MENTAL ARCHITECTURE

Understanding Human Vision, Computer Vision, and Deep Learning

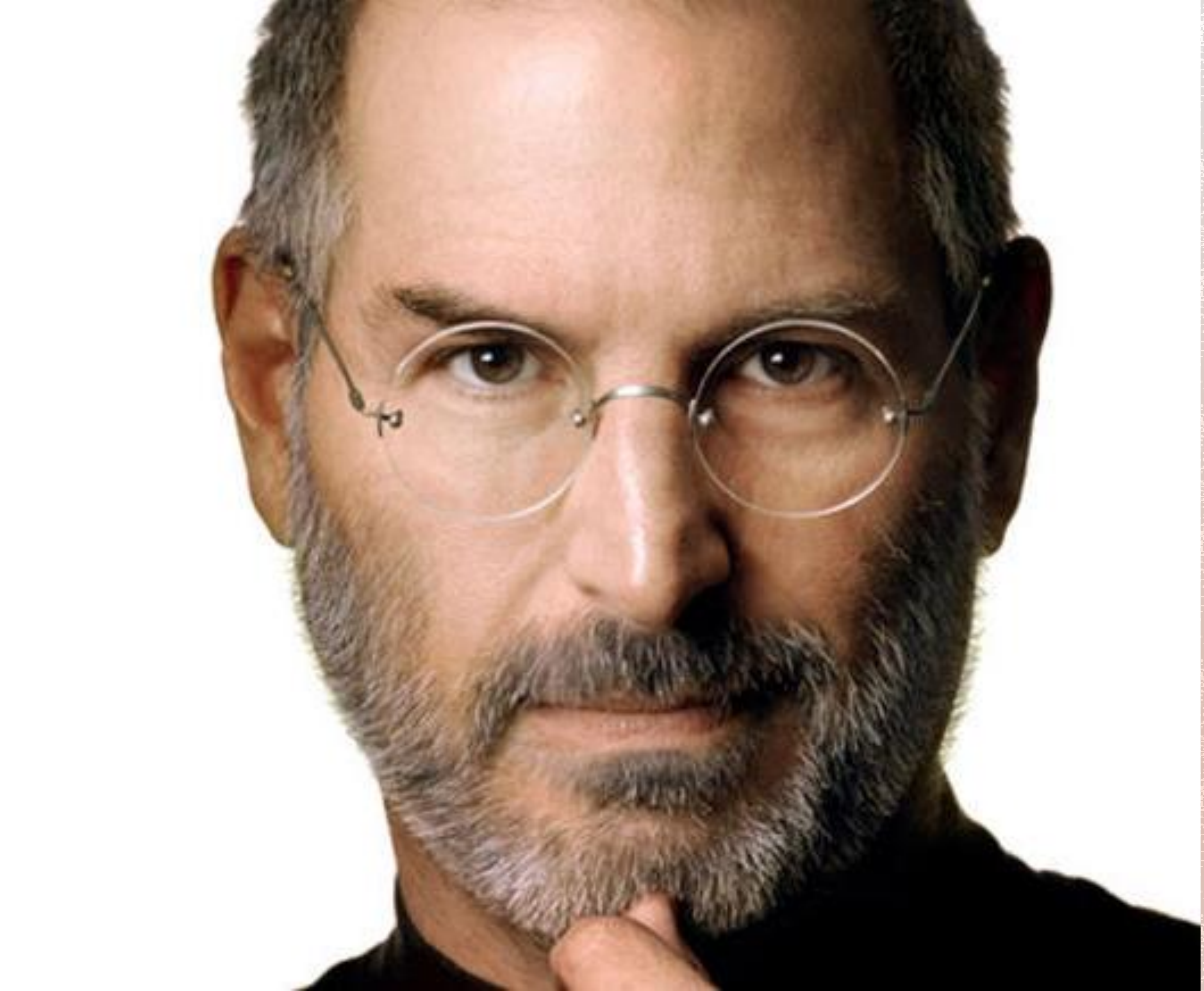
Tuesday, October 22, 2019 (2:00 PM – 3:50 PM, Room 1003)

Presented by Howard Blumenfeld

PLEASE REMEMBER TO SIGN IN ON THE OFFICIAL FLEX DAY ATTENDANCE SHEET (BACK OF ROOM), TAKE A SURVEY TO COMPLETE LATER, AND GRAB YOUR RAFFLE TICKET!

MA

Mental Architecture



STEVE JOBS

"When you grow up you tend to get told the world is the way it is and your life is just to live inside the world. Try not to bash into the walls too much. Try to have a nice family, have fun, save a little money. That's a very limited life.

Life can be much broader once you discover one simple fact: Everything around you that you call life was made up by people that were no smarter than you and you can change it, you can influence it, you can build your own things that other people can use.

Once you learn that, you'll never be the same again."

AUTHOR BACKGROUND

Born and raised in the Santa Clarita Valley, CA

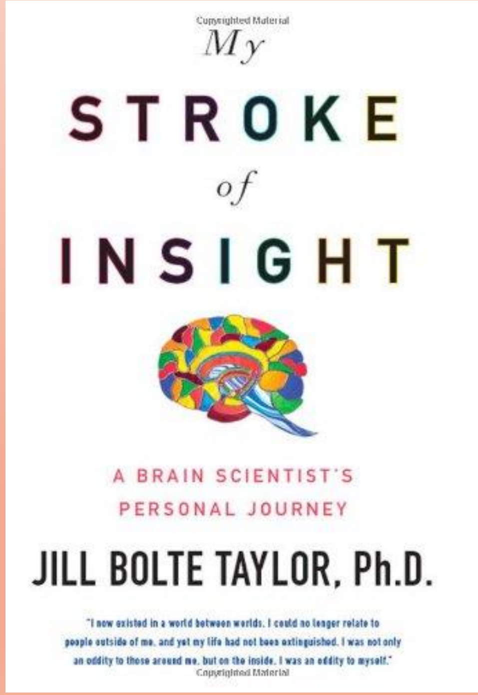
Academic background in mathematics, psychology, and literature

Strong interest in cognitive science and philosophy

16 years experience teaching mathematics in higher education (11 of those at LPC)

INSPIRATION FOR BOOK

Jill Bolte Taylor's Book which narrates her experience with a major stroke in her left hemisphere.



Jill Bolte Taylor is an American neuroanatomist, motivational speaker, author, and CEO of My Stroke of Insight, Inc.

JILL BOLTE TAYLOR

"And I lost my balance and I'm propped up against the wall. And I look down at my arm and I realize that I can no longer define the boundaries of my body. I can't define where I begin and where I end. Because the atoms and the molecules of my arm blended with the atoms and molecules of the wall. And all I could detect was this energy."

OTHER SOURCES OF INSPIRATION

"The beginning is the end, and the end is the beginning."

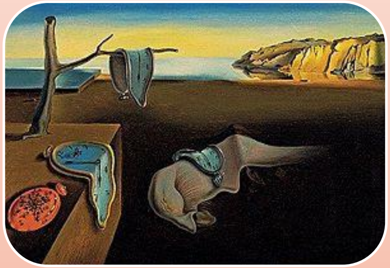
"Dark" co-creators Baran bo Odar and Jantje Friese

"The question is not where. But when."

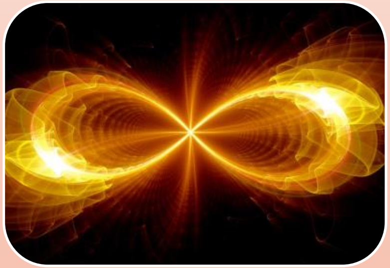
WHAT IS MENTAL ARCHITECTURE?

THE SEVEN CHAPTERS OF MENTAL ARCHITECTURE

No Space



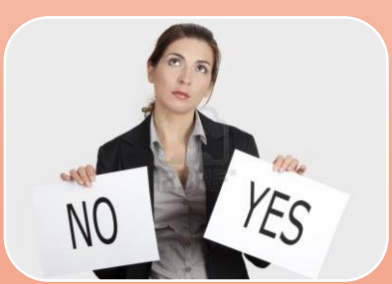
No Point in Time



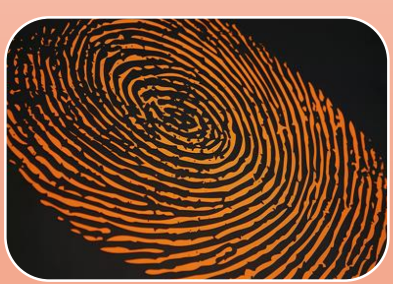
No Infinity

No Perspective

No Normal

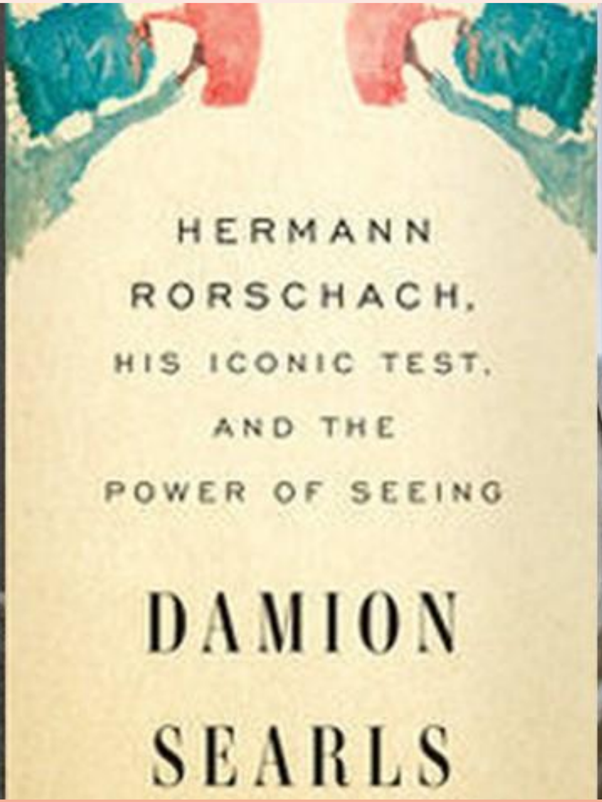


No Decisions



No Identity

NO SPACE



"Vision is the sense that both operates at a distance, unlike touch and taste, and can be focused and directed, unlike hearing and smell. We can pay attention to certain noises or odours, or try to ignore them, but we cannot blink our ears or aim our nose: the eye is far more active, under far more control. Seeing is our best perceptual tool – our foremost way to engage with the world."

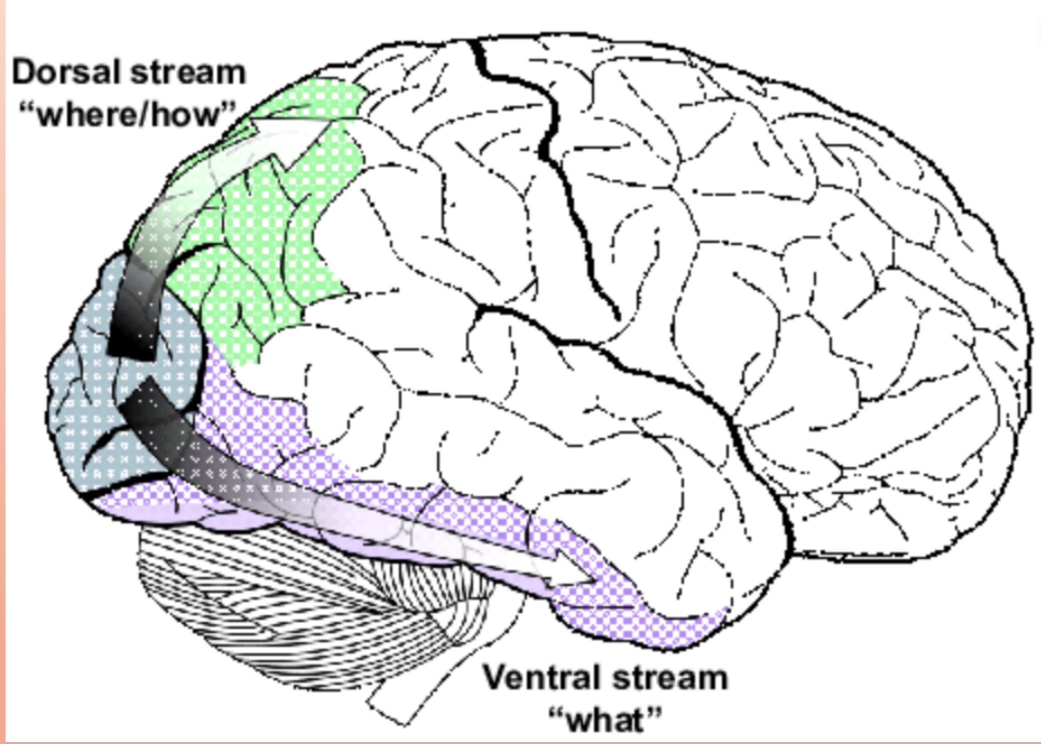
Damion Searls

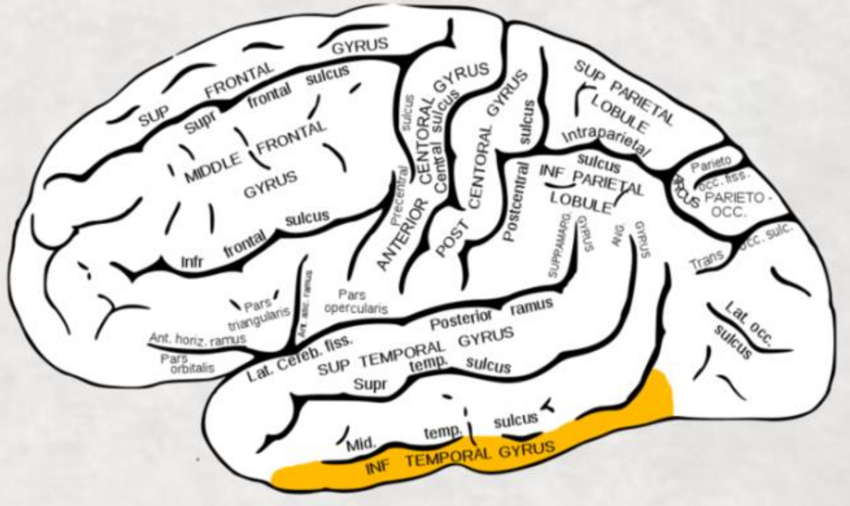
NO SPACE: OBJECT RECOGNITION

How do you distinguish objects from one-another? How does your brain know the difference between an apple and an orange?

How does your brain filter out irrelevancies and other types of “noise” in your visual field to allow you to see what you need to see?

THE VENTRAL VISUAL STREAM & INFERIOR TEMPORAL CORTEX





INFERIOR TEMPORAL CORTEX (IN YELLOW).

Information is communicated hierarchically from the visual cortex toward the inferior temporal cortex (IT) – regions heavily implicated in object recognition (including faces). The IT is also implicated in filtering out irrelevant sensory data.

THE VENTRAL VISUAL STREAM

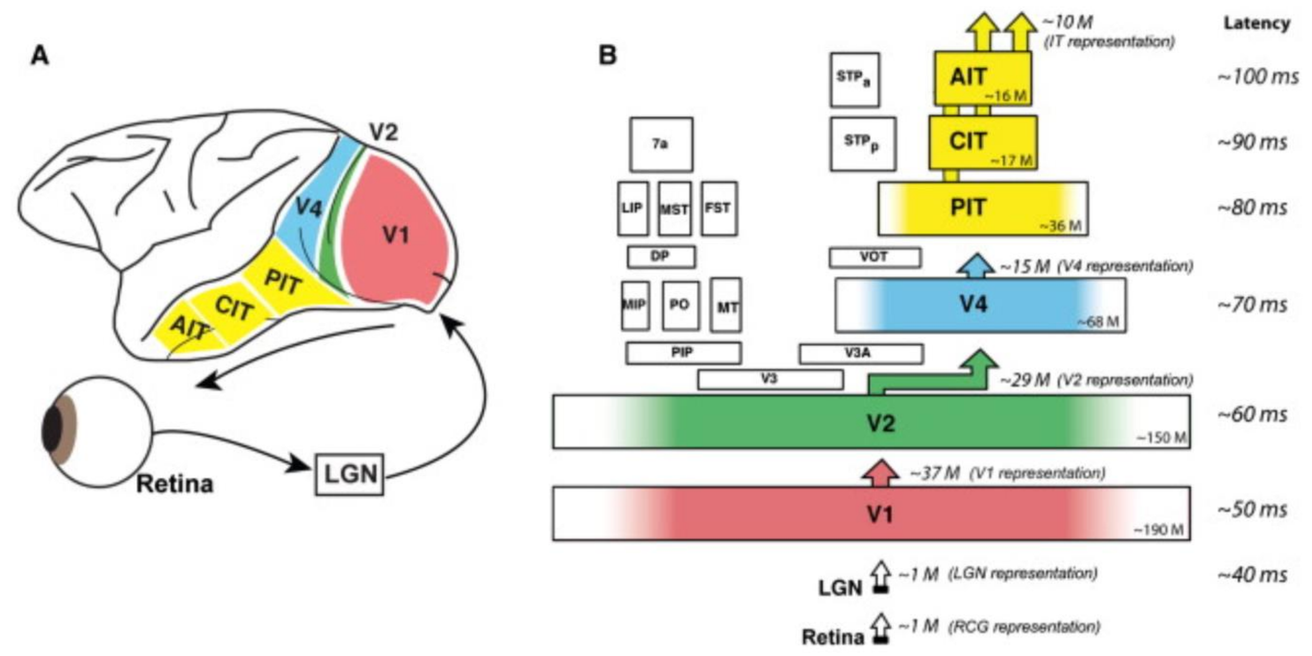
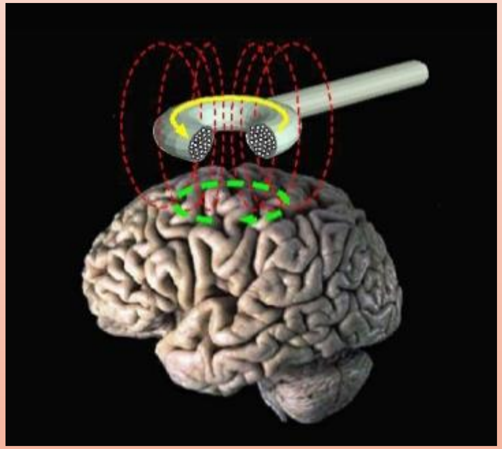


Figure 3 The Ventral Visual Pathway

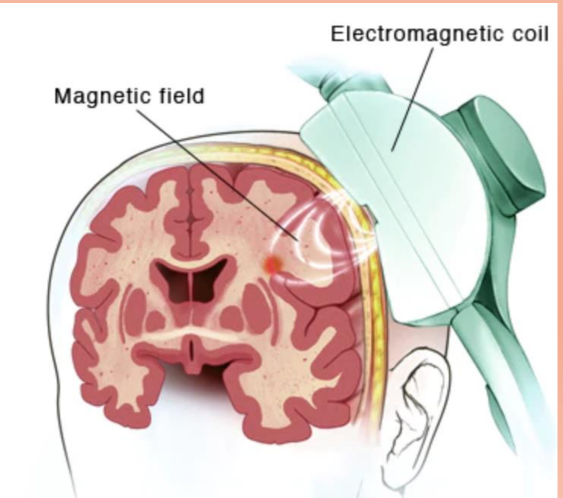
This figure illustrates the Ventral Visual Stream in a macaque monkey brain.

