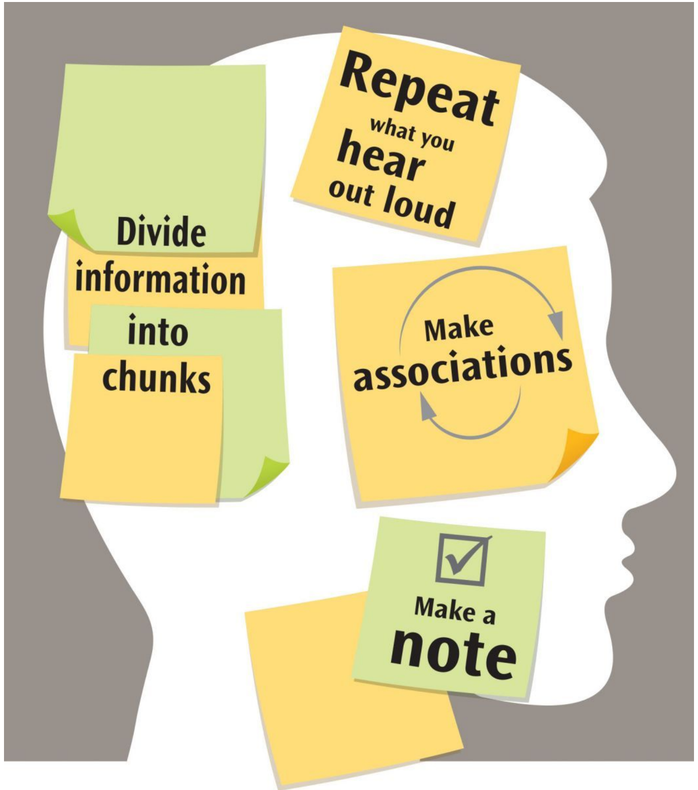
- For perceptrons, all layers are done one by one
- This is done because of the feedforward and feedback approaches for adjusting weights
- Perceptron layers don’t connect with layers that aren’t neighbors
- Humans can process information in different sections of the brain simultaneously
- Brain can fire neuron signals asynchronously
- So, artificial neural networks can’t multitask


dendrites
nucleus
cell
body
axon
axon
terminals
in_1
in_2
in_n
Σ
f
out
bias

Differences

Memory

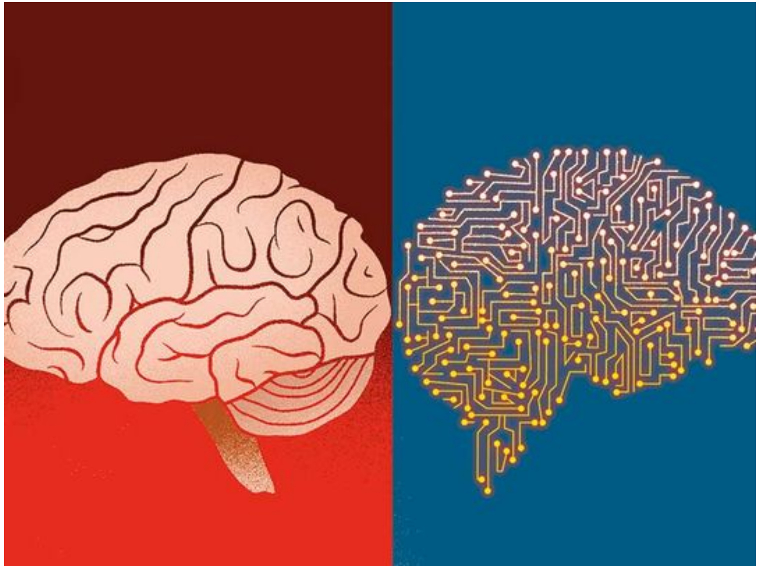
- Neurons have fault tolerance to store redundant information to prevent memory loss
- Neuron tolerance depends on where they are in the brain and require no central part for all memory storage
- Perceptrons have no built in fault tolerance or self regeneration
- There are options to recover deleted memory in artificial neural networks, but it's not already included



