

Study Portfolio

COMPUTER VISION

Varit Kobutra

Professor Patricia McManus

Houston Community College

3 August 2024

Portfolio GitHub Link: <https://github.com/henrykobutra/ital-1378-portfolio>



Computer vision is not
about teaching computers
to see, it's about teaching
computers to understand
the world around us.

- Jensen Huang, CEO of NVIDIA



WHAT WE LEARNED

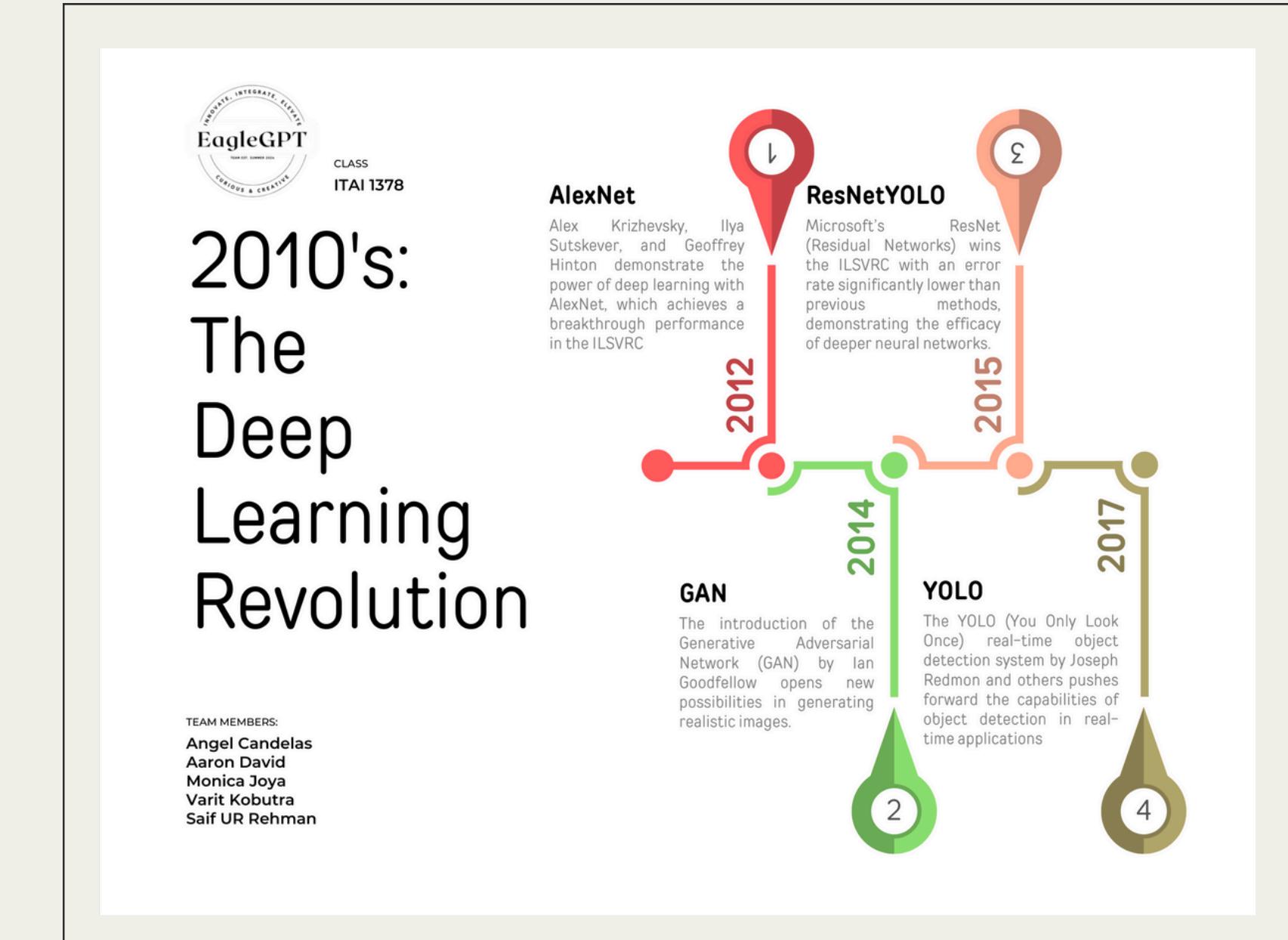
- Historical Timeline of Computer Vision
- Jupyter Notebooks and GitHub
- Image Processing Fundamentals
- Machine Learning in Computer Vision
- Neural Network Building Blocks
- Convolutional Neural Networks
- Advanced CNN Architectures
- Object Detection and Recognition
- The Future of Generative AI and AI Agents