TRANSCRANIAL MAGNETIC STIMULATION (TMS)



Noninvasive procedure involving the production of magnetic fields to stimulate neuron activity in specific brain regions.

An electromagnetic coil is placed on top of the scalp to initiate this procedure, which is often used as a clinical treatment for major depressive disorders.



EFFECTS OF TMS ON THE VENTRAL VISUAL STREAM

TMS applied to the ventral visual stream causes temporary object recognition deficits, particularly with faces (prosopagnosia)

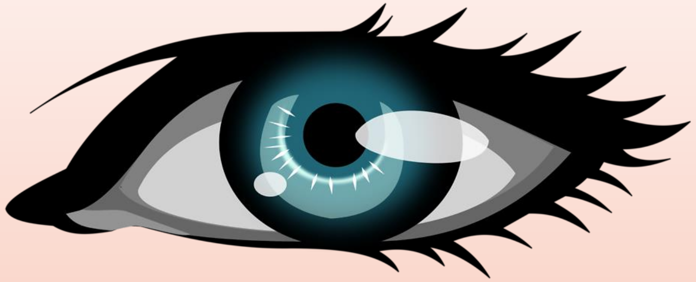
CORE OBJECT RECOGNITION

- James J. Dicarlo, Davide Zoccolan, and Nicole C. Rust sought to discover evidence of how the brain solves the visual object recognition problem.
- Their task was to find direct evidence of computational algorithms used for object recognition.
- Their study involved examining how the architecture and plasticity of the ventral visual stream worked together with retinal ganglion cells in the eye, visual cortex neurons, single IT neuron and IT neural network responses to form an accurate solution for object representation in IT.

CORE OBJECT RECOGNITION

- Core object recognition is the ability in primates to rapidly identify objects in a very short amount of time (typically 200-500 ms).
- The major problem with object recognition is being able to identify objects across a wide range of visual conditions (variations in position, scale, pose, illumination, context, deformations). If the brain were not able to correctly perform this task, people would see the same object viewed from different perspectives or in different conditions as entirely different objects.
- The fascinating issue is that each different appearance (or perspective) of an identical object creates a different response in the retina, but somehow the visual cortex is able to sort through this data and correctly and quickly identify the object each time (usually in under 500ms).

CORE OBJECT RECOGNITION



"Each encounter of the same object activates an entirely different retinal response pattern and the task of the visual system is to somehow establish the equivalence of all of these response patterns while, at the same time, not confuse any of them with images of all other possible objects."

-DiCarlo et al.

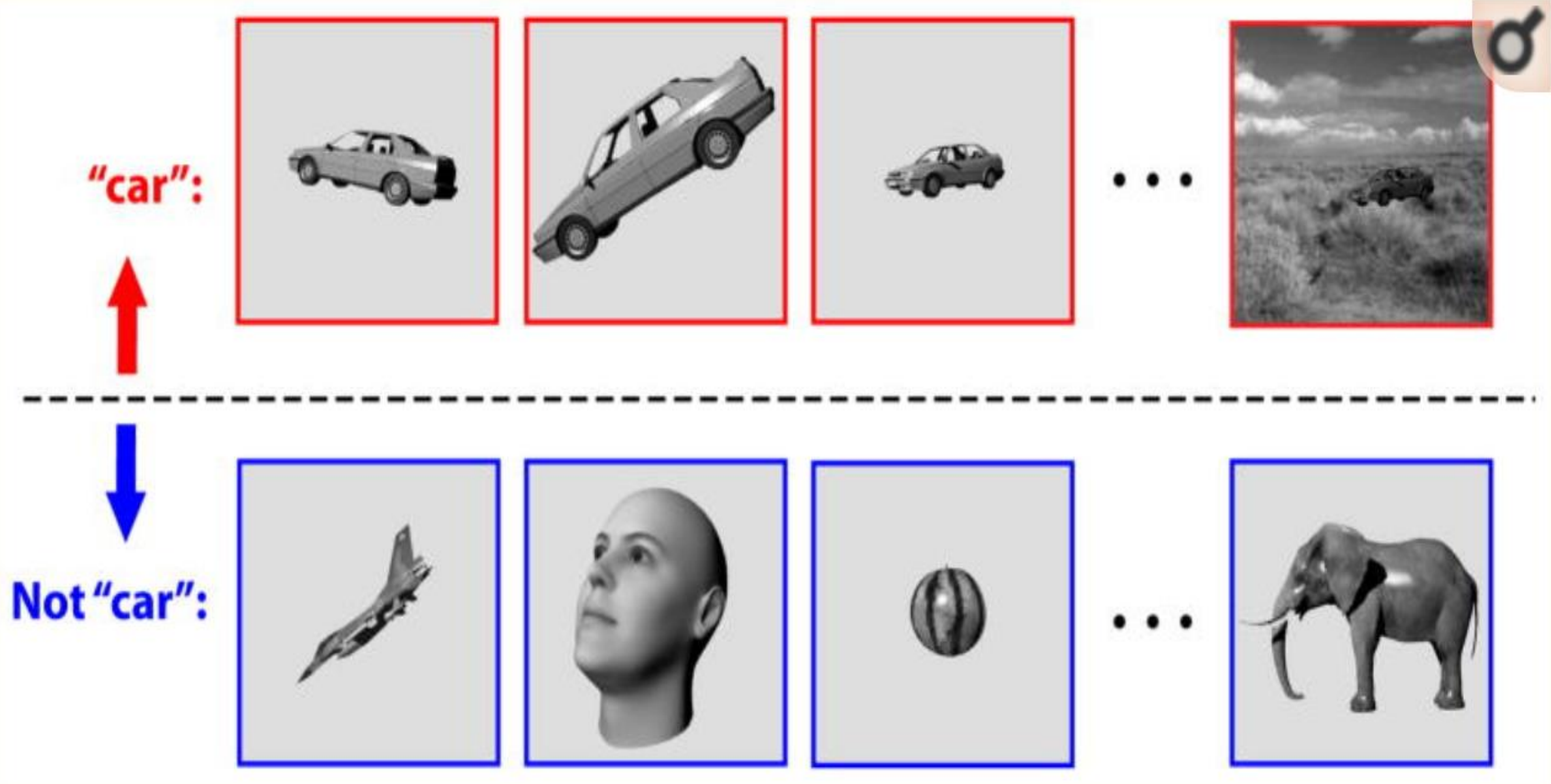


Figure 1

Core object recognition

is the ability to rapidly (<200 ms viewing duration) discriminate a given visual object (e.g., a car, top row) from all other possible visual objects (e.g. bottom row) without any object-specific or location-specific pre-cuing (e.g. (DiCarlo and Cox, 2007). Primates perform this task remarkably well, even in the face of identity-preserving transformations (e.g., changes in object position, size, viewpoint, and visual context).

OBJECT INVARIANCE

The ability to correctly identify objects rapidly is a necessary tool for human survival.

How does our brain correctly and seamlessly identify objects despite the seemingly endless variation in object parameters and contexts?

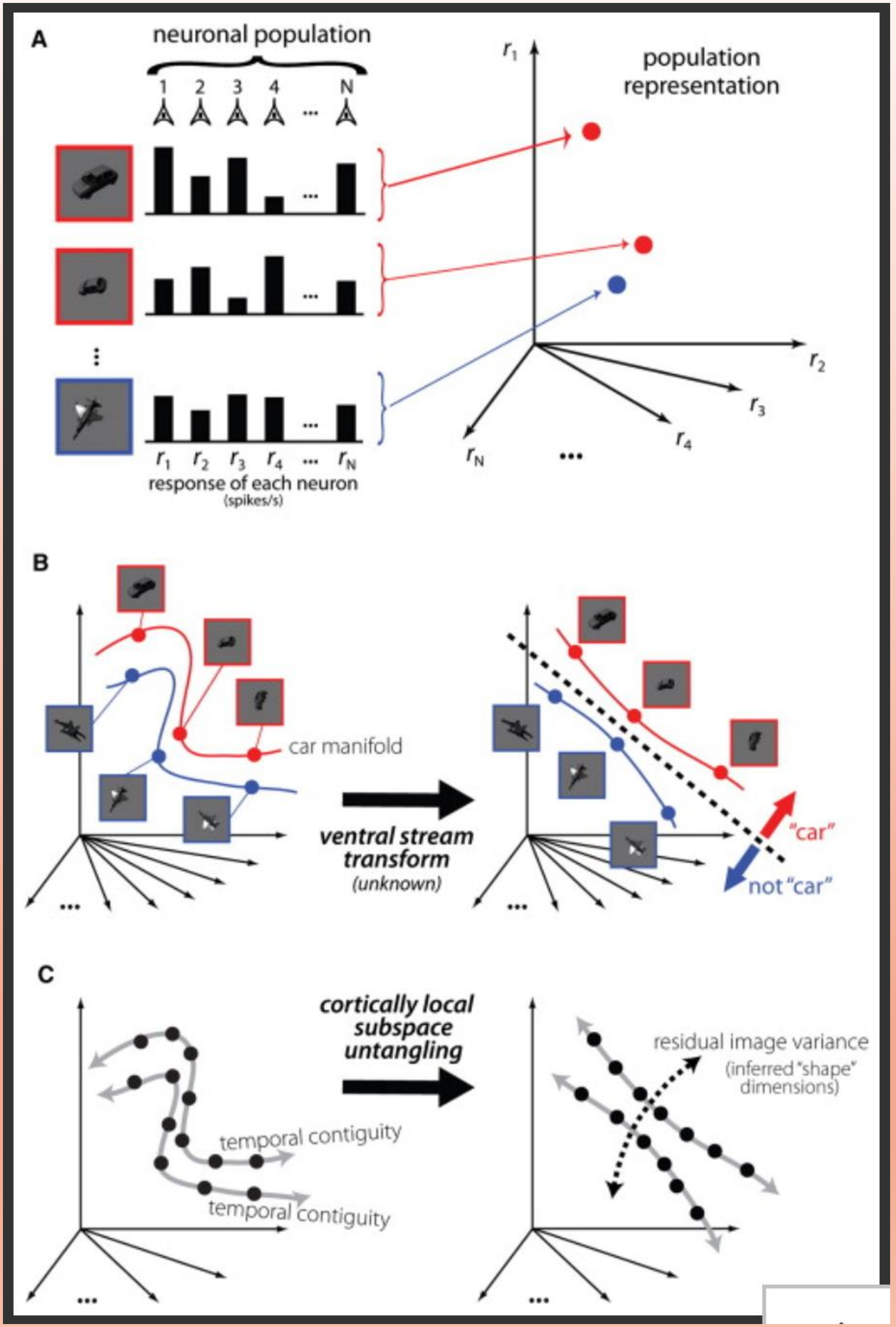
The brain groups objects into different classes, so that they can be correctly categorized. Not all cars look the same, but they share enough common features to be labeled as such.

Interesting question - At what point does a car stop being a car and become something else entirely?

- Figure A illustrates the result of a collection of neurons responding to a specific orientation of an object. Each identity-preserving transformation (rotations, transformations, scalings, etc.) to the object results in a different neuronal "response vector" (illustrated by the red and blue rays). The red rays are for cars and the blue rays are for objects which are "not cars," in this example.

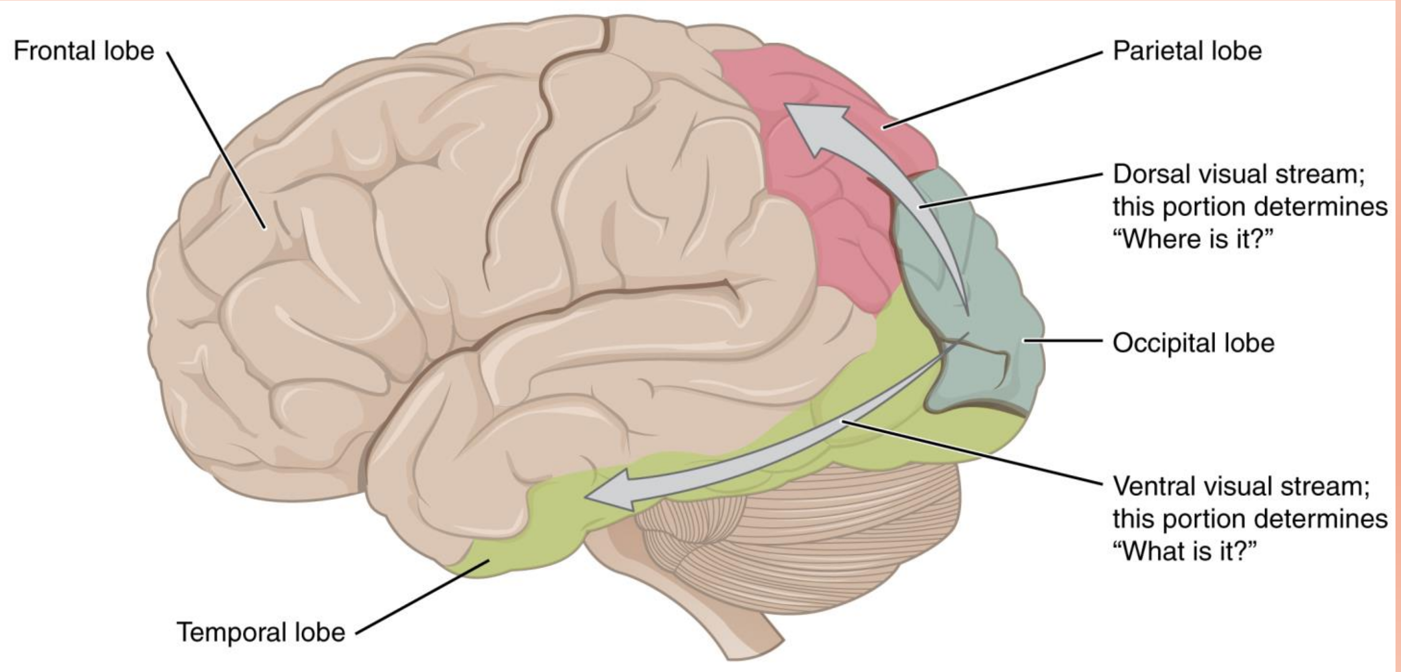
- Figure B shows a "car manifold" and a "non-car manifold" in which all possible orientations of a particular object are represented graphically. These distinct manifolds may be tangled up together. A ventral stream transformation function (most likely iterative) along with statistical functions (i.e., simple weighted summation) are applied to the manifolds to separate out the cars from the non-cars.

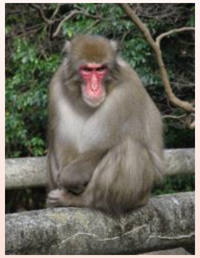
- Figure C illustrates how the brain untangles representations of objects without labels by using temporal and contextual cues to disambiguate different objects in your frame of reference.



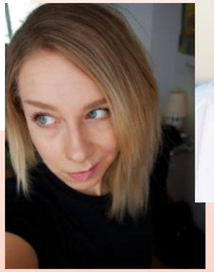
THE VENTRAL STREAM

- Lesions in the posterior ventral stream can cause partial blindness in part of the visual field.
- Lesions in the anterior ventral stream can cause object recognition difficulties, especially for complicated objects.
- Damage to the temporal lobe can also result in problems with object recognition.





THE VENTRAL STREAM

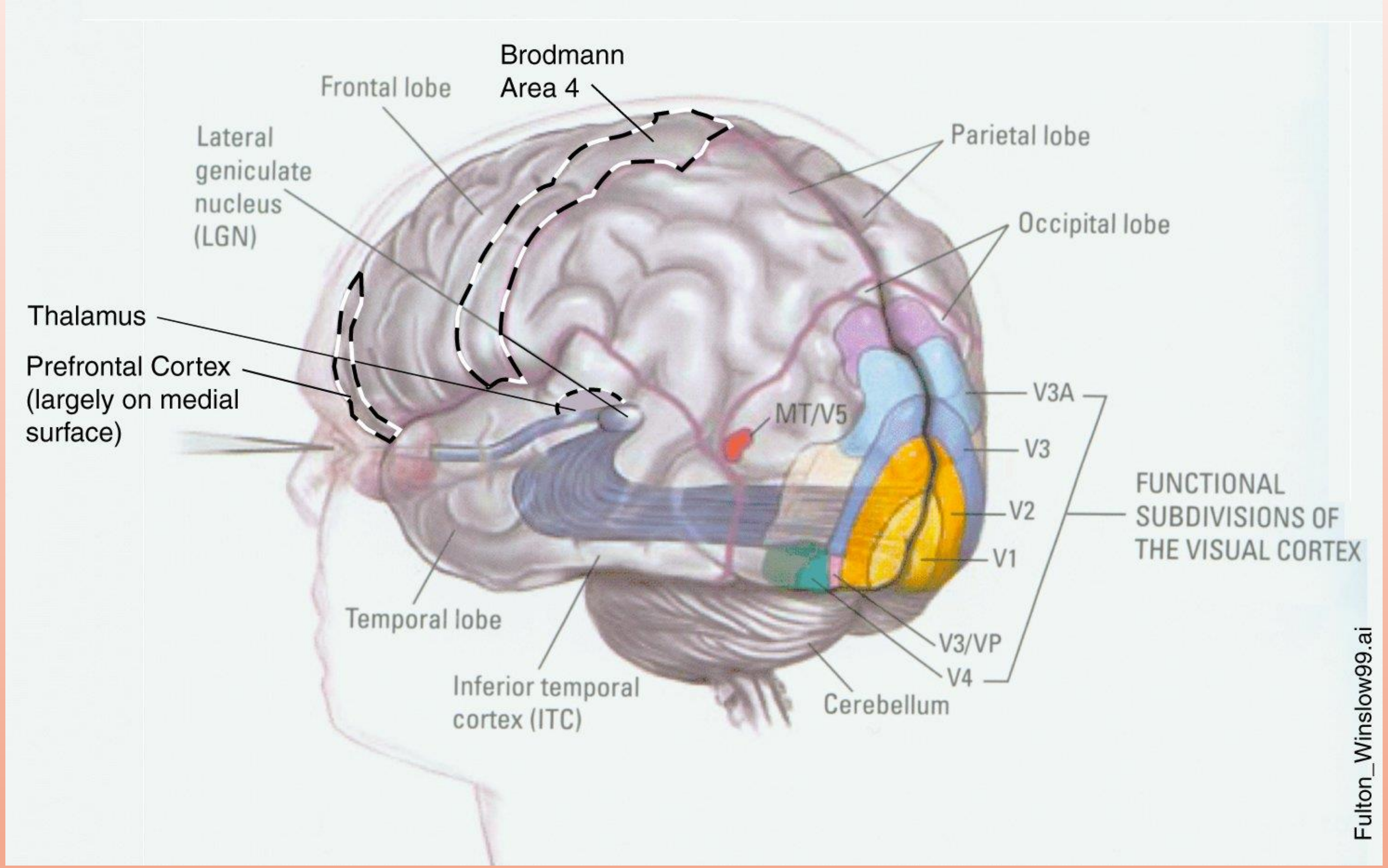


There is some uncertainty about how the IT is organized. The common practice is to break it down into posterior (pIT), central (cIT), and anterior (aIT), but that may not be a detailed-enough characterization.