Differences

Memory - Continued

- Artificial neural networks also have the issue of replacing parts
- Removing “memory” (CPU) in a computer can mess with neural network knowledge
- Humans don’t replace their brains or intentionally wipe their memories



Differences

Power

- Brain activity takes up 20 percent of all human energy, which is about 20 watts of electricity
- This makes the brain very efficient in terms of power
- Artificial neuron networks, using a typical Nvidia GeForce Titan X GPU, requires 250 watts
- Can lead to system heating up to 50 to 80 degrees Celsius
- So the brain is far more efficient

Differences

Power - Continued

- Long term, overheating a computer can cause internal damage
- In the (unlikely) event the CPU needs to be replaced, this builds on the issue of missing fault tolerance
- High temperature also affects computer performance, throttling it to prevent overheating

Differences

Signal

- Since perceptrons are non-linear, some layers approximate activation functions
- Leads to vanishing gradient problem, affecting the output
- Vanishing Gradient Problem - earlier layers have less influence over the output than later layers because of compression
- While some functions (ReLU) mitigate this, there is still compression and the output is synchronous

Differences

Signal - Continued

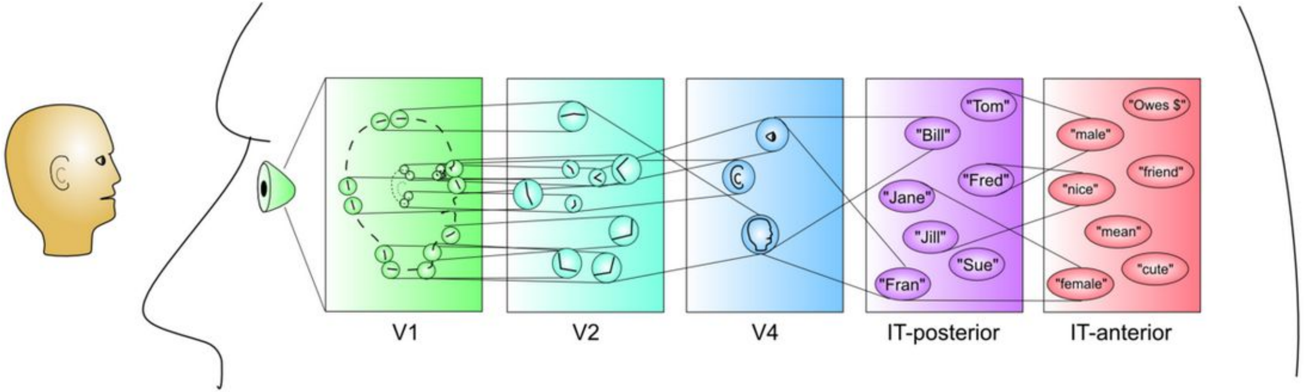
- Artificial neuron networks have a predefined structure
- Perceptrons cannot be added or removed, leading to fixed number of signals
- Weights are also initially randomized, which doesn’t match real neuron expectations
- Only advantage artificial neural networks have is the lack of neural fatigue, though there is a risk of overheating

Hierarchy

- By 1980, Kunihiko Fukushima proposed a model that would focus on two hierarchy problems
  1. Spatial Invariance - different ventral visual pathways integrates over a range of locations for the features of the previous pathway layer
  2. Pattern Discrimination - each of the following layers would deal with more complex combinations of features compared to their previous layers