HISTORICAL TIMELINE OF COMPUTER VISION

The success of these methods on benchmark competitions has inspired a growing number of researchers and companies to adopt and further develop deep learning methods for computer vision. This has arguably led to a breakthrough in the effectiveness of computer vision in several areas" (Voulodimos et al., 2018, p. 2)

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). *Deep learning for computer vision: A brief review*. Computational Intelligence and Neuroscience, 2018, 1-13. <https://doi.org/10.1155/2018/7068349>



JUPYTER NOTEBOOKS AND GITHUB

Jupyter Notebooks and GitHub form a powerful tandem in computer vision workflows, with Notebooks providing an interactive environment for code execution, data visualization, and documentation, while GitHub offers version control, collaboration features, and a platform for sharing and discovering cutting-edge computer vision projects and datasets.

Learn more:

- <https://jupyter.org/>
- <https://github.com>

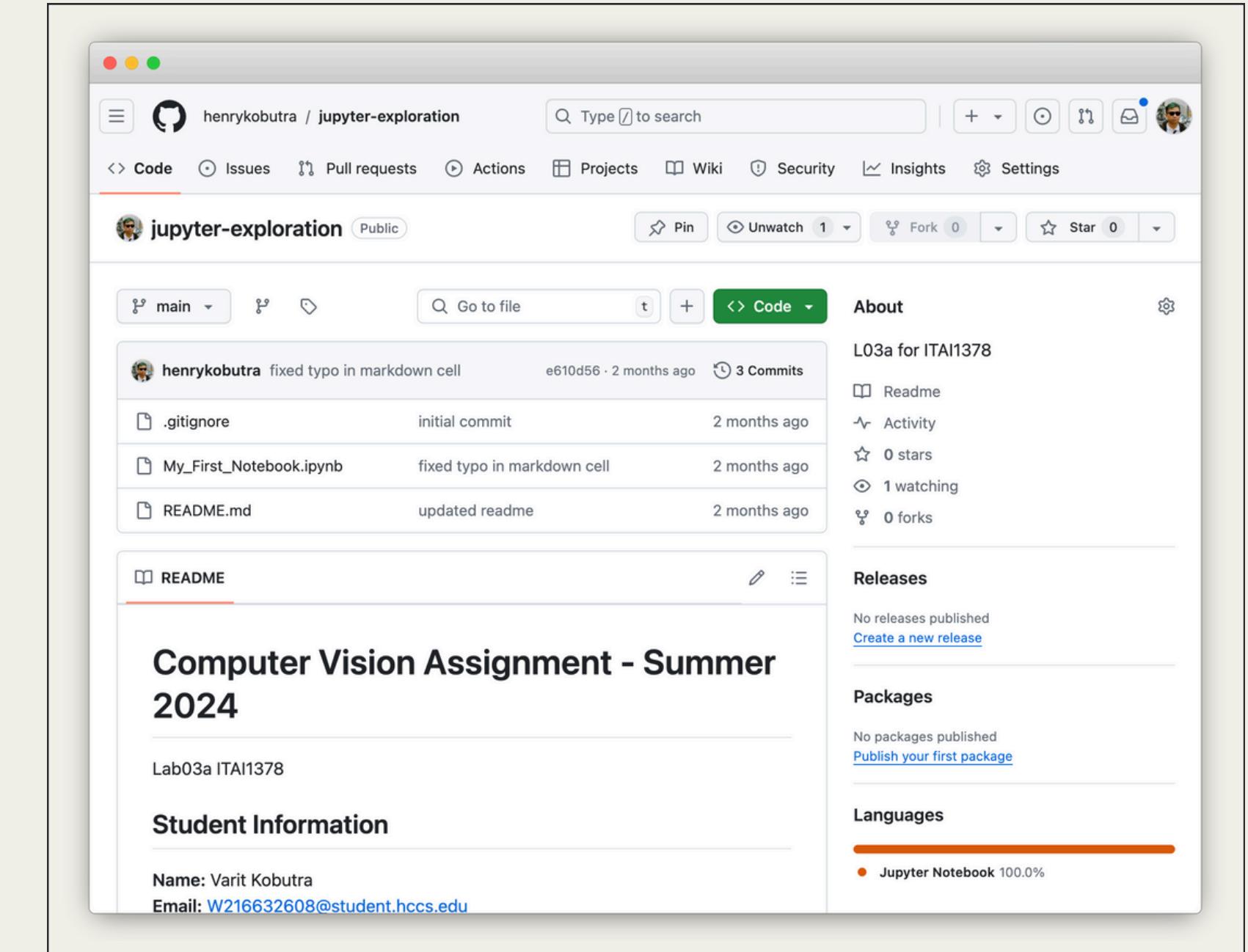


IMAGE PROCESSING FUNDAMENTALS

As Gonzalez and Woods (2018) note, "Image processing is a core technology in computer vision, forming the foundation upon which higher-level vision tasks are built" (p. 1). This underscores the critical role of image processing techniques in enabling sophisticated computer vision applications.

Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing* (4th ed.). Pearson.

How Pixels Form Digital Images and Color Models Representation

Prepared by Angel Candela, Aaron David, Monica Joya, Varit Kobutra, Salf UR Rehman
Prepared for Prof. Patricia McManus
Class of ITAI 1378 Computer Vision: 12461 Summer 2024

EagleGPT x HCC

● Introduction to Pixels

A pixel, short for "picture element," is the smallest unit of a digital image or display. Each pixel represents a specific color and brightness value, collectively forming the image.



Resolution of an image is determined by the number of pixels it contains (e.g., 1920x1080 pixels).

● RGB Color Model

RGB stands for Red, Green, and Blue. It is an additive color model where colors are created by combining these three primary colors of light at various intensities.

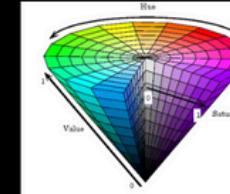


Explanation: Each color in the RGB model is represented by a triplet (R, G, B), where each component can range from 0 to 255. When all three colors are at their maximum intensity (255, 255, 255), the result is white. When all are at zero intensity (0, 0, 0), the result is black.

Application: Used in digital screens, cameras, and image editing software.

● HSV Color Model