In one study of monkey brains, researchers found that there were at least 3-6 regions in the IT that were associated with the recognition of faces.

The idea that cells from the retina directly map to images in the brain in a straightforward way may be outdated, and object categorization (influenced by social and behavioral cues) may play an indirect role in object recognition.

There is also much evidence in the human brain that the IT and visual cortex are both heavily implicated in processing faces, perhaps moreso than any other class of objects.



THE VISUAL CORTEX

The visual cortex is organized into different regions, each of which has a six-layered structure.

Each visual area shares the same neuronal connectivity patterns, and they appear to have a hierarchical structure to them.

The most likely scenario supported by research is that visual sensory data (related to object recognition) is processed by the retina (via retinal ganglion cells) and then the information is sent to the lateral geniculate nucleus (LGN) (inside the thalamus), and then it passes through V1 to V2, and from V2 to V4, and then along the ventral visual stream to the IT.

It is important to note that information flow is not entirely linear and there are likely some feedback loops encountered along the way.

THE OBJECT RECOGNITION PROCESS IS LIKE AN ASSEMBLY LINE

"Your job, as a local cortical subpopulation, is to take all your neuronal afferents (your input representation) and apply a set of nonlinearities and learning rules to adjust your input synaptic weights based on the activity of those afferents. These nonlinearities and learning rules are designed such that, even though you do not know what an object is, your output representation will tend to be one in which object identity is more untangled than your input representation."

-DiCarlo et al.

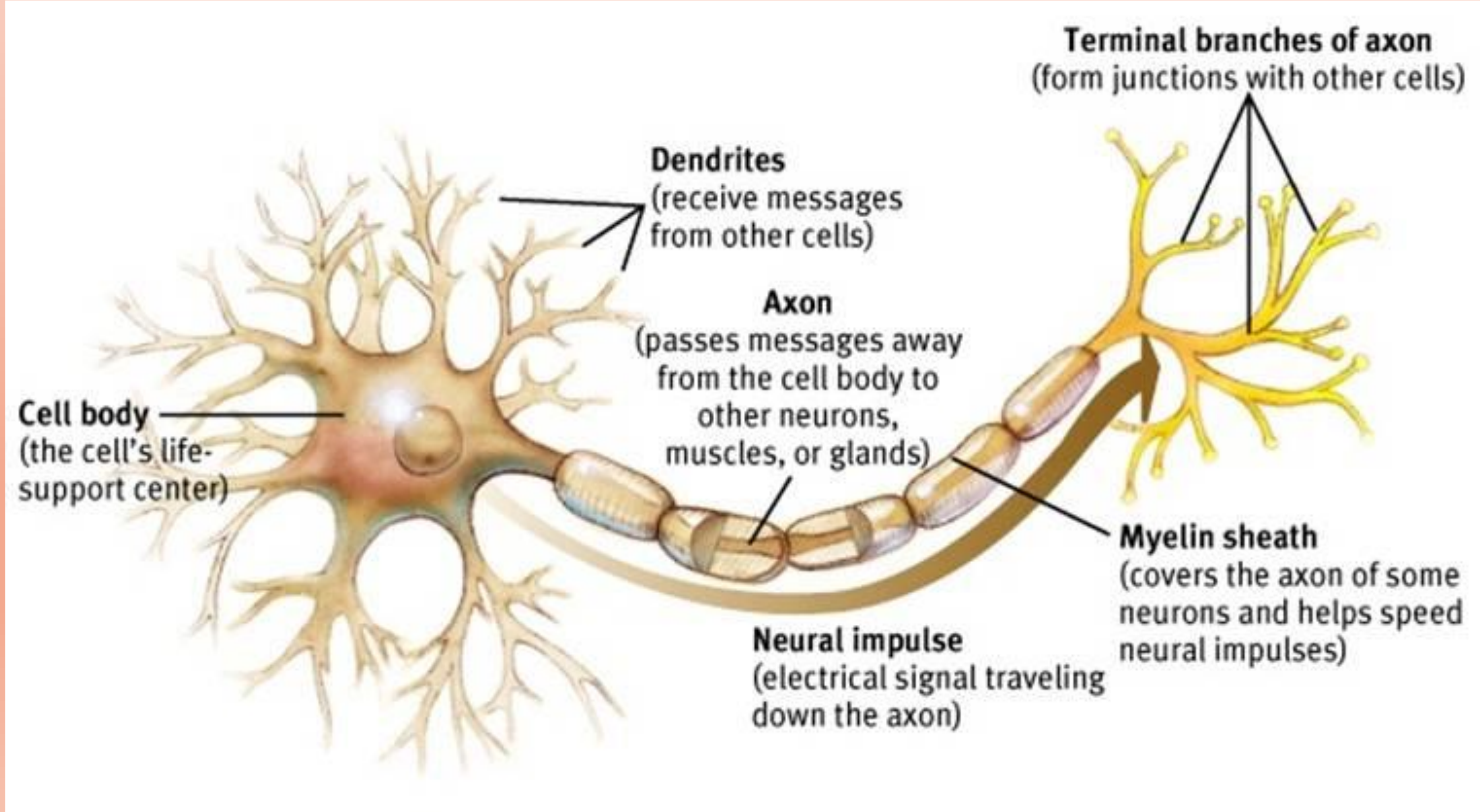
VENTRAL STREAM PROCESSING

On average, it only takes about 100 ms from the time visual imagery (photons) enters the eye and is processed by the retinal ganglion cells to the time that visual data is processed hierarchically by the visual cortex and enters into the IT. Within 50 ms of additional processing by IT (to the original 100 ms delay), core object recognition occurs.

There is evidence to support that the ventral stream generates neuronal activity that results in robust *invariant* core object recognition abilities. Low-level visual processing occurs in area V1, where information from the retina is mapped and basic object features can be detected. By the time visual data makes its way down the ventral stream to IT, that data is heavily processed and fine-tuned.

SO HOW ARE OBJECTS REPRESENTED NEUROLOGICALLY?

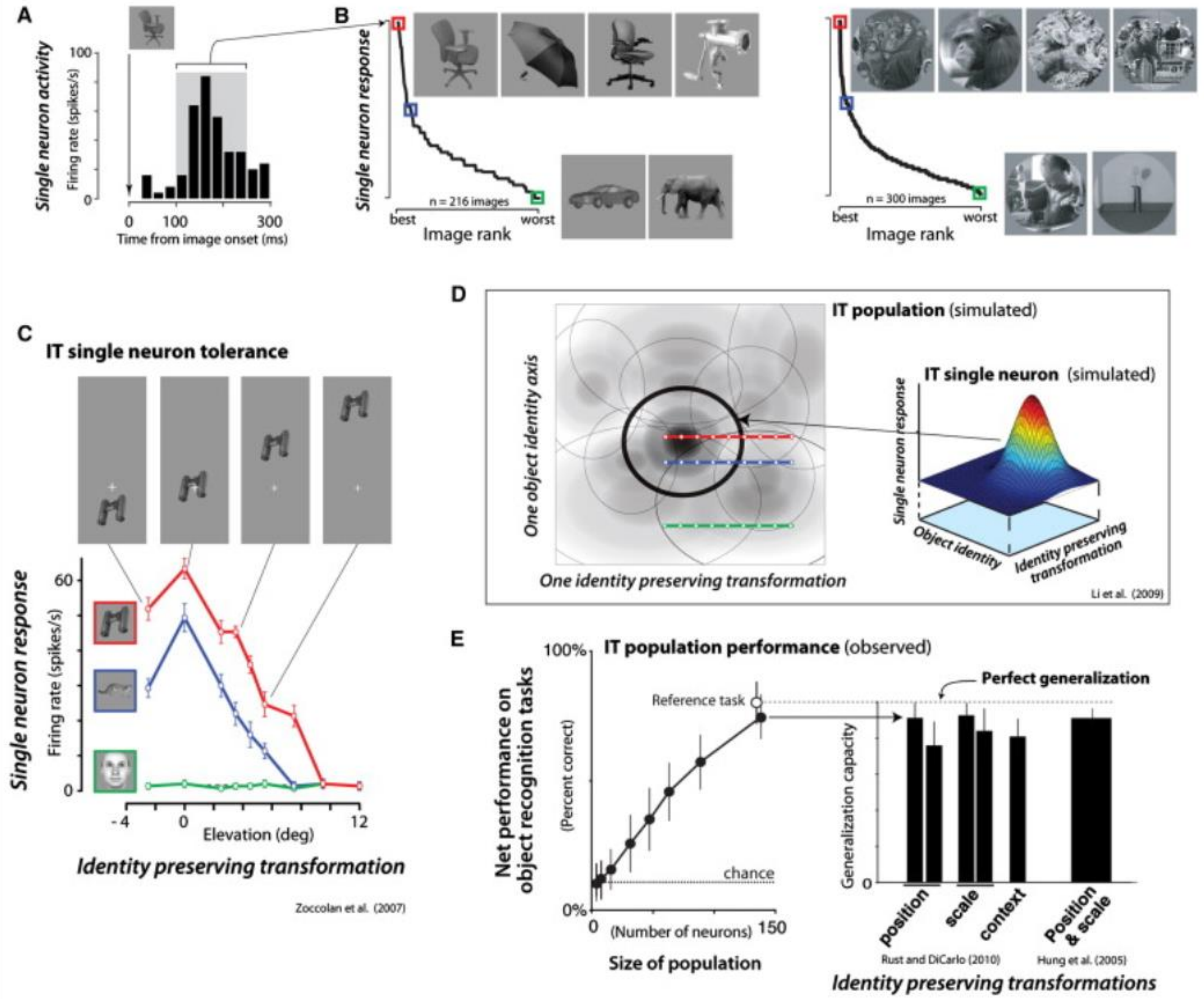
- Before we can answer this question, we must understand how signal processing between neurons occurs.



VENTRAL STREAM PROCESSING

"The only known means of rapidly conveying information through the ventral pathway is via the spiking activity that travels along axons. Thus, we consider the neuronal representation in a given cortical area (e.g., the "IT representation") to be the spatiotemporal pattern of spikes produced by the set of pyramidal neurons that project out of that area (e.g., the spiking patterns traveling along the population of axons that project out of IT."

-DiCarlo et al.

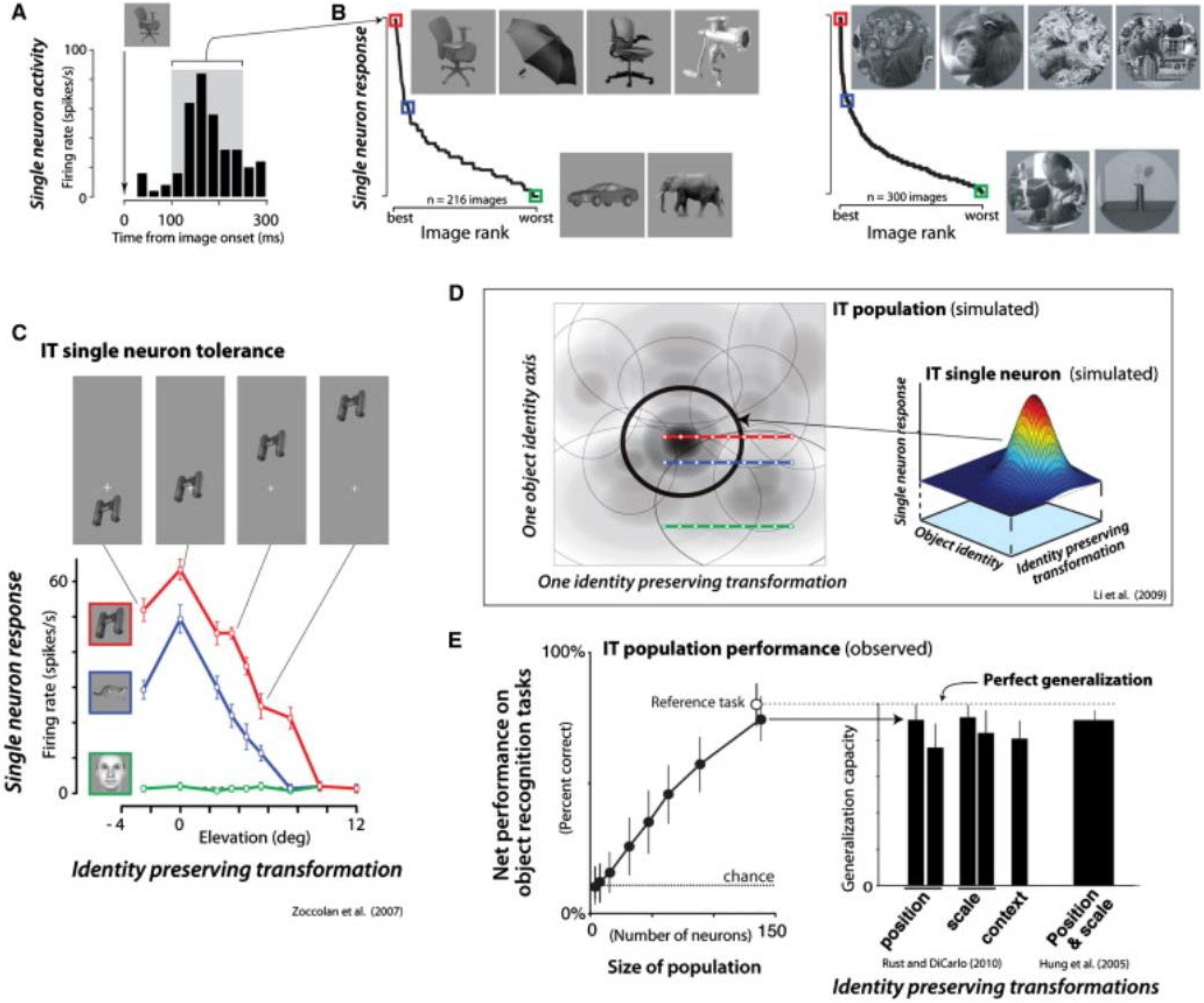


VENTRAL STREAM PROCESSING

Figure A – Electrical activity from a single IT neuron in response to a specific object (a chair).

Figure B – Electrical spiking response from a single neuron (same as in Figure A) to a series of images (the chair elicited the strongest response) and then electrical spiking responses from a different IT neuron (graph to the right) (a photograph of people elicited the strongest response.

Figure C – Electrical activity from a single neuron exposed to a high-response, medium-response, and low-response series of images subject to changes in their visuospatial orienatations.

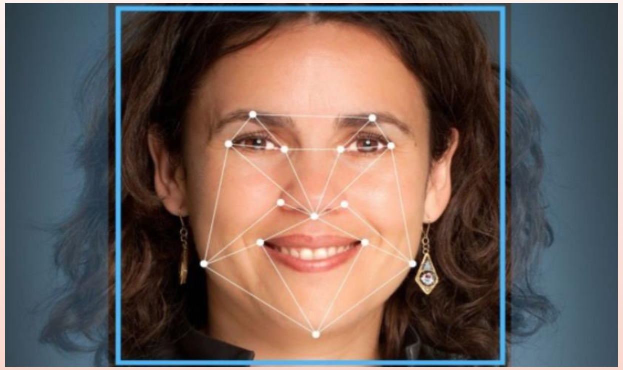


VENTRAL STREAM PROCESSING

Figure D – A single neuron isolated from a neural network showing its electrical response pattern to a particular object under shape variables and position/orientation variables.

Figure E – The first graph illustrates the accuracy of object recognition compared to the size of the neuronal population (just a few hundred neurons). The second graph demonstrates, for this particular number of neurons, that the network is highly accurate with object identification across different variables in position, scale, context, or even position and scale together.

THE INFERIOR TEMPORAL CORTEX



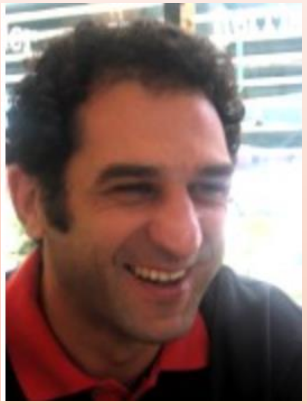
- Small collections of randomly chosen neurons (a few hundred) will elicit different, but reasonably similar patterns of electrical "spiking" responses over brief time intervals.
- These responses are averaged via "simple weighted summations."
- The IT weighted summation scheme can account for core object recognition, including under invariant conditions (scaling, contextual, rotations, etc.)
- Other studies using fMRI-targeted clusters of IT neurons showed that the IT is also responsible for facial recognition and face discrimination also across invariant conditions.

THE LAWS OF RAD IT

- In 2012, Najib J. Majaj, Ha Hong, Ethan A. Solomon, and James J. DiCarlo published a research paper *Simply Weighted Sums of Inferior Temporal Neuronal Firing Rates Accurately Predict Human Core Object Recognition Performance.*

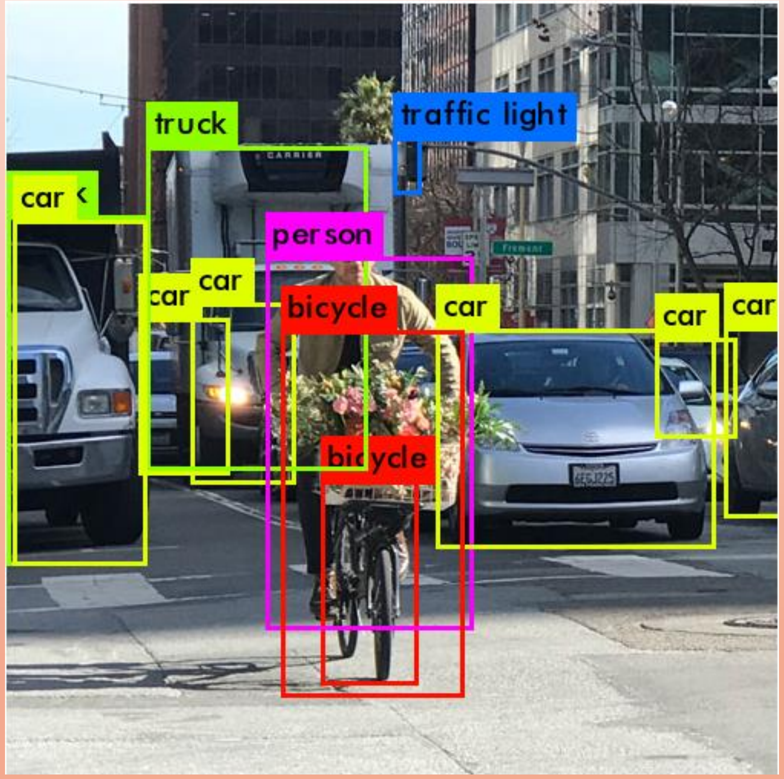
- LaWS of RAD IT stands for "simple, learned weighted sums of randomly selected average neuronal responses spatially distributed over monkey IT."

- This study was the first of its kind to provide a link between the way monkeys' IT identifies objects and humans do. They found that monkey's object recognition neurologically in their IT across 64 different trials consisting of thousands of images subject to invariant conditions accurately predicted humans' object recognition neurologically in the IT.