Hierarchy
Continued

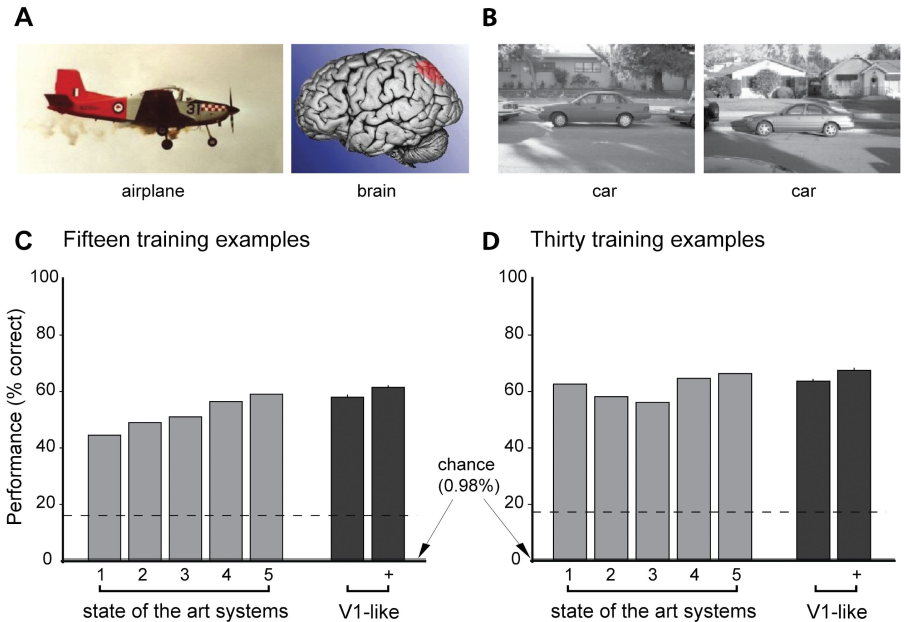
- Hierarchy layers:
  1. First layer (V1) focusing on edges
  2. V2 focusing on combining edges into lines
  3. V4 looking into complex visual features
  4. IT layers determining what is being perceived



Adaptation And Results

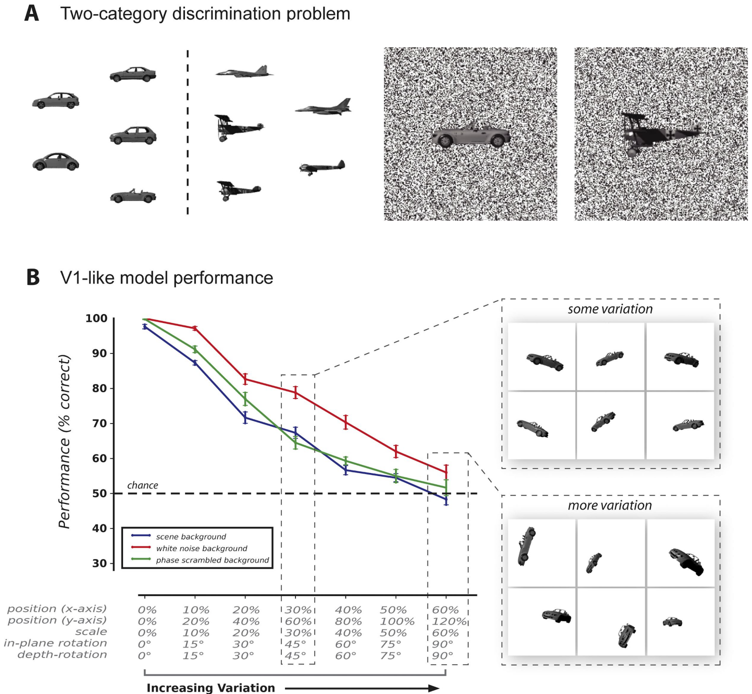
V1 Only

- Some “null” models focusing on the V1 layer are able to best the leading object recognition systems when it comes to looking into standard image recognition tests