HSV stands for Hue, Saturation, and Value. It is a cylindrical color model that remaps the RGB primary colors into dimensions that are easier for humans to understand.



Hue: Represents the type of color and is measured in degrees (0° for red, 120° for green, 240° for blue).
Saturation: Represents the intensity or purity of the color (0% is grayscale, 100% is the purest color).
Value: Represents the brightness of the color (0% is black, 100% is the brightest color).

Application: Used in image analysis, computer vision, and color selection tools.

● Comparison of RGB and HSV

RGB: Best for devices that emit light (screens, cameras).
HSV: Useful in image processing in Computer Vision.



RGB Image HSV Image

● Practical Examples

RGB: Primary color model for TV, Monitors and Smartphones.
HSV: Artists and designers often use this model because it allows them to manipulate colors more easily.
YCbCr: Widely used in computer vision for tasks like object detection and image segmentation.
HSL: When creating visualizations that represent data with color, this model is better than RGB because it more perceptually uniform color spaces, meaning that variations in data are accurately represented by variations in color.

● Conclusion

Understanding pixels and color models is crucial in computer vision as it enables accurate image analysis, object detection, and segmentation by effectively manipulating and interpreting color information.

Other Notable Color Models Related to Computer Vision

YCbCr: A color model that represents colors in terms of hue, lightness, and saturation. It is similar to HSV but uses lightness instead of value. It is widely used for tasks that require color-based image segmentation and compression, as it allows for more efficient compression by separating the luminance channel from the chrominance channels without significantly affecting perceived image quality, making it ideal for video compression and broadcast transmission.

References

- Hand Map: "HSV vs. RGB." Hand Map, 26 Oct. 2016, handmap.github.io/hsv-vs-rgb/.
- MathWorks, "Understanding Color Spaces and Color Space Conversion." MATLAB 4 Product Documentation, MathWorks, www.mathworks.com/help/vision/understanding-color-spaces.html. Accessed 24 June 2024.
- Northwestern University, "Converting Colorspaces." MATLAB Image Processing Toolbox Documentation, MathWorks, www.mathworks.com/help/images/converting-colorspaces.html. Accessed 24 June 2024.
- "Color Modes and Color Spaces." Programming Design System, www.apple.com/appstore/ios/ios15/color-modes-and-color-spaces/index.html. Accessed 24 June 2024.