COMPUTER-BASED OBJECT RECOGNITION

COMPUTER VISION AND DEEP LEARNING



- Machine-learning algorithms for computer vision have come a long way in the past decade.
- An application of deep learning involves training neural networks to recognize images through mathematical and statistical analyses of extensive data.
- The main applications of computer vision presently include object detection and object classification.

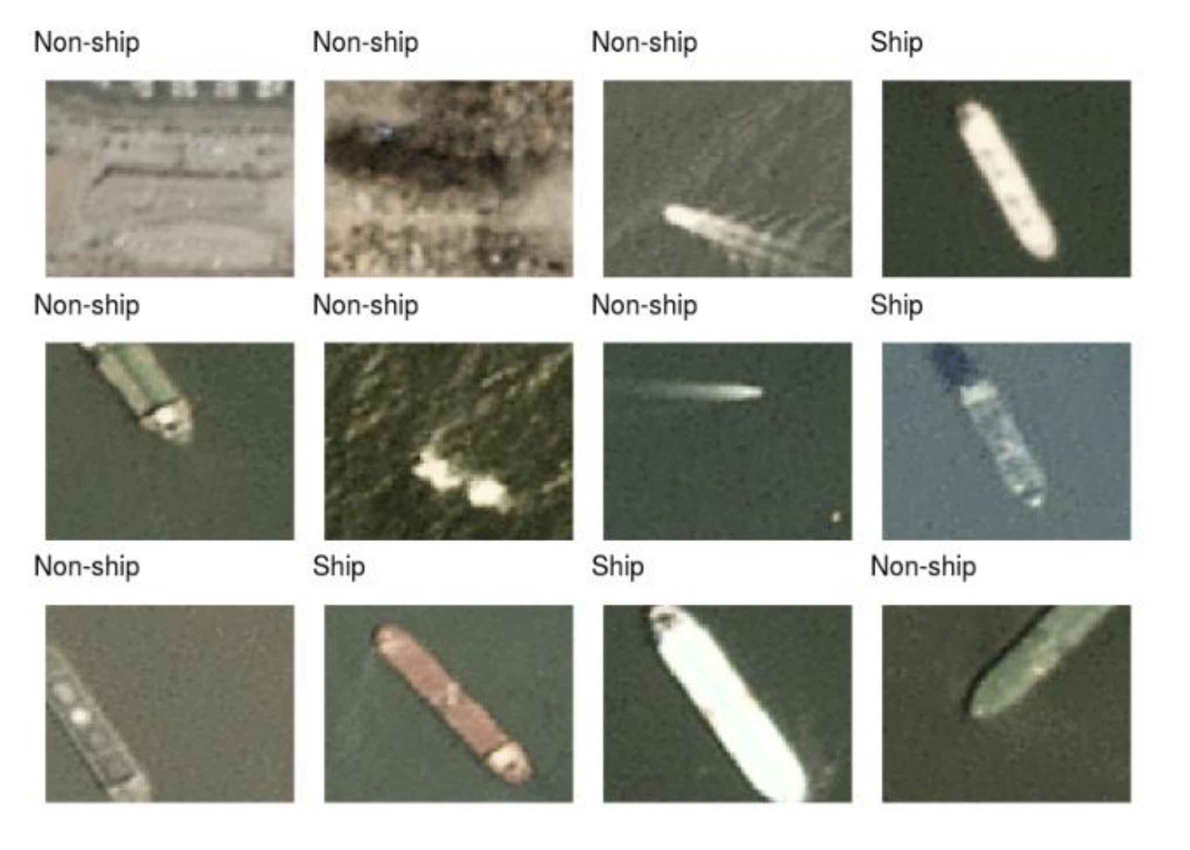
CONVOLUTIONAL NEURAL NETWORKS

- CNNs are a type of deep-learning neural network that can be used for object classification purposes.

- Tensor Flow was created by the "Google Brain" research team as an open-source machine-learning library.

- An API is an Application Programming Interface that is responsible for the communication between two software interfaces.

- You can think of an API as allowing a programmer to access packages from another provider to easily create complex and sophisticated programs without having to develop everything from scratch.

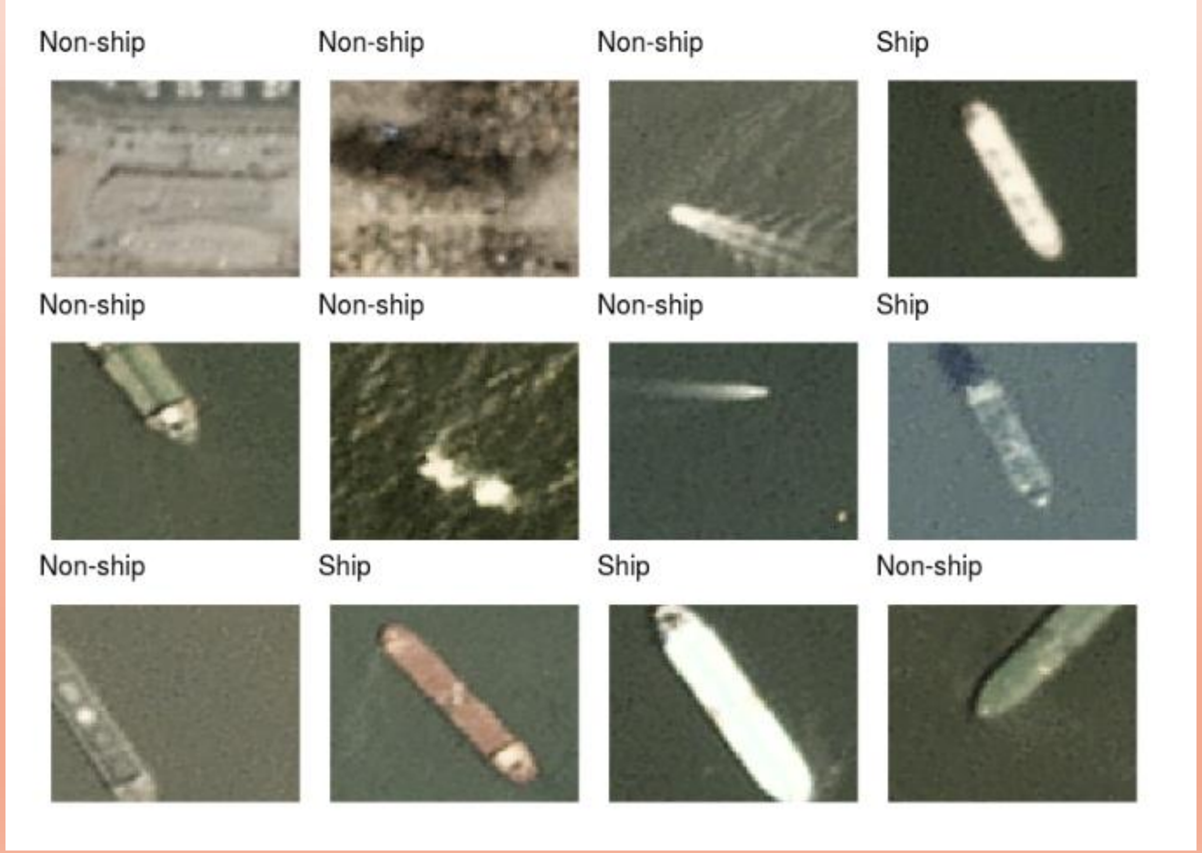


CONVOLUTIONAL NEURAL NETWORKS

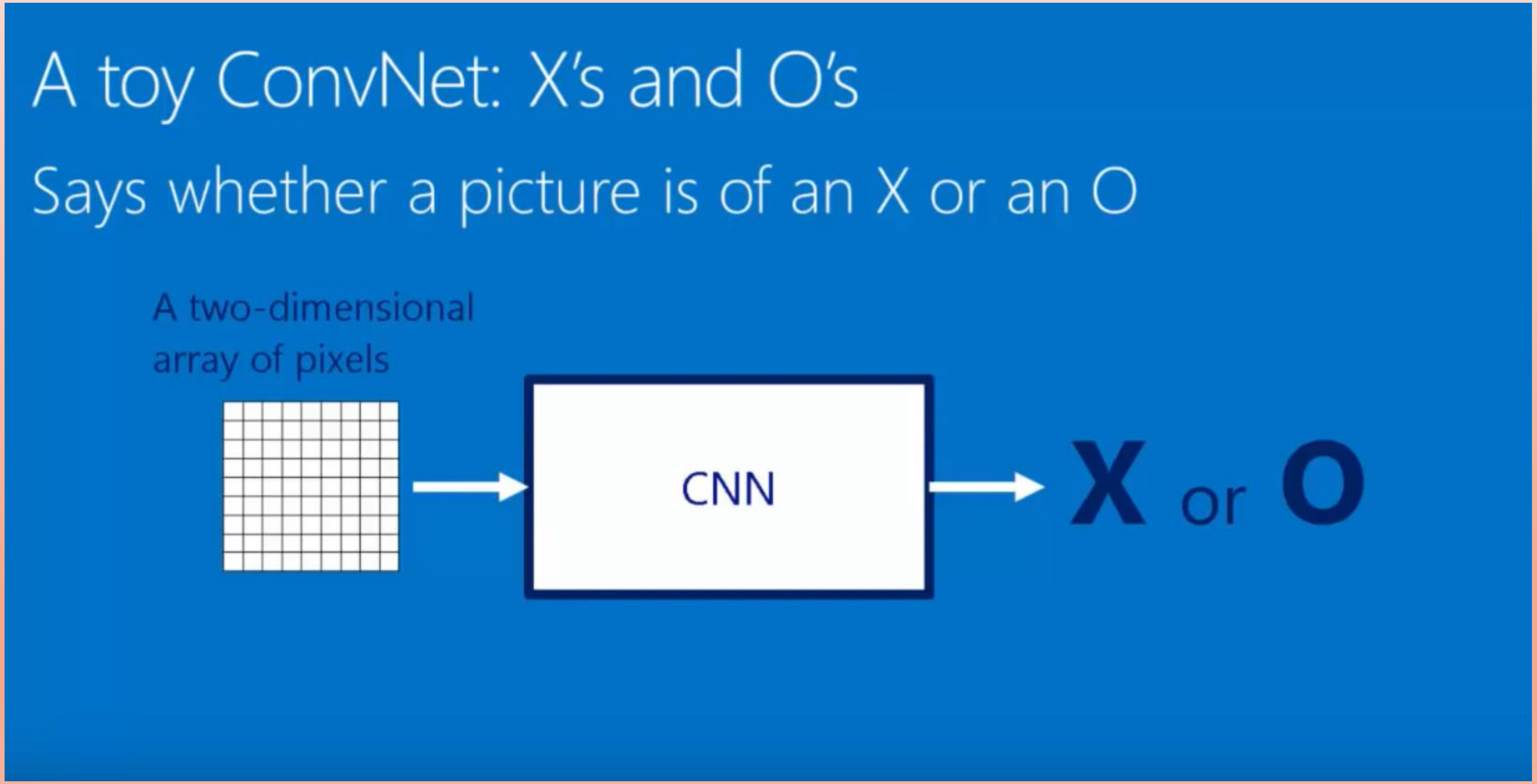
- Keras is a high-level deep learning API that can be run on multiple backends including TensorFlow.

- Keras can be used to build a CNN for object classification such as the one in the picture to the right. These images were part of a data set of close to 3000 images of ships and objects that were not ships.

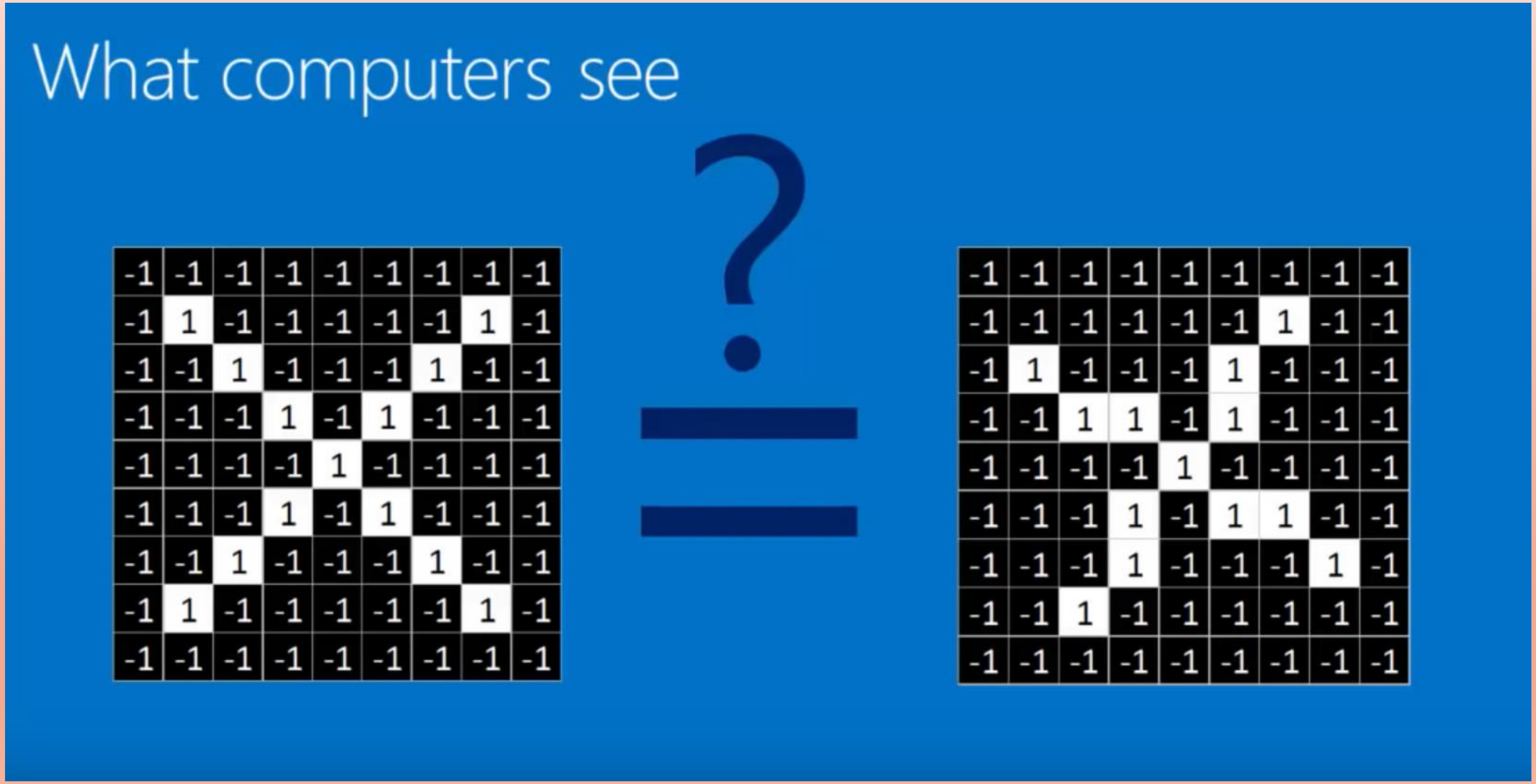
- The architecture is built for a CNN using a series of computational layers.