Adaptation and Results
V1 Only - Continued

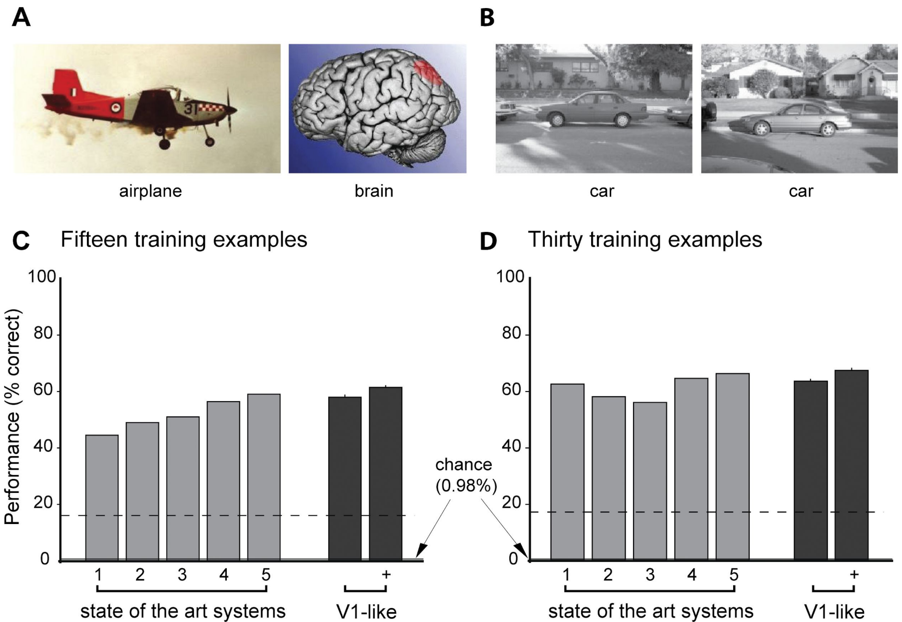
- But these struggle when it comes to real-world recognition due to only working with little variation compared to all possible combinations

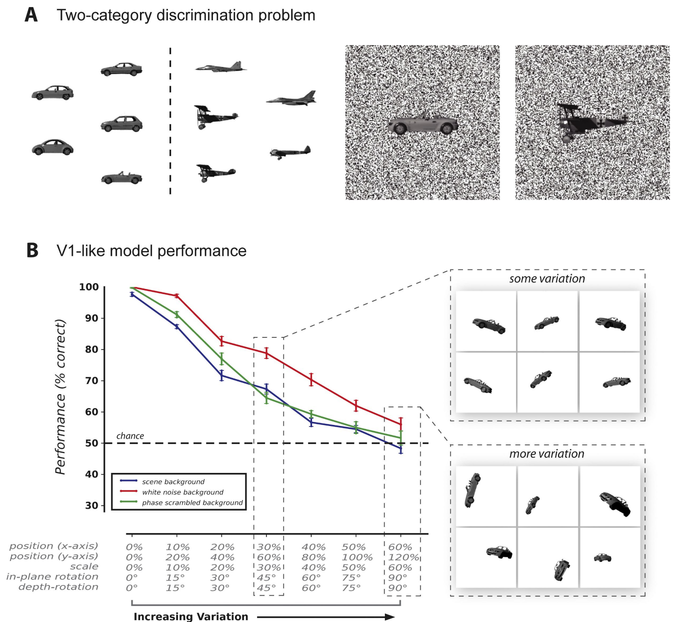


Adaptation and Results

V1 Only - Continued

- When testing such a system on the Caltech101 test set, the accuracy ranged from 61 to 67 percent, but dropped to 24 percent when tested on Caltech256 test set with more real-world variation