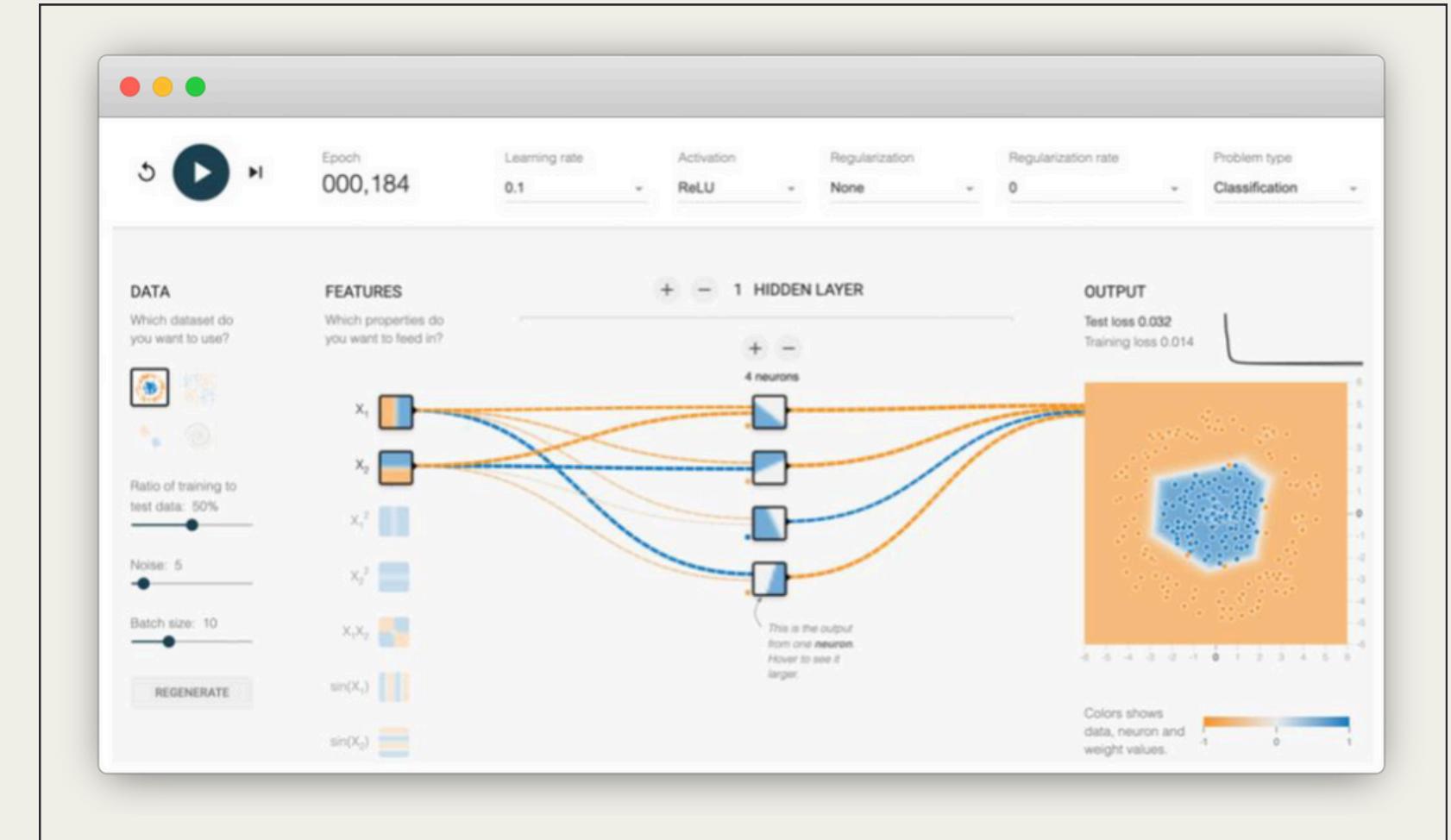
MACHINE LEARNING IN COMPUTER VISION

Machine learning forms the backbone of modern computer vision systems, enabling tasks like image classification, object detection, and scene understanding through algorithms that learn from data. In our course, we explored this synergy hands-on by implementing an SVM classifier for the CIFAR-10 dataset, achieving 31.2% accuracy across 10 diverse object categories and gaining practical insights into challenges like small image sizes and class diversity.



NEURAL NETWORK BUILDING BLOCKS

TensorFlow Playground provided an interactive platform to visualize and experiment with neural network architectures, allowing us to gain intuitive understanding of key concepts like layers, neurons, and activation functions – a approach that aligns with Smilkov et al.'s (2017) finding that "interactive visualizations can help build an intuitive understanding of machine learning concepts" (p. 1).