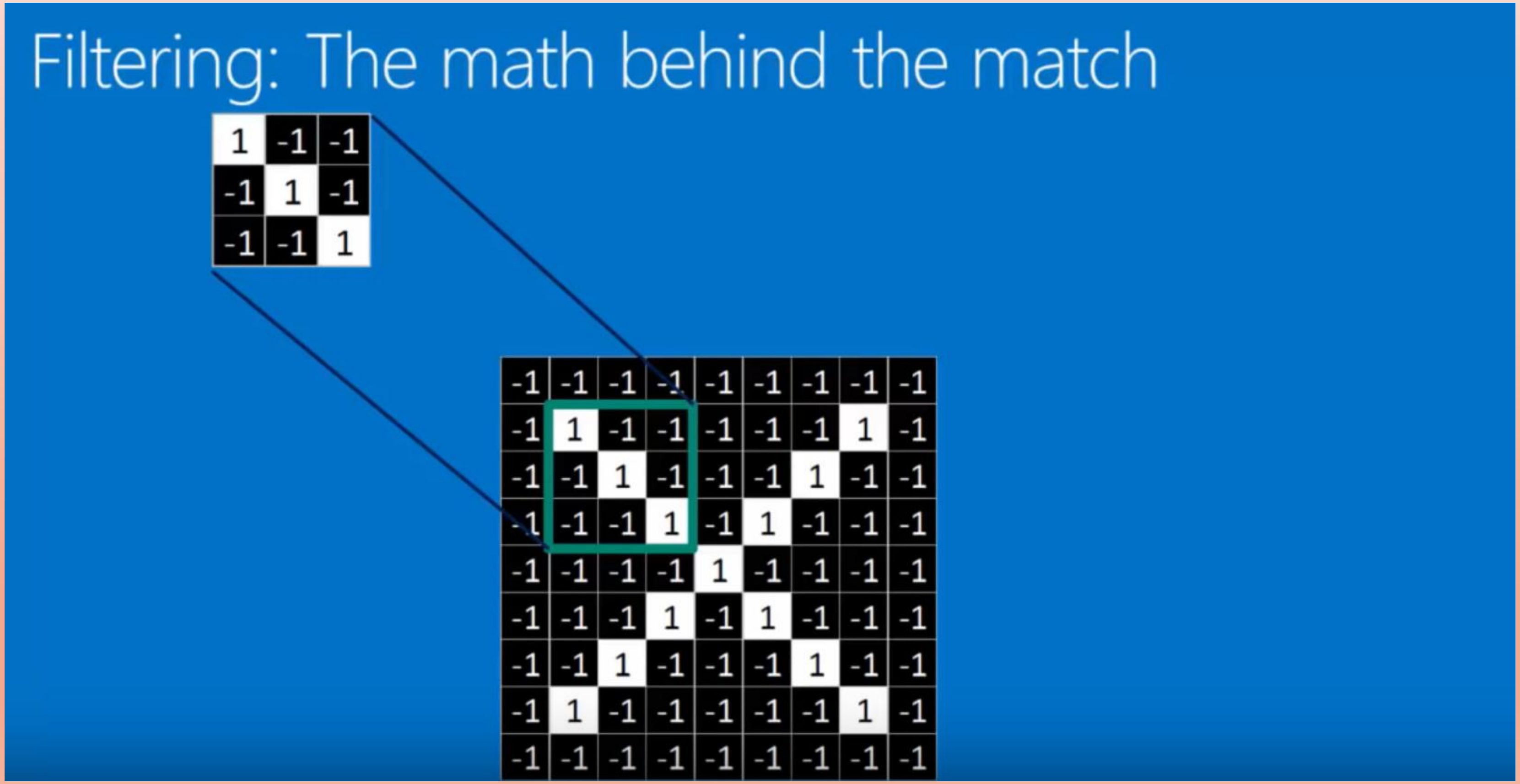
CONVOLUTIONAL NEURAL NETWORKS



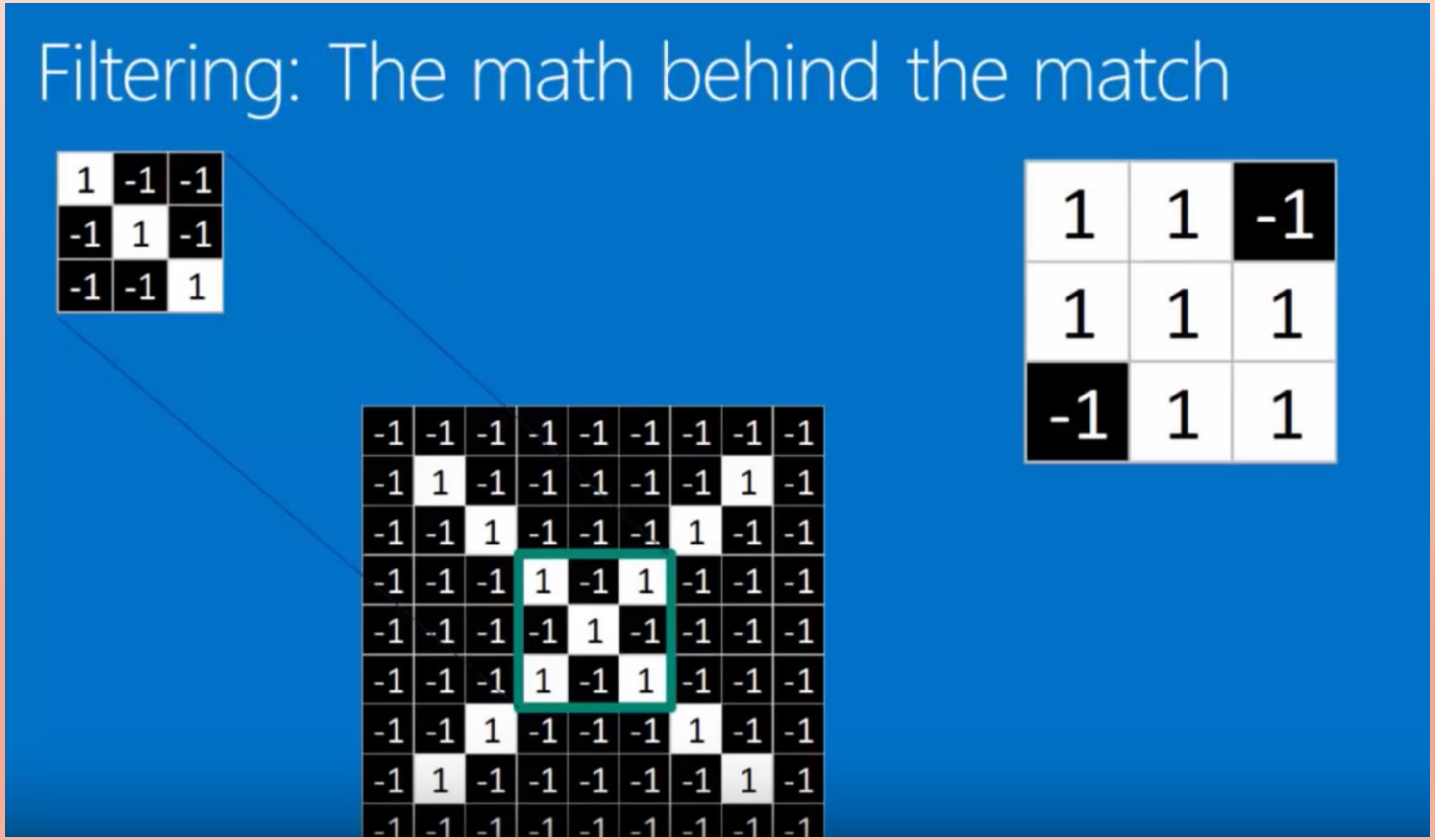
CONVOLUTIONAL NEURAL NETWORKS



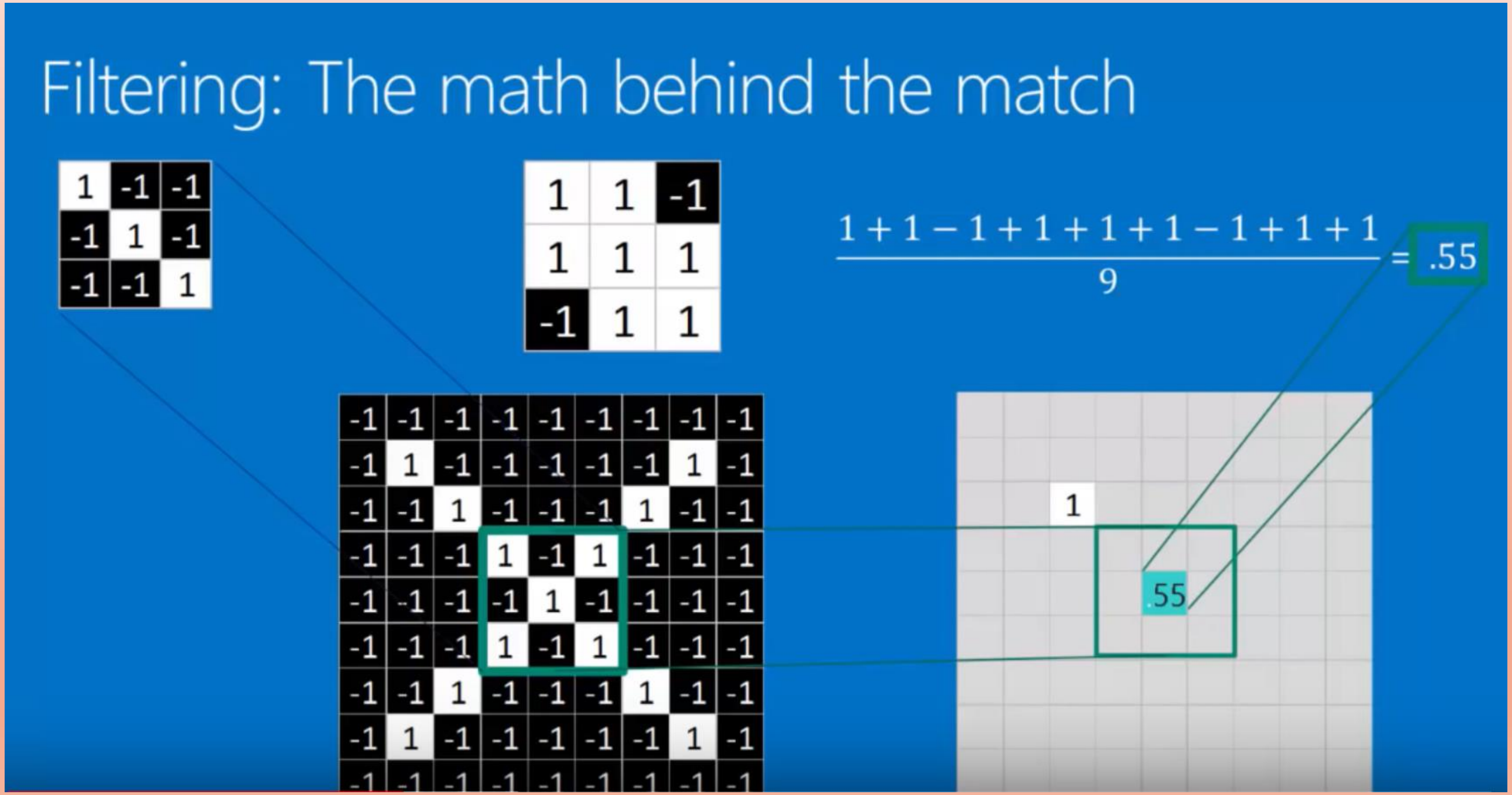
FILTERING AN IMAGE WITH A KERNEL



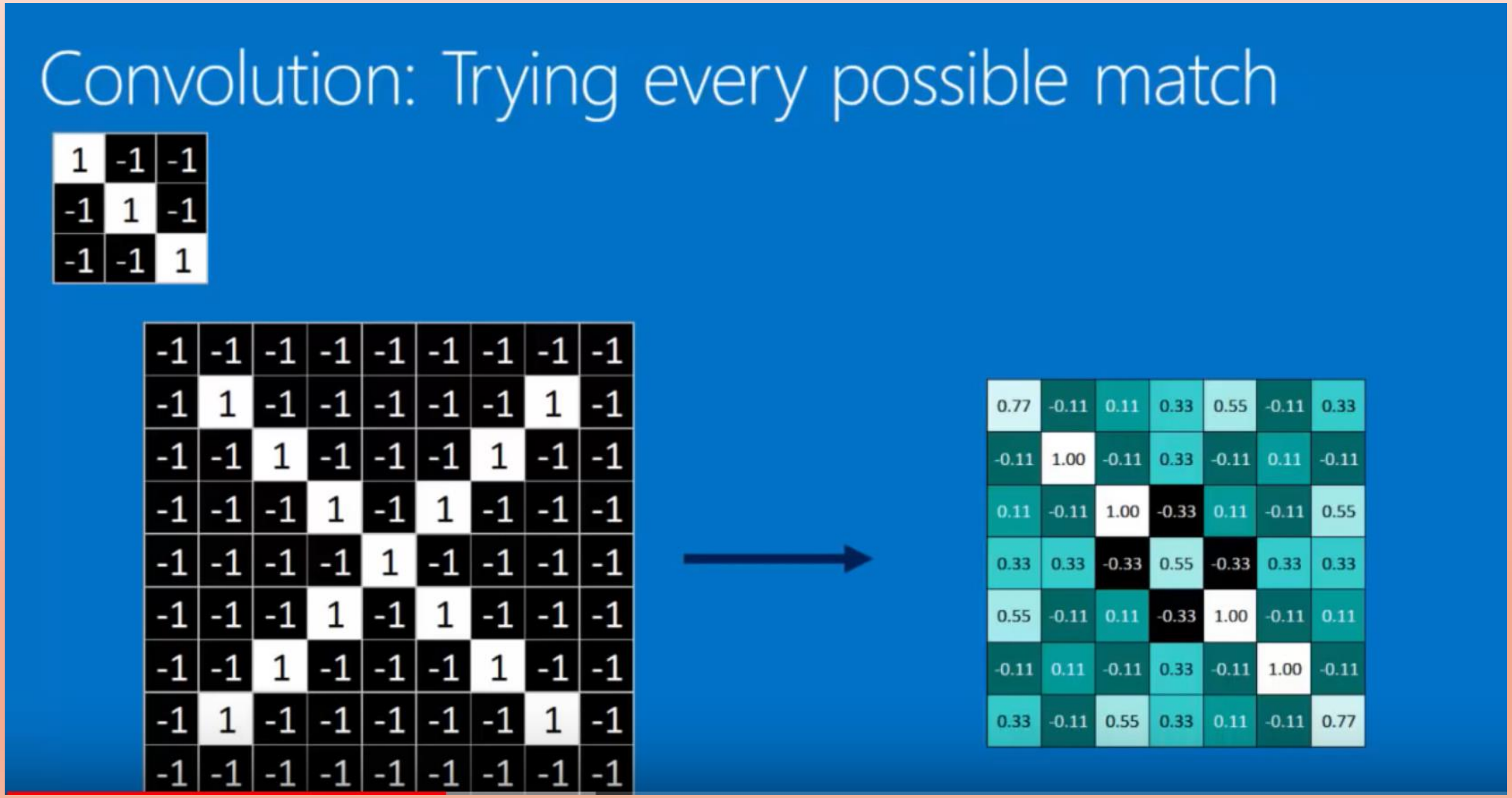
FILTERING AN IMAGE WITH A KERNEL



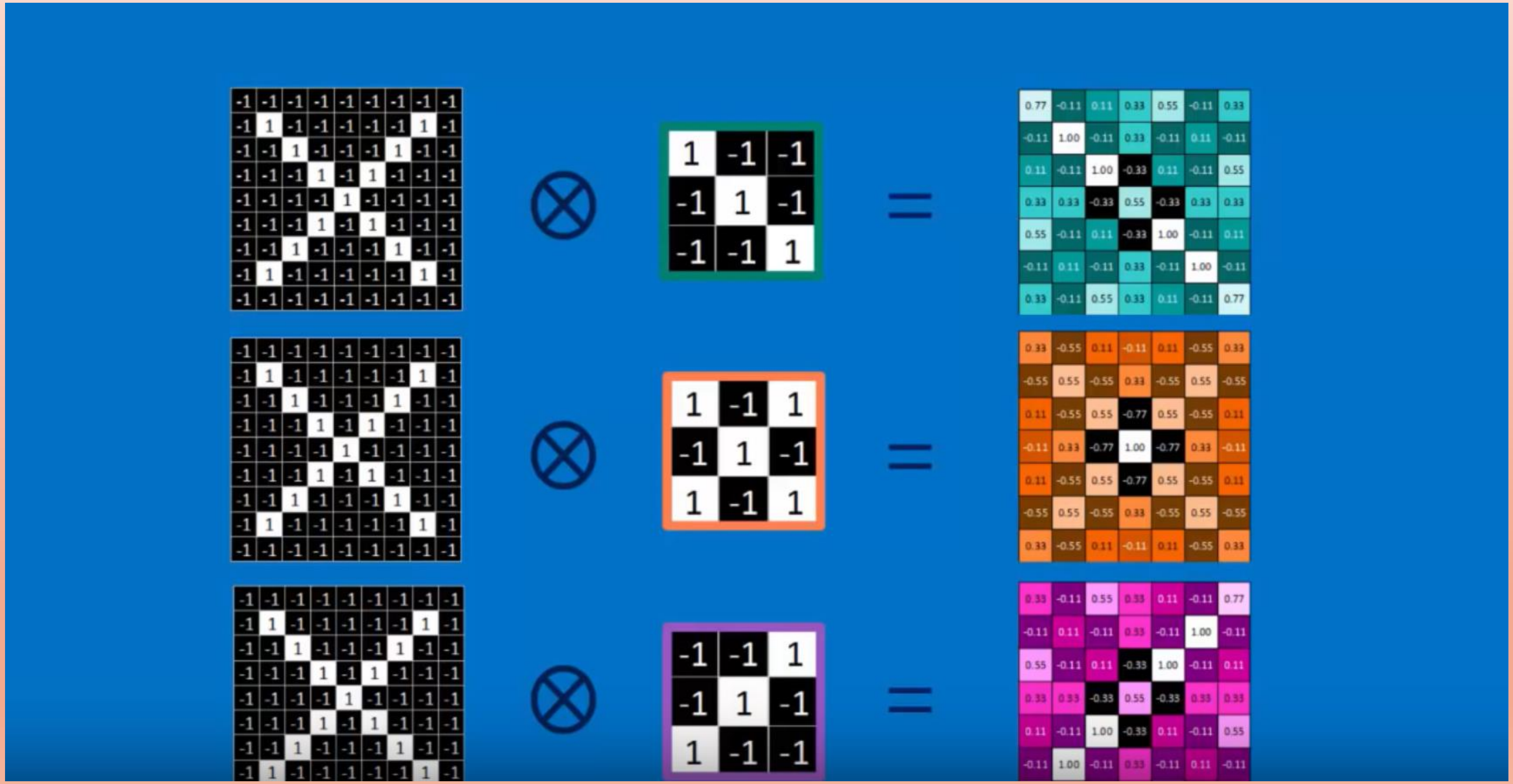
FILTERING AN IMAGE WITH A KERNEL



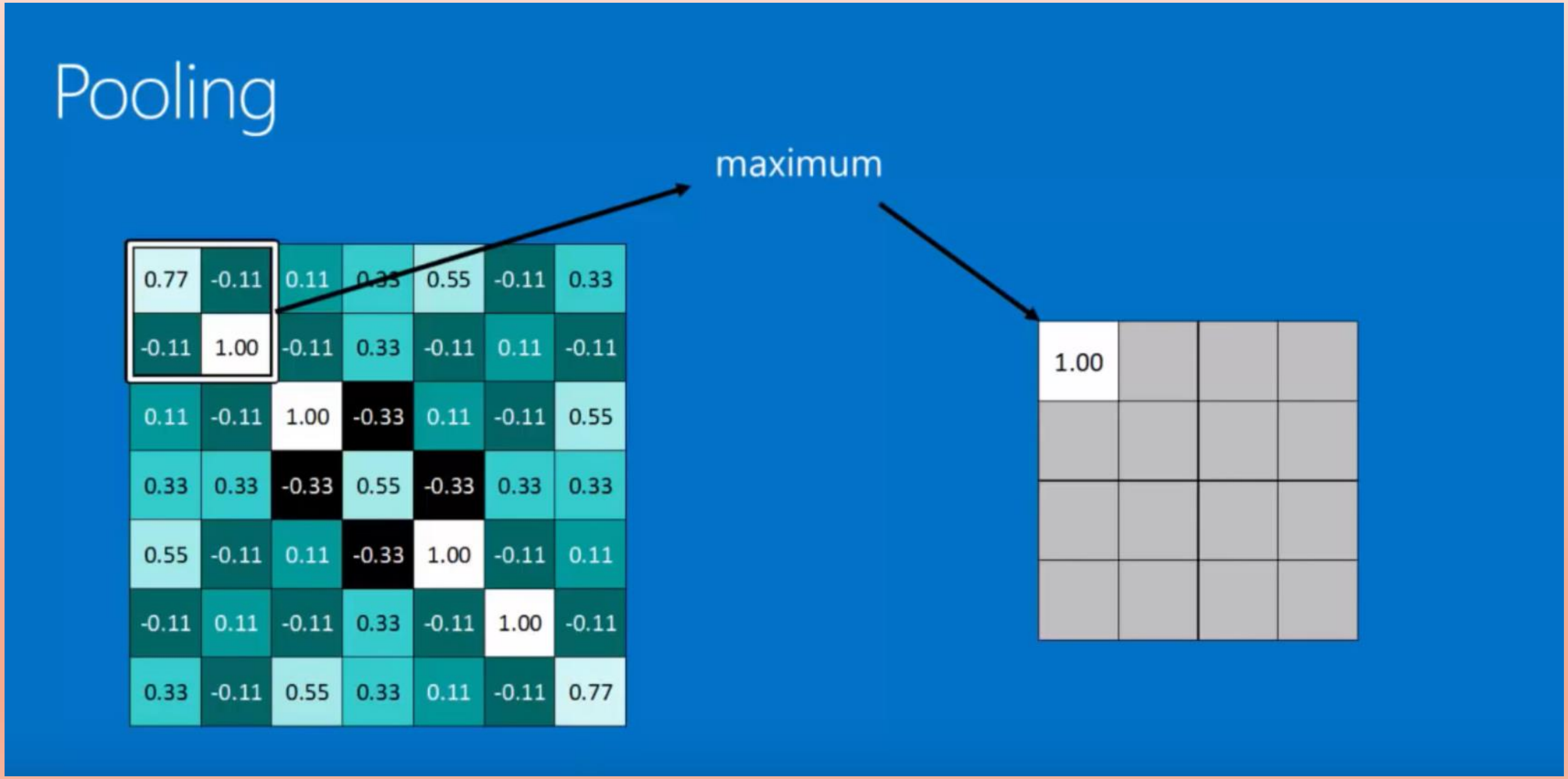
CONVOLUTION



CONVOLUTIONAL LAYERING

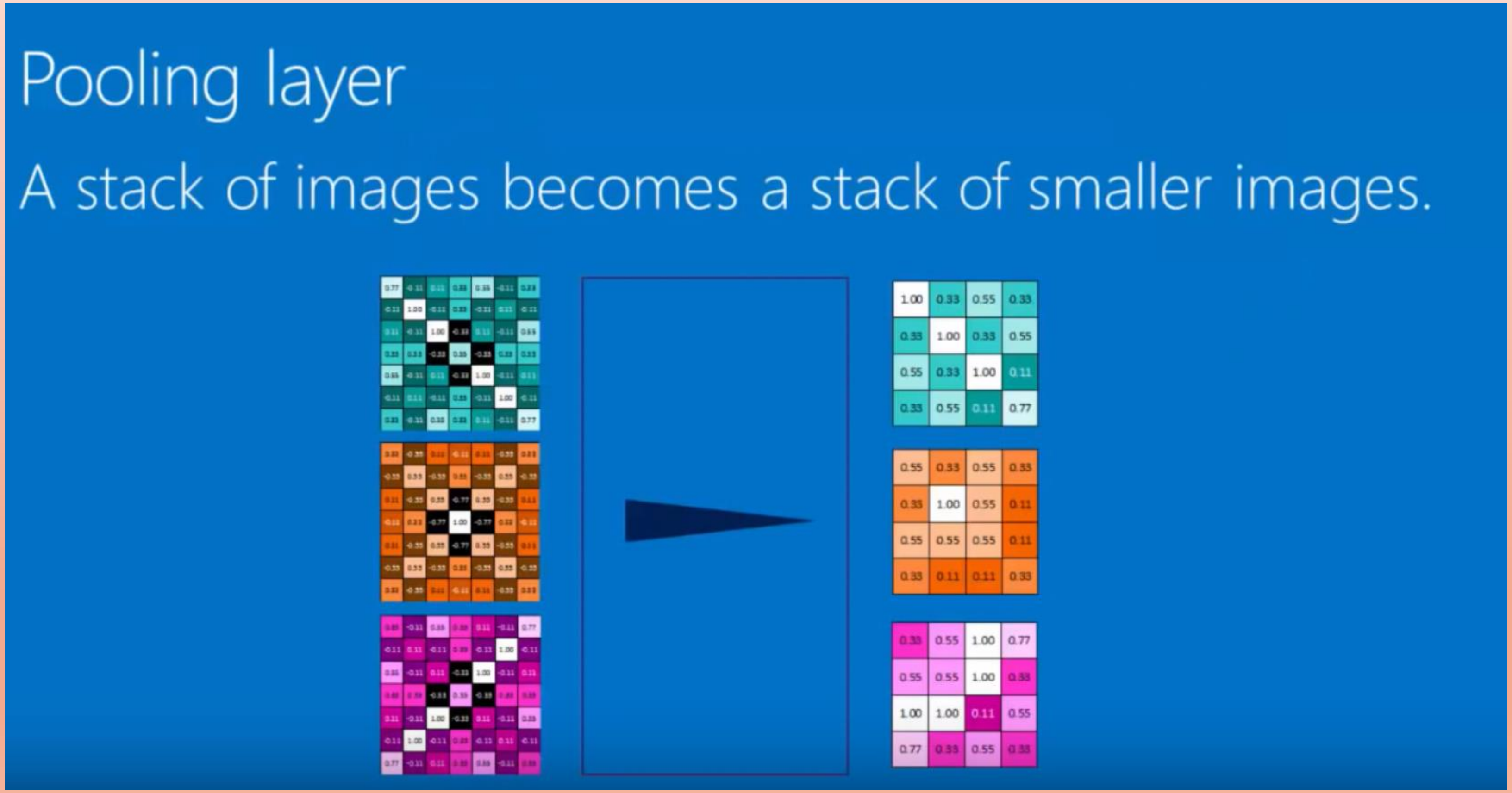


MAX POOLING

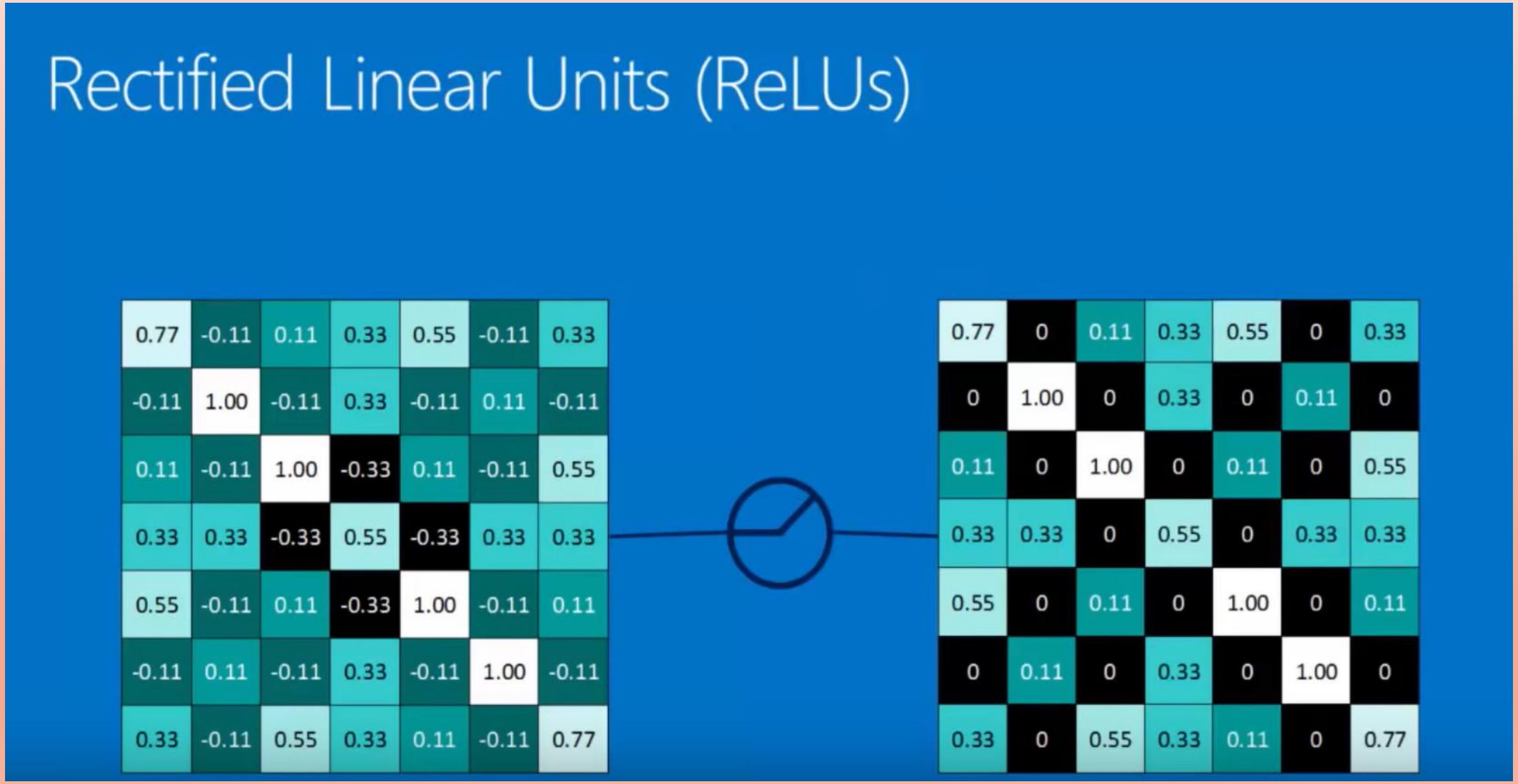


This example illustrates max pooling with a stride of 2.