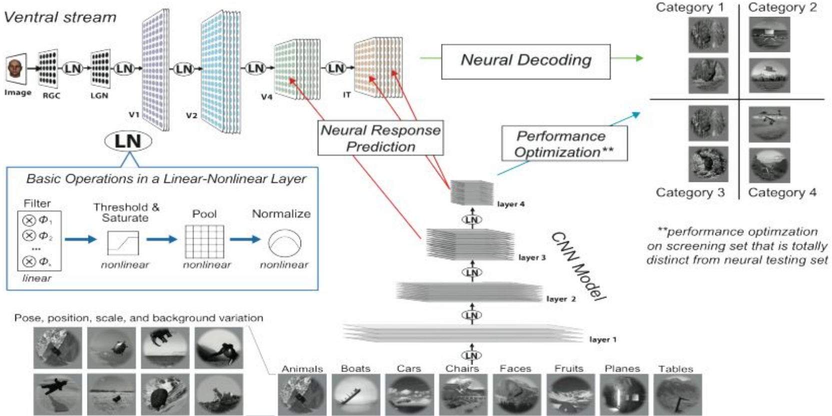




Adaptation and Results

V4 and IT Implementation

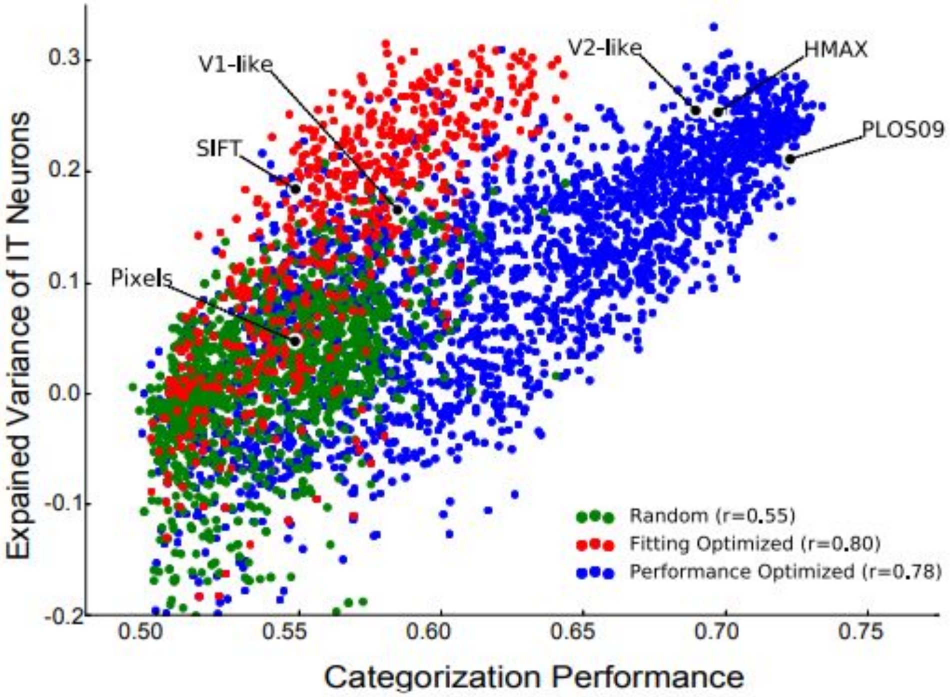
- In an alternative project, thousands of architectures classified as biologically-plausible feedforward neural network models were evaluated to find out the significance of the V4 and IT layers
- In areas of categorization performance and IT neural explained-variance, there was found to be a connection between IT predictivity and model performance
- With this data, a highly optimized feedforward network model was made that can match IT performance across multiple recognition tasks



Adaptation and Results

V4 and IT Implementation - Continued

- While the IT layer project model wasn't a completely custom recreation, it managed to reach levels of accuracy that were superior to V1 models
- This model obtained a 90 percent improvement when compared to other models, including the state-of-the-art models focusing only on V1
- Further testing also made it closer to higher and intermediate cortical areas by working with the "bottom-up" approach with a "top-down perspective", thus setting biologically consistent restraints