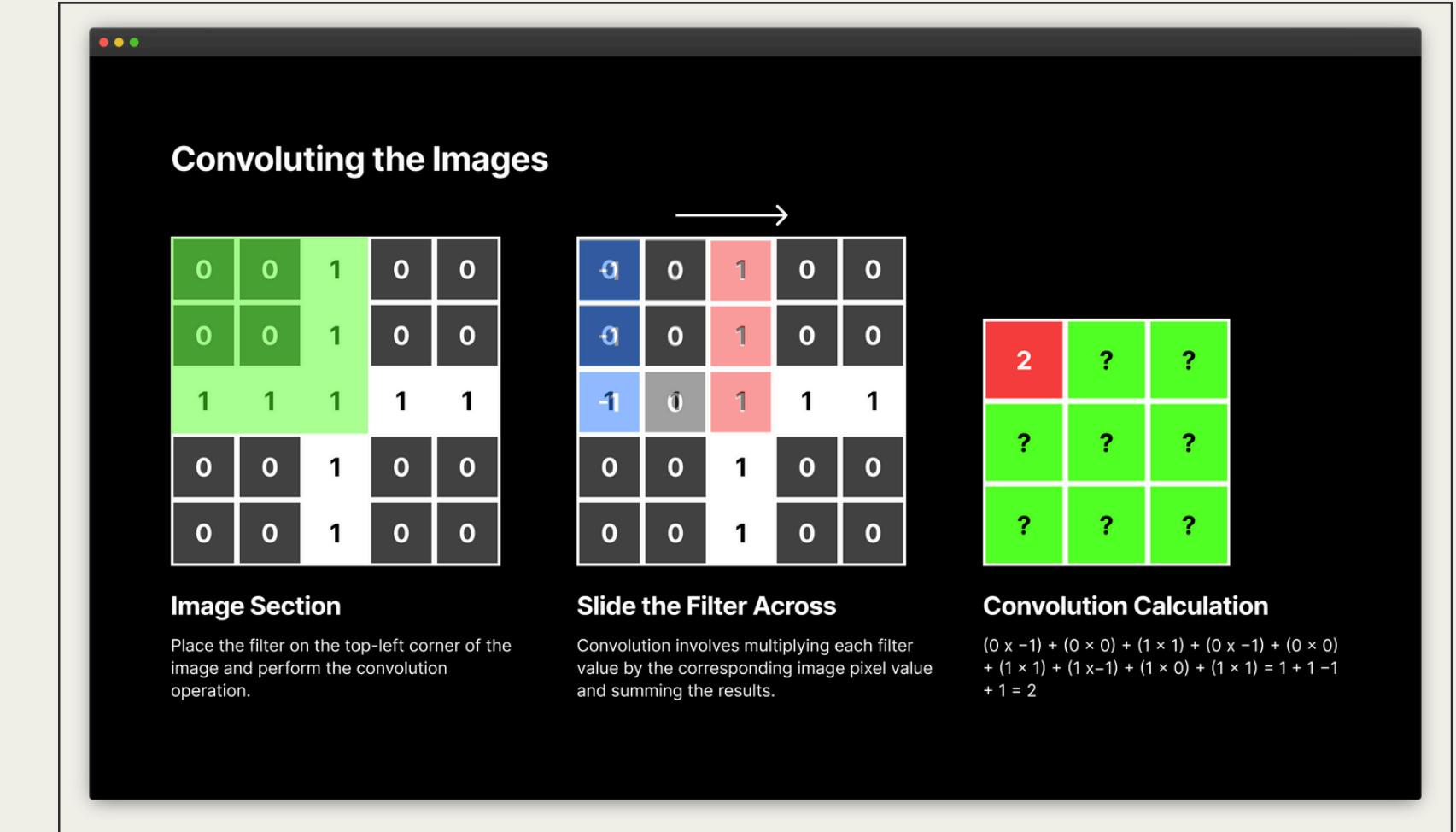


Smilkov, D., Carter, S., Sculley, D., Viégas, F. B., & Wattenberg, M. (2017). *Direct-manipulation visualization of deep networks*. arXiv preprint arXiv:1708.03788.



CONVOLUTIONAL NEURAL NETWORKS

Convolutional Neural Networks revolutionized computer vision by introducing specialized layers that mimic the human visual cortex, enabling efficient processing of grid-like data such as images. As LeCun et al. (2015) note, "CNNs have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent networks have shone light on sequential data such as text and speech" (p. 436).