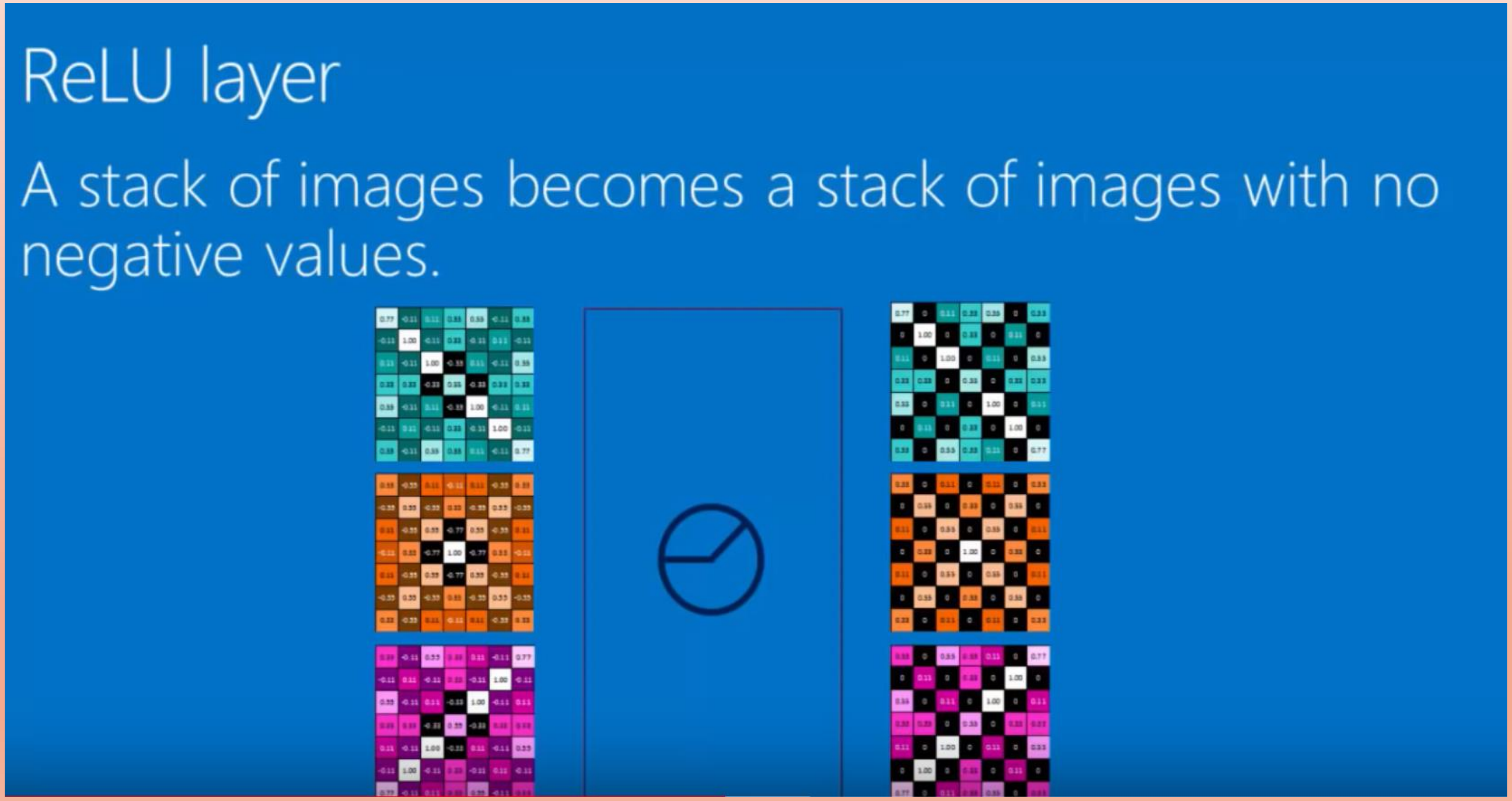
POOLING LAYER



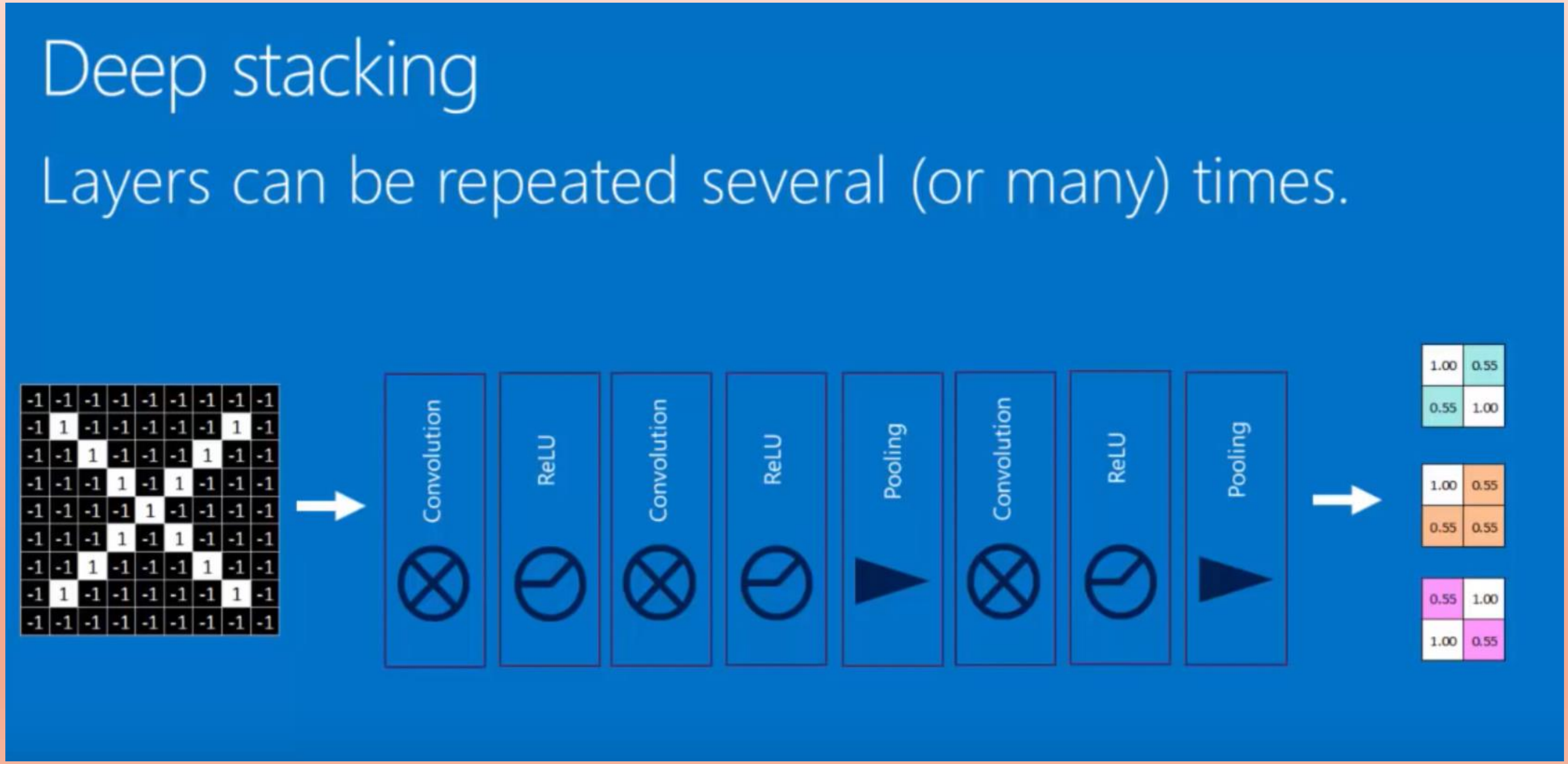
RELUS


Rectified Linear Units (ReLUs)
0.77 -0.11 0.11 0.33 0.55 -0.11 0.33
-0.11 1.00 -0.11 0.33 -0.11 0.11 -0.11
0.11 -0.11 1.00 -0.33 0.11 -0.11 0.55
0.33 0.33 -0.33 0.55 -0.33 0.33 0.33
0.55 -0.11 0.11 -0.33 1.00 -0.11 0.11
-0.11 0.11 -0.11 0.33 -0.11 1.00 -0.11
0.33 -0.11 0.55 0.33 0.11 -0.11 0.77
0.77 0 0.11 0.33 0.55 0 0.33
0 1.00 0 0.33 0 0.11 0
0.11 0 1.00 0 0.11 0 0.55
0.33 0.33 0 0.55 0 0.33 0.33
0.55 0 0.11 0 1.00 0 0.11
0 0.11 0 0.33 0 1.00 0
0.33 0 0.55 0.33 0.11 0 0.77