Adaptation and Results

SUVM

- Another method - structural unsupervised viewlets model (SUVM)
- Created for automated visual object discover and detection
- Requirements:
  1. Access to large-scale perceptual data
  2. Flexible representations of objects in the data
  3. All done with an effective unsupervised learning algorithm
- Meant to combat the existing deep neural network (DNN) limitations of supervised frameworks for data analysis and the lack of a formal framework meant to utilizing higher levels of abstraction
- Viewlets were meant to act like neurons that fire asynchronously at different stimuli

Adaptation and Results

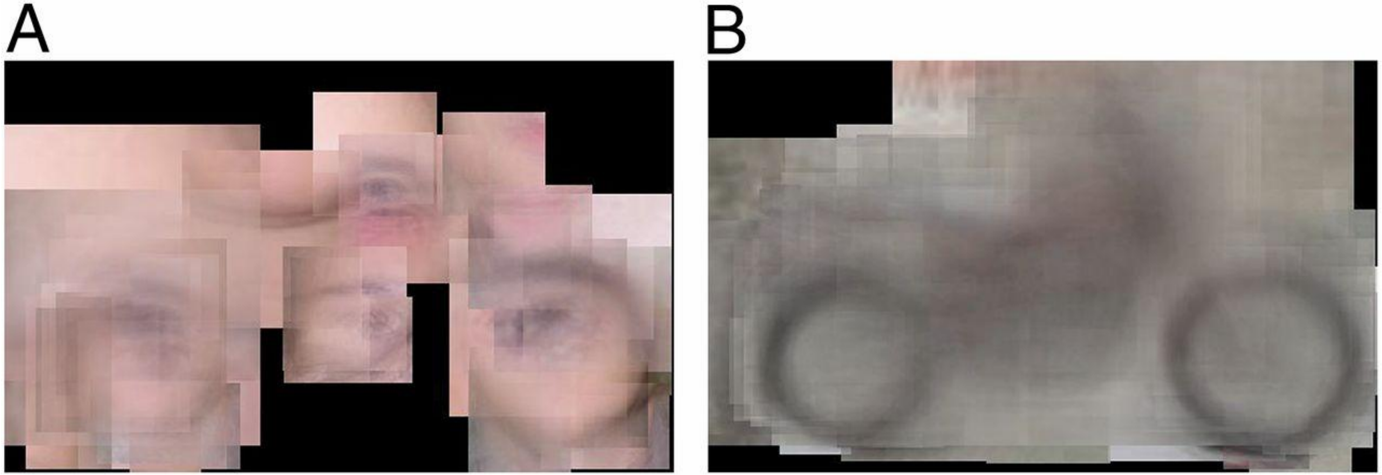
SUVM - Continued

- Models/methods included in the SUVM model:
  1. Spatial relationship network
  2. Configuration-independent parts clustering
  3. Global positional embedding
  4. Positive-only learning setup for estimation
  5. Localization alongside object detection