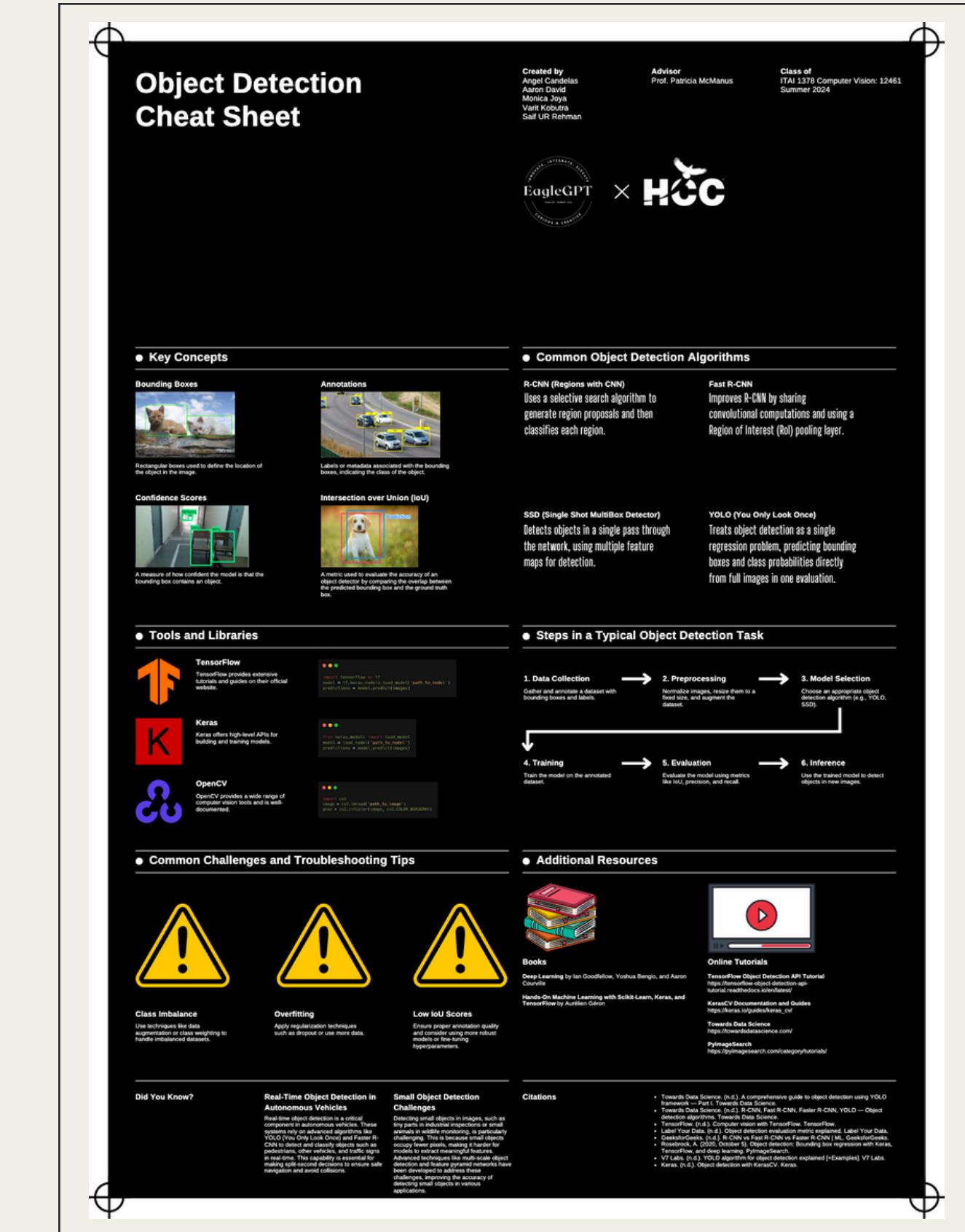


LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning*. Nature, 521(7553), 436–444.

ADVANCED CNN ARCHITECTURES

The rapid evolution of object detection architectures, from R-CNN to YOLO, demonstrates the field's progress in balancing accuracy and speed. As Zhao et al. (2019) observe, "Object detection has achieved significant breakthroughs in recent years with the rapid development of deep convolutional neural networks" (p. 1).