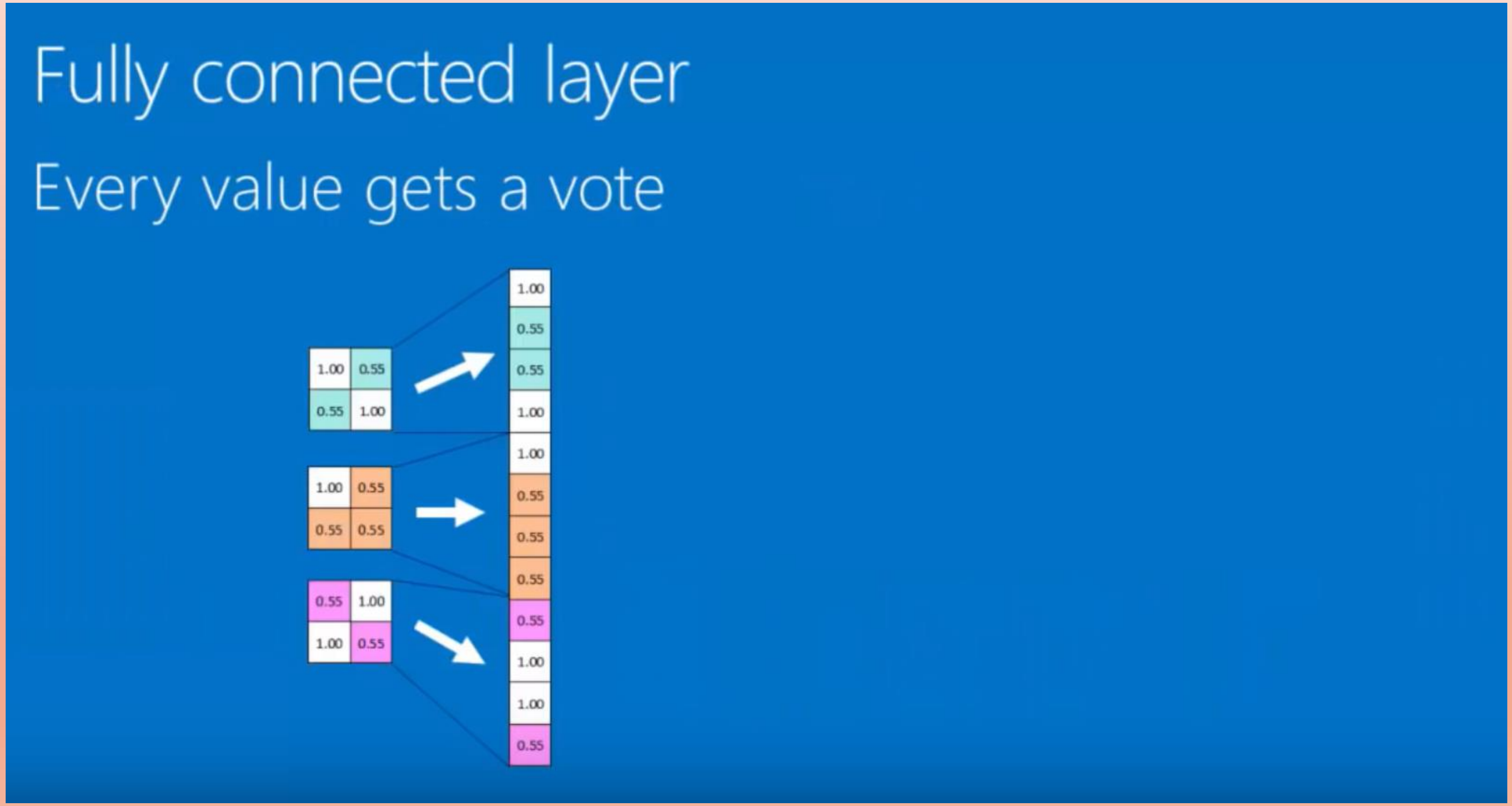
RELU POOLING



DEEP STACKING



FULLY (DENSELY) CONNECTED LAYER

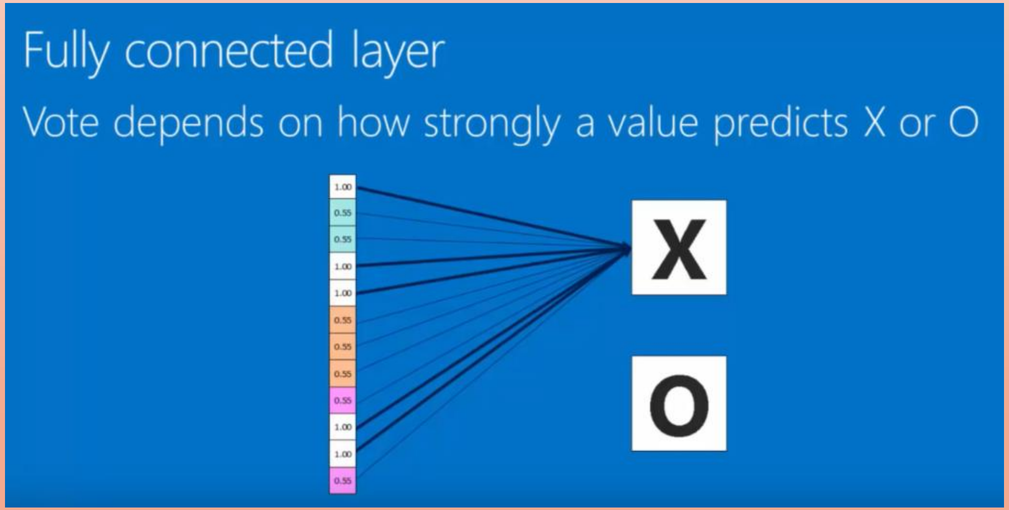


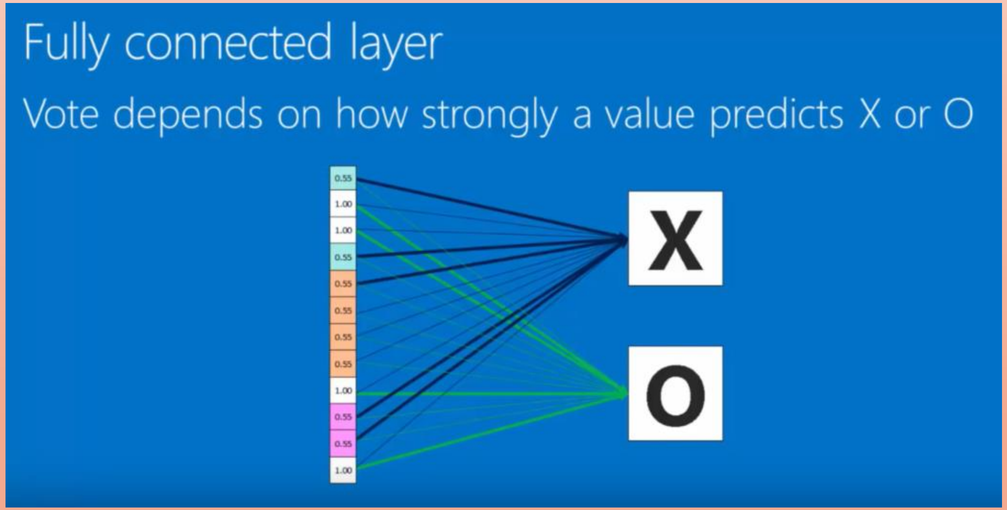
The neurons in the three boxes are tiled out vertically at this step.

FULLY (DENSLEY) CONNECTED LAYER

Predicting an X (input to CNN is an "X")

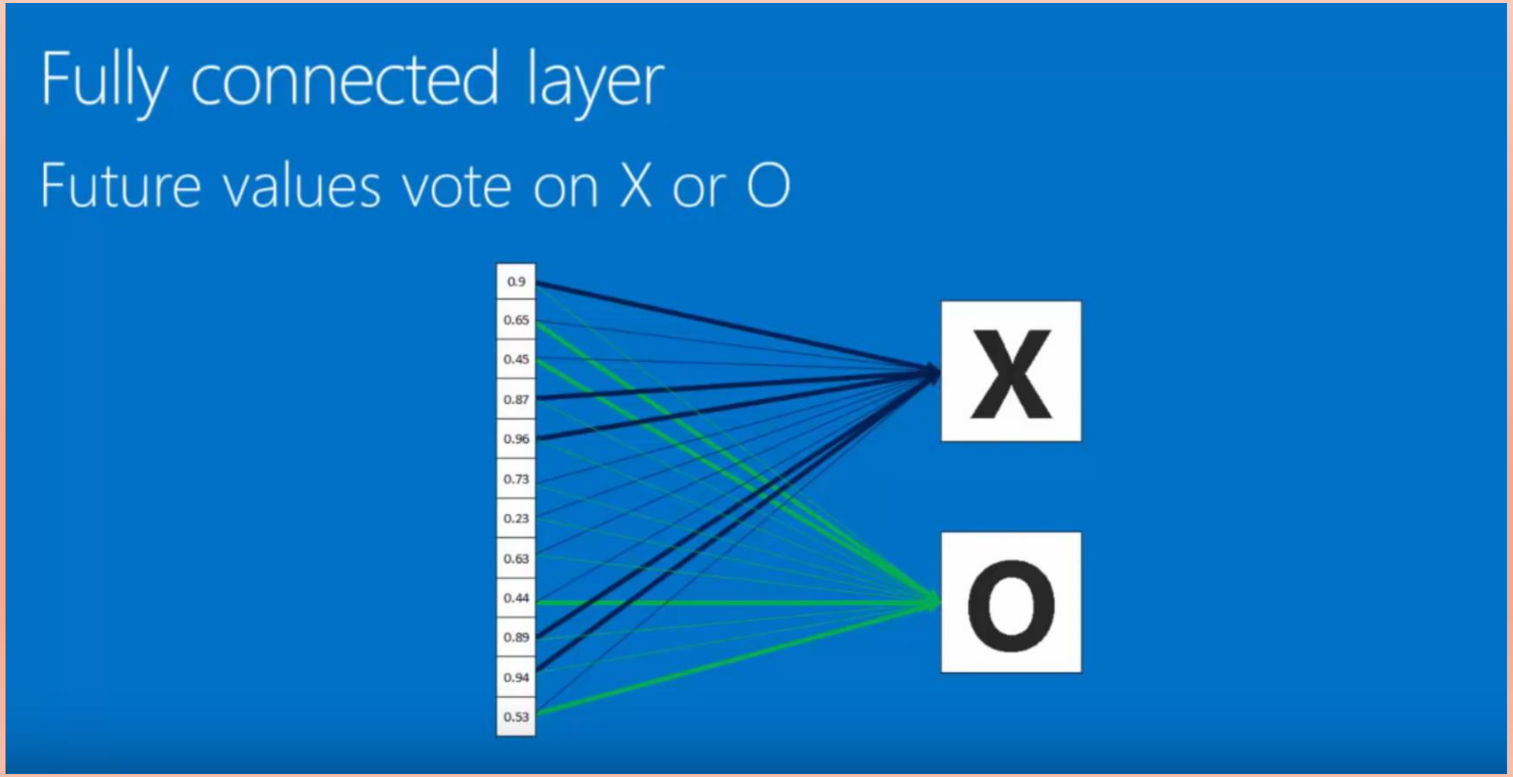
Predicting an O (input to CNN is an "O" instead of an X)



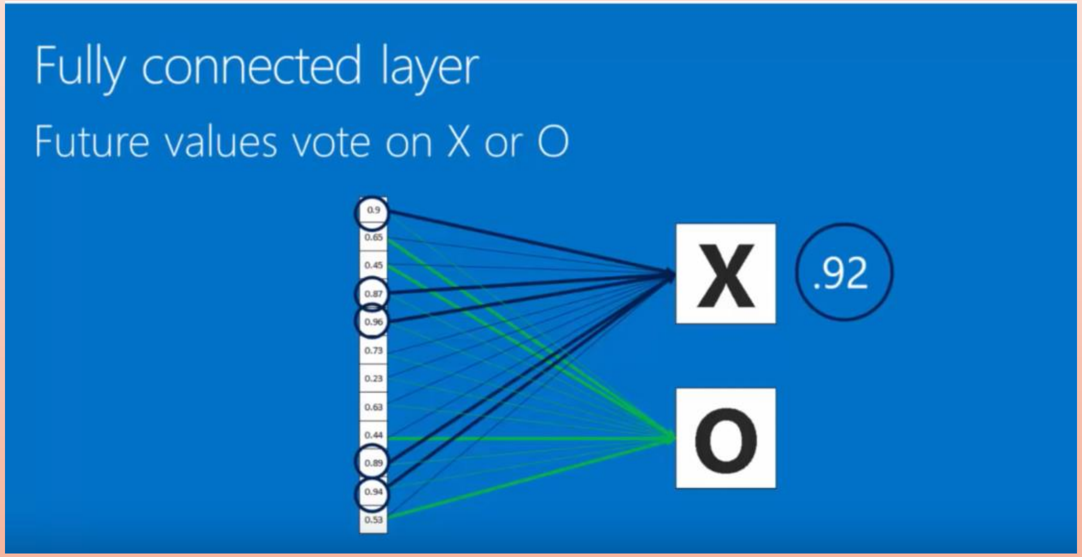


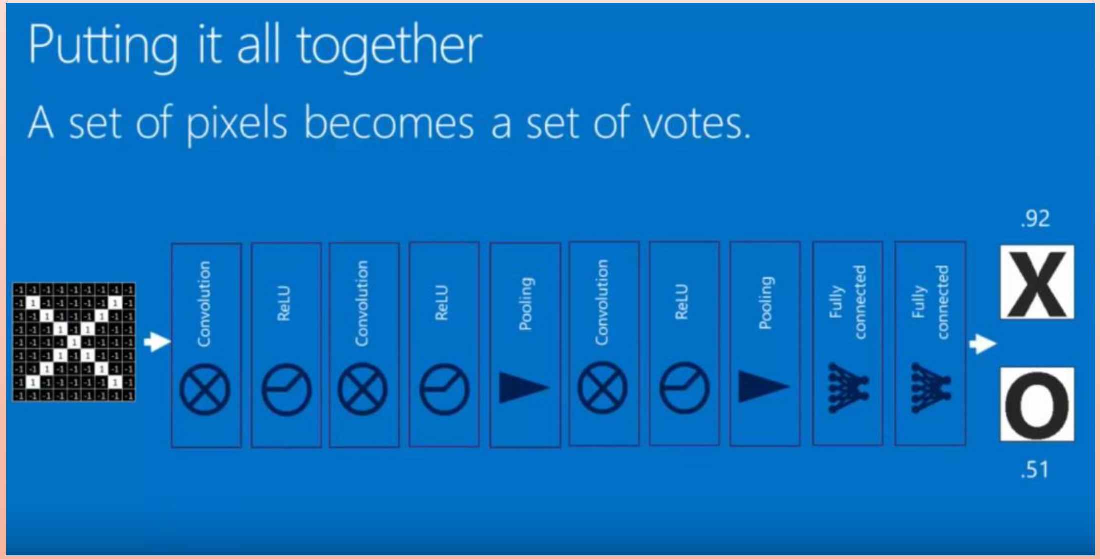
TRAINING (DEEP LEARNING)

Input image is unknown. We wish to run it through the CNN we "trained" to find out whether the image is an X or an O.



TRAINING (DEEP LEARNING)

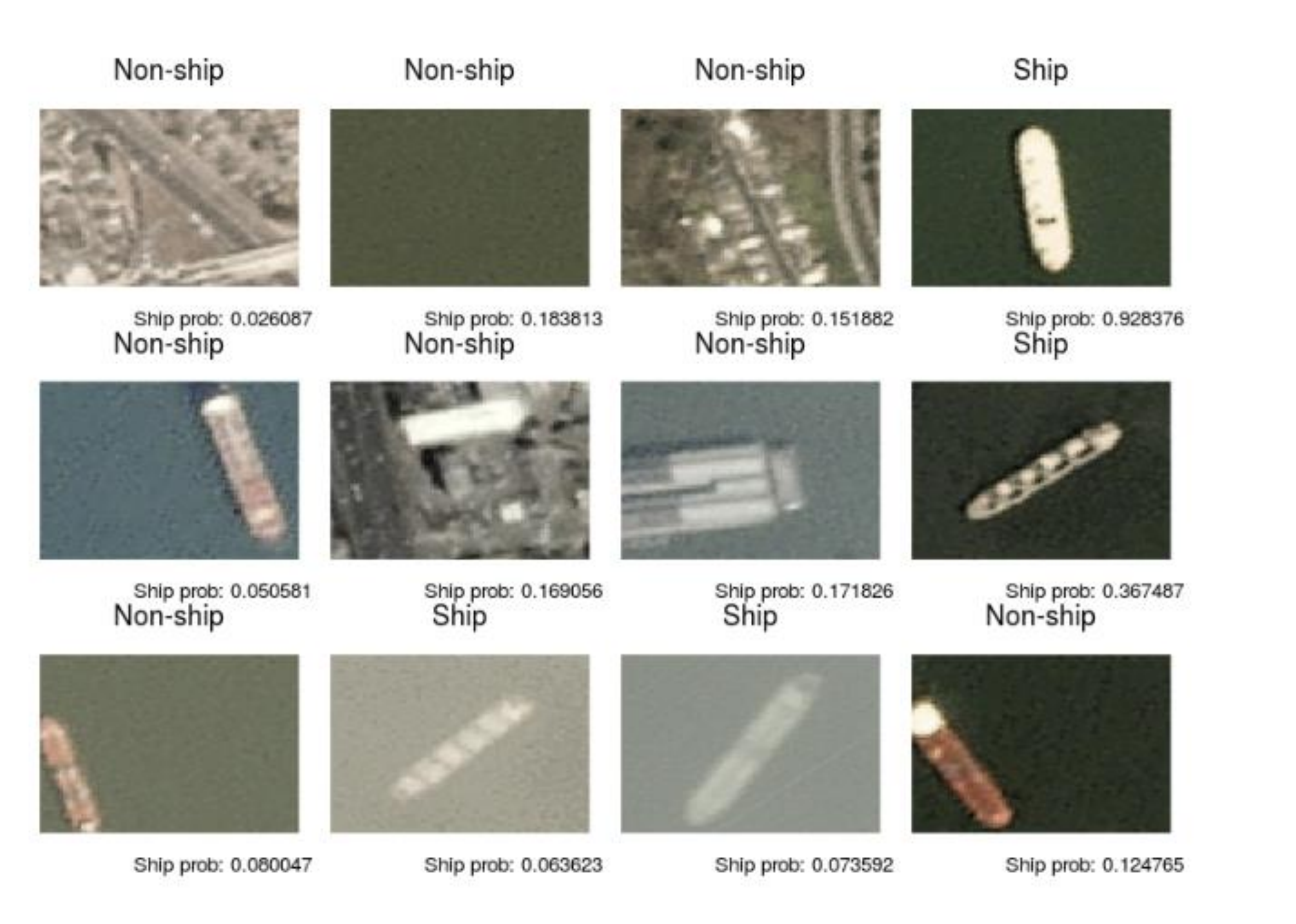




The weighted values ("voting" values) come from a process known as back propogation, which is a mathematical process via which a neural network is "trained" to identify *any* image as an X or an O with a certain degree of confidence.

These output layers can become the input layers for future iterations of the object identification scheme (sometimes these layers are known as *hidden layers.*

DEEP LEARNING



Training a neural network consists of the specification of a loss function, optimizer, and other metrics to evaluate the objects. Back propogation, as mentioned earlier, is a useful training technique.

Once all parameters and configuration settings are in place, it is time to fit the model to the data.

IMPROVEMENTS TO CONVOLUTIONAL NEURAL NETWORKS

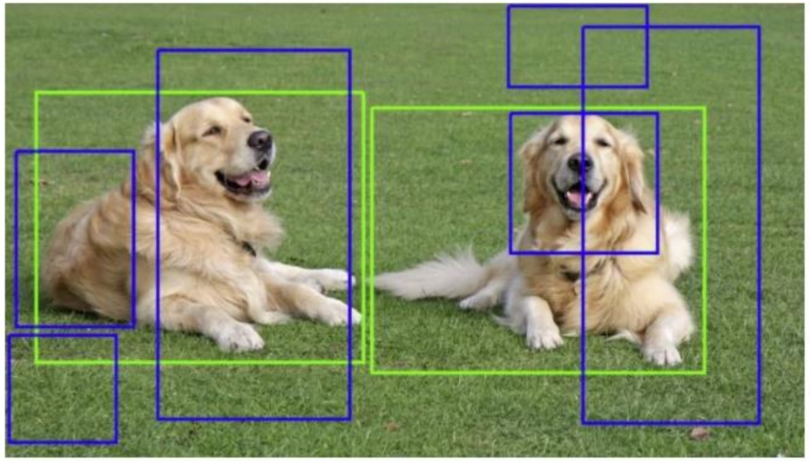
In 2014, Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik wrote a paper called *Rich feature hierarchies for accurate object detection and semantic segmentation*.

In the paper, he described combining region proposals with CNNs, a method he coined R-CNN.

A region proposal algorithm is a sophisticated method that uses an image as its input and as the output it produces a myriad of bounding boxes that correspond to regions in the picture that have a high likelihood of being an image. These regions may overlap, contain only parts of an image, or be completely off-base, but there will be proposals that offer a high probability of an image.