Adaptation and Results

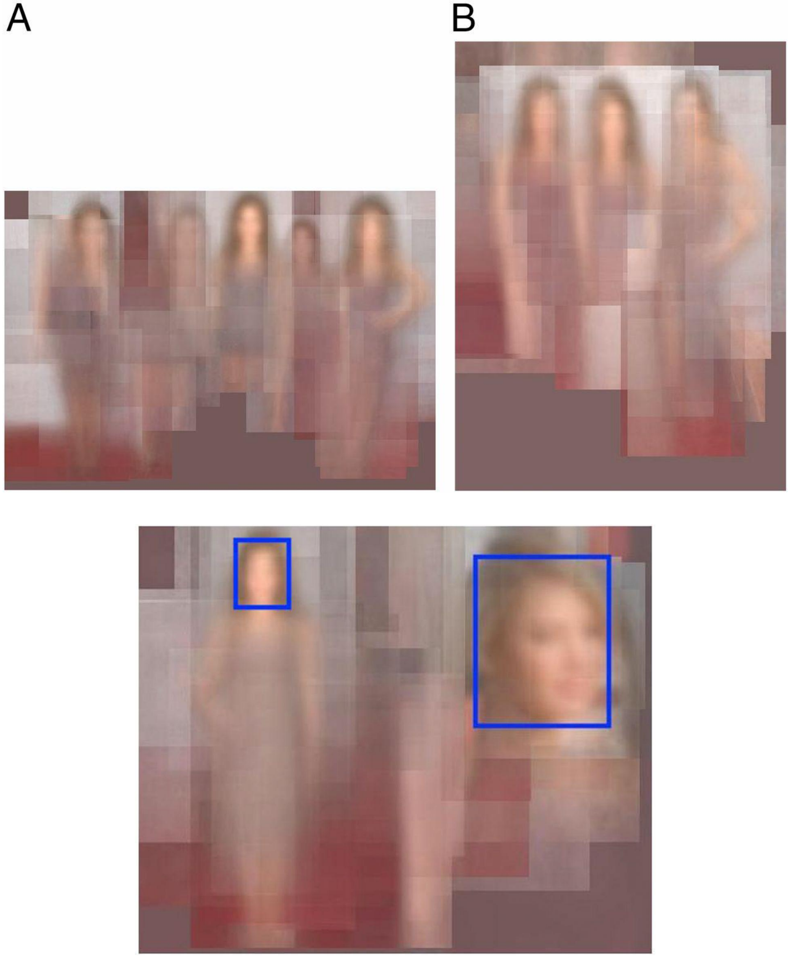
SUVM - Continued

- Results were obtained using the CalTech-4 dataset as well as a celebrity dataset of pictures from the Internet
- When learning solely from face images, it was able to determine where the faces were without error while also being able to detect faces in 50% of the test motorbike images



Adaptation and Results
SUVM - Continued

- However, collective structural integrity was not detected as instances of "face", with airplanes and car categorizations not being as discriminative either when tested individually
- When all the dictionaries were combined, the SUVM was able to show far better improvement, showing that being trained on multiple types can allow for such a system to better determine what an object is
- when focusing strictly on humans, the face was easily able to be found while the torso parts were less successful



Summary

- While artificial intelligence is making great strides, it has yet to match human object recognition quality
- Artificial neural networks still struggle with object variation
- By referencing the human brain hierarchy model, the results of artificial neural networks models can improve their accuracy significantly, even when only implementing the V1 layer

Related Work

Chen, Lichao, et al. Brain-Inspired Automated Visual Object Discovery and Detection. *PNAS*, *National Academy of Sciences*, 2 Jan. 2019, URL: www.pnas.org/content/116/1/96.

Kumar, Ambeshwar & T.M, Rajesh & Tech, M & Student, & Professor, Assistant. (2017). *A Moving Object Recognition using Video Analytics*. URL: https://www.researchgate.net/publication/318598502_A_Moving_Object_Recognition_using_Video_Analytics

Nagyfi, Richard. The Differences between Artificial and Biological Neural Networks. *Medium*, *Towards Data Science*, 4 Sept. 2018, URL: towardsdatascience.com/the-differences-between-artificial-and-biological-neural-networks-a8b46db828b7.

Related Work (continued)

O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., and Contributors (2012). *Computational Cognitive Neuroscience*. Wiki Book, 1st Edition. URL: http://ccnbook.colorado.edu

Pinto, N., Cox, D. D., & DiCarlo, J. J. (2008). Why is real-world visual object recognition hard?. *PLoS computational biology*, 4(1), e27. URL: https://doi.org/10.1371/journal.pcbi.0040027

Verschae, Rodrigo, Ruiz-del-Solar, Javier. Object Detection: Current and Future Directions. *Frontiers in Robotics and AI*, 2. URL: https://www.frontiersin.org/article/10.3389/frobt.2015.00029

Yamins, Daniel, et al. *Predicting IT and V4 Neural Responses With Performance-Optimized Neural Networks*. 2013, URL: https://neuroailab.stanford.edu/uploads/2/4/9/2/24920889/it_v4_prediction.pdf

Questions?