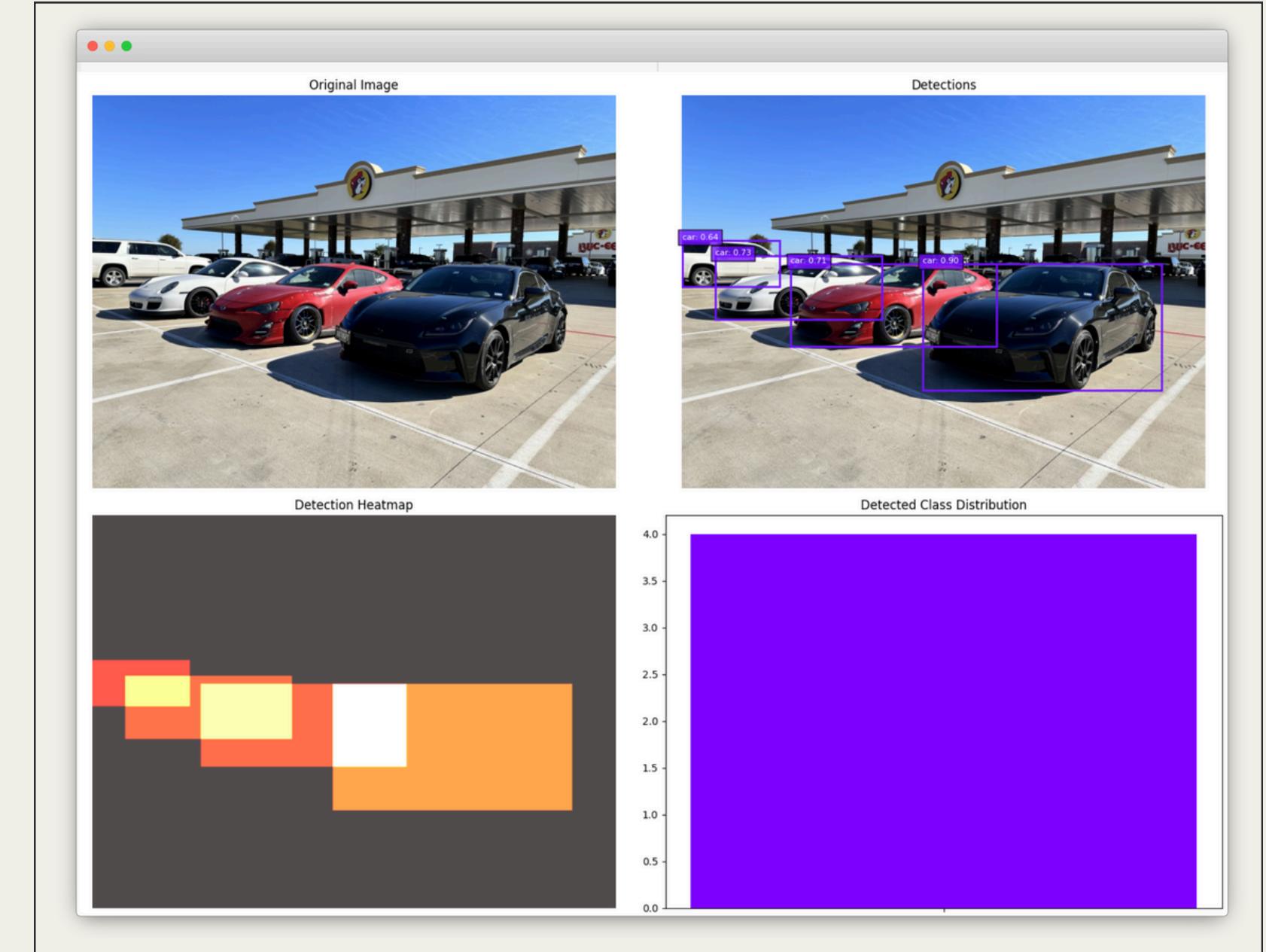
Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). *Object detection with deep learning: A review*. IEEE transactions on neural networks and learning systems, 30(11), 3212–3232.