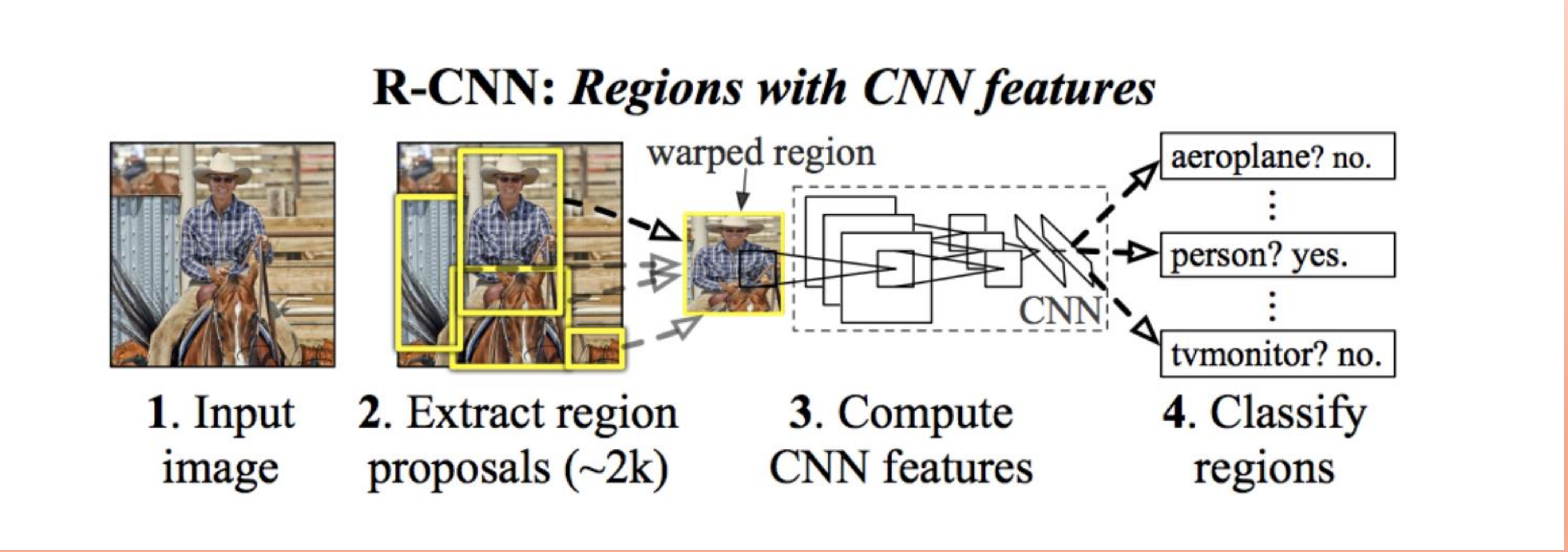
Although many false positive images are generated by this method, the advantage of using it is that there is a very "high recall" of images, meaning that the region proposal algorithm has a high probability of finding a real image.

R-CNN addresses the "object detection" problem.



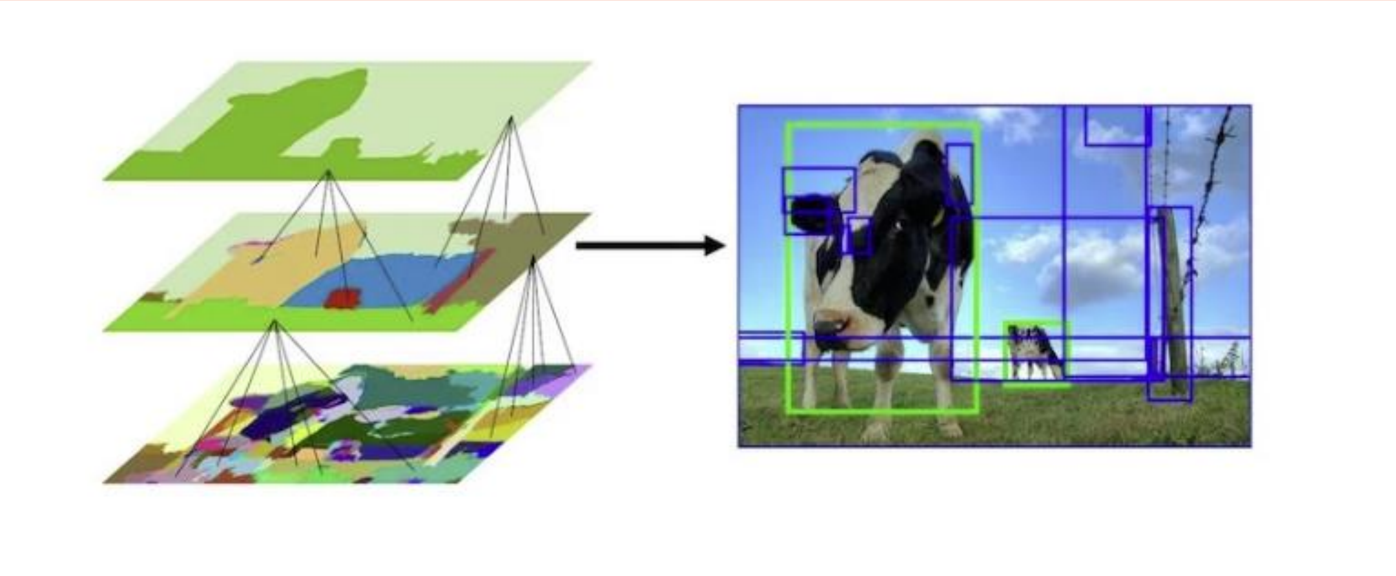
Blue Boxes: False Positives; Green Boxes: True Positives

R-CNN


R-CNN: Regions with CNN features
warped region
aeroplane? no.
person? yes.
tvmonitor? no.
CNN
1. Input image
2. Extract region proposals (~2k)
3. Compute CNN features
4. Classify regions

IMPROVEMENTS TO CONVOLUTIONAL NEURAL NETWORKS

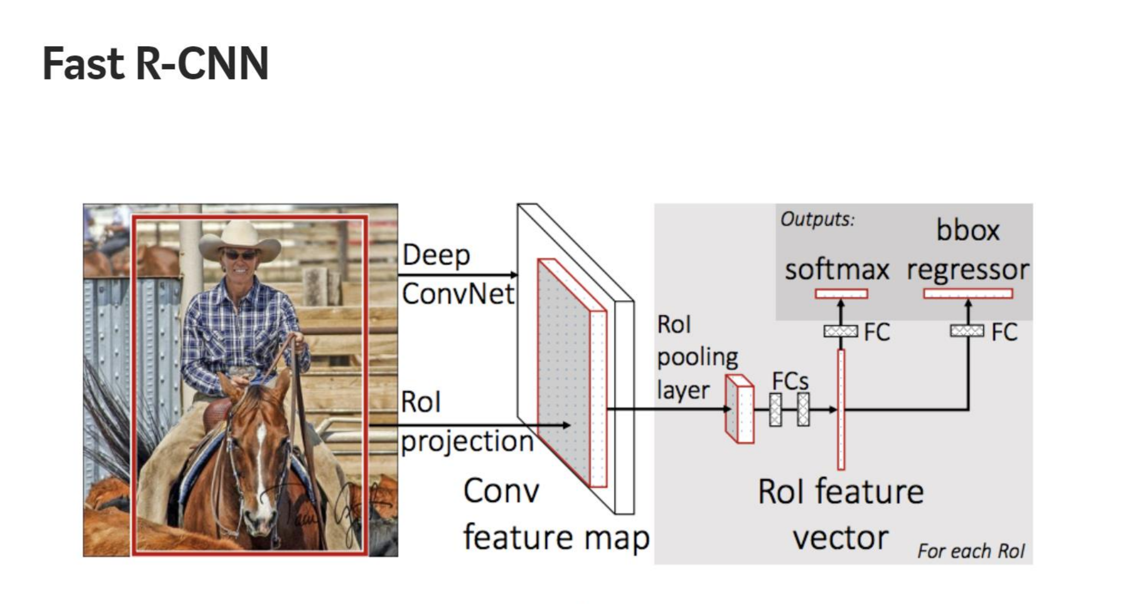
- One of the most common region proposal algorithms is *selective search*, which groups regions based on color, texture, size, and shape compatibility.



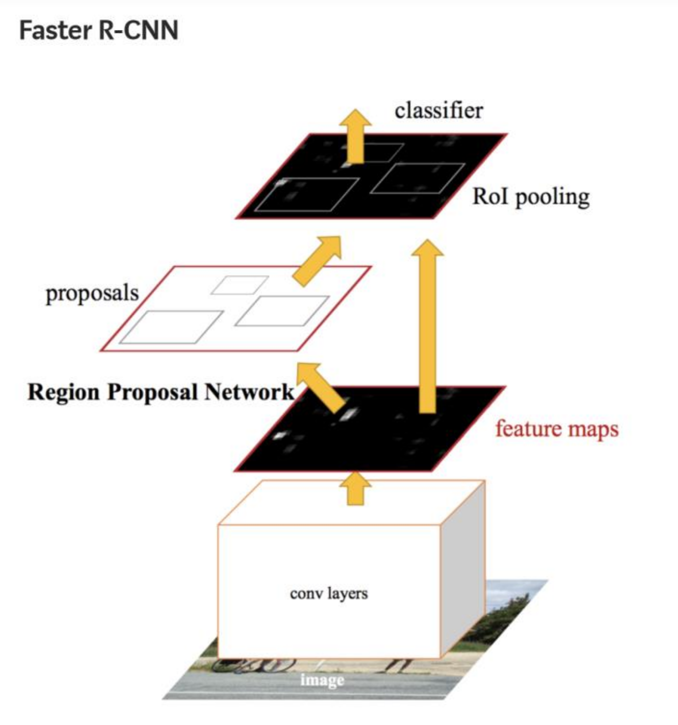
- Begin with an oversegmented image (bottom) and group the adjacent regions based on similarity. Continue hierarchically until the algorithm converges on all of the relevant objects in the image.

IMPROVEMENTS TO CONVOLUTIONAL NEURAL NETWORKS

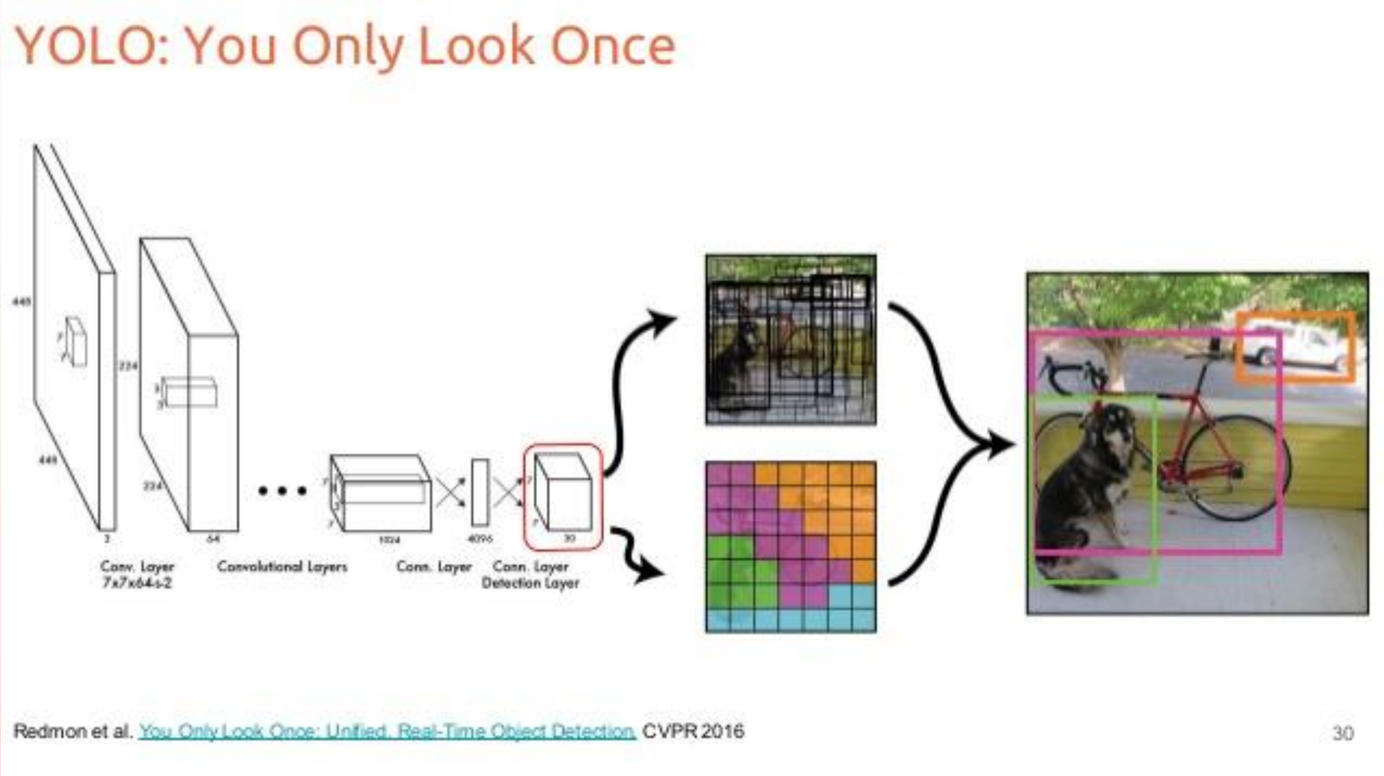
- Improvements to the R-CNN were made recently, with the development of two new algorithms – one being the Fast R-CNN and the other being the Faster R-CNN algorithm.



Fast R-CNN



Faster R-CNN



YOLO

- YOLO is an algorithm that is primarily used for real-time object detection and tracking.
- YOLO is different from R-CNN in that it does not use specific regions to converge on probable images.
- Instead, YOLO divides an image into a rectangular grid consisting of hundreds of individual cells and runs a CNN on the cells to obtain a probability that there is an object in each bounding box.
- Boxes that do not contain objects are eliminated from consideration and then the boxes with the maximum shared area are bounded together via *non-max suppression*.

REFERENCES

- "A Guide to GPU-Accelerated Ship Recognition in Satellite Imagery Using Keras and R (Part I) - Appsilon Data Science: End to End Data Science Solutions." Appsilon Data Science | End to End Data Science Solutions, October 15, 2018. https://appsilon.com/ship-recognition-in-satellite-imagery-part-i/?nabe=5807278576631808:0&utm_referrer=https://www.kdnuggets.com/2018/09/object-detection-image-classification-yolo.html.
- Cessac, Bruno, and Adrian G. Palacios. "Spike Train Statistics from Empirical Facts to Theory: The Case of the Retina." *Modeling in Computational Biology and Biomedicine*, June 2012, 261–302. https://doi.org/10.1007/978-3-642-31208-3_8.
- Chinellato, Eris, Javier Felip, Beata J. Grzyb, Antonio Morales, and Angel P. Del Pobil. "Hierarchical Object Recognition Inspired by Primate Brain Mechanisms." *2011 IEEE Workshop on Computational Intelligence for Visual Intelligence (CIVI)*, 2011. https://doi.org/10.1109/civi.2011.5955017.
- Dicarlo, James J., Davide Zoccolan, and Nicole C. Rust. "How Does the Brain Solve Visual Object Recognition?" *Neuron* 73, no. 3 (2012): 415–34. https://doi.org/10.1016/j.neuron.2012.01.010.
- Gandhi, Rohith. "R-CNN, Fast R-CNN, Faster R-CNN, YOLO - Object Detection Algorithms." Medium. Towards Data Science, July 9, 2018. https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e.
- Grimaldi, Piercesare, Kadharbatcha S. Saleem, and Doris Tsao. "Anatomical Connections of the Functionally Defined 'Face Patches' in the Macaque Monkey." *Neuron* 90, no. 6 (2016): 1325–42. https://doi.org/10.1016/j.neuron.2016.05.009.

REFERENCES (CONTINUED)

- *How Convolutional Neural Networks Work*, 2016. https://www.youtube.com/watch?v=FmpDIaiMIeA&list=LLBWruYZLNIyDNGzMMJ8GM7A&index=4&t=951s.
- Majaj, N. J., H. Hong, E. A. Solomon, and J. J. Dicarlo. "Simple Learned Weighted Sums of Inferior Temporal Neuronal Firing Rates Accurately Predict Human Core Object Recognition Performance." *Journal of Neuroscience* 35, no. 39 (2015): 13402–18. https://doi.org/10.1523/jneurosci.5181-14.2015.
- Mihajlovic, Ilija. "Everything You Ever Wanted To Know About Computer Vision." Medium. Towards Data Science, October 13, 2019. https://towardsdatascience.com/everything-you-ever-wanted-to-know-about-computer-vision-heres-a-look-why-it-s-so-awesome-e8a58dfb641e.
- "Object Detection and Image Classification with YOLO." KDnuggets. Accessed October 22, 2019. https://www.kdnuggets.com/2018/09/object-detection-image-classification-yolo.html.
- OpenStax. Anatomy and Physiology. OpenStax, March 6, 2013. https://opentextbc.ca/anatomyandphysiology/chapter/14-2-central-processing/.
- *Steve Jobs' Vision of the World*, 2011. https://www.youtube.com/watch?v=UvEiSa6_EPA.
- Taylor, Jill Bolte. *My Stroke of Insight: a Brain Scientists Personal Journey*. New York: Penguin Books, 2016.
- *TEDTalks: Jill Bolte Taylor--My Stroke of Insight*, 2008. https://www.ted.com/talks/jill_bolte_taylor_s_powerful_stroke_of_insight.

BOOK & WEBSITE INFORMATION

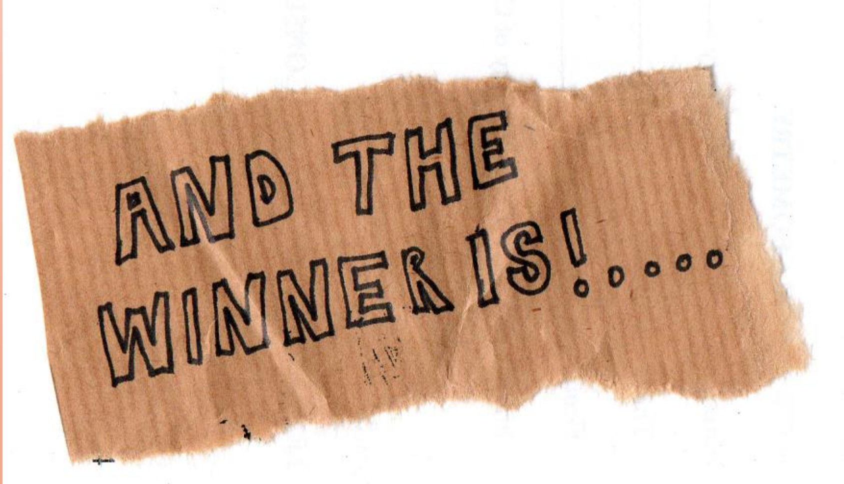
- Visit http://www.thementalarchitect.com/ for information about the book
- The book is available anywhere books are sold (Amazon, Kindle, Barnes & Noble, Smashwords, etc.)
- Published in August 2019 and released publically on October 17.
- Can also purchase autographed copies directly from me (limited quantity available) for $24.95 plus tax.
- We are also selling merchandise on the website (T-shirts, hoodies, mugs, backpacks, drawstring bags) with 5% of net proceeds donated to the Brain & Behavior Research Foundation.
- There is also a blog on the website.

WEBSITE CONTEST

- We will be giving away one Mental Architecture package each month for the next three months.
    - Signed copy of *Mental Architecture*
    - *Mental Architecture* T-Shirt
    - $25 gift card to Barnes & Noble
- You can enter the contest online on the MA website or by going here: https://www.thementalarchitect.com/contest.html

AUTOGRAPHED BOOK DRAWING

- Let's find out who our winner is…



Q&A AND SURVEYS

- If we have time, we will take some questions.
- Please fill out the surveys and return to me.
- Please make certain you signed in to the event on the official FLEX day attendance sheet.

- Thank you for your interest & support for *Mental Architecture*!!!!
- Stay tuned for future talks!