OBJECT DETECTION AND RECOGNITION

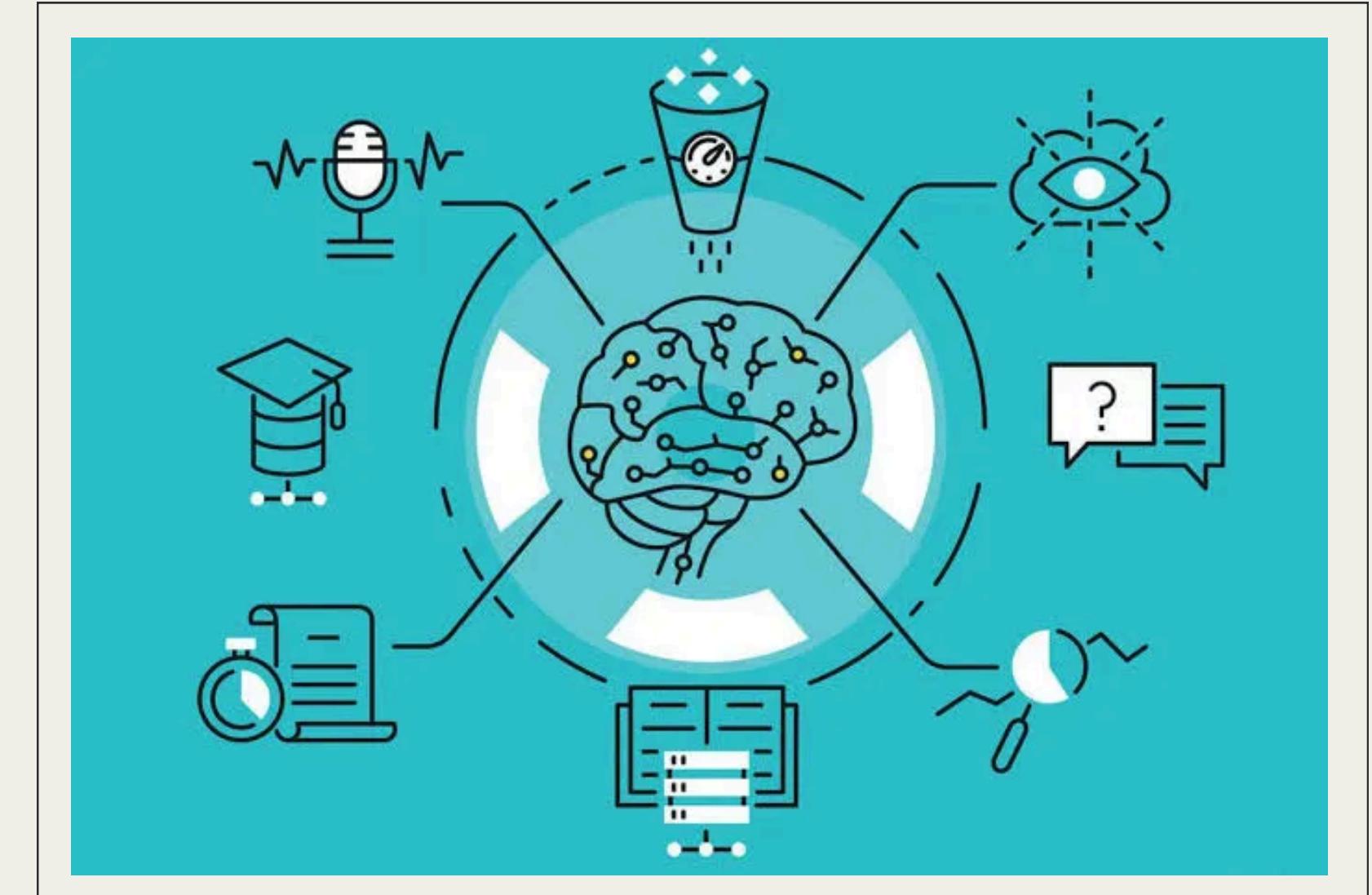
Object detection and recognition extend image classification by not only identifying what objects are in an image, but also where they are located. As Zou et al. (2019) explain, "Object detection can be viewed as a combination of classification and localization tasks, significantly increasing the complexity and utility of computer vision systems" (p. 2).

Zou, Z., Shi, Z., Guo, Y., & Ye, J. (2019). *Object detection in 20 years: A survey*. arXiv preprint arXiv:1905.05055.



THE FUTURE OF GENERATIVE AI AND AI AGENTS

The convergence of Generative AI and AI Agents with computer vision promises to revolutionize how machines perceive and interact with the world. As Ramesh et al. (2022) note, "Large language models have recently been shown to be capable of performing a wide range of tasks, from generating coherent text to manipulating images, by following natural language instructions" (p. 1), suggesting a future where AI systems can not only understand visual data but also create and manipulate it with unprecedented sophistication.



Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). *Hierarchical text-conditional image generation with clip latents*. arXiv preprint arXiv:2204.06125.

Thank you!

VISION ACHIEVED

Varit Kobutra

Professor Patricia McManus

Houston Community College

3 August 2024

Portfolio GitHub Link: <https://github.com/henrykobutra/ital-1378-portfolio>

