Computer Vision for Object Detection

Lab 58 Technology Research Brief

December 2020

Computer Vision (CV) is a use of artificial intelligence (AI) that mirrors human vision's ability to analyze complex image and video data. **Object Detection** algorithms focus on the capability for computers to segment and understand specific areas of an image. These technologies in tandem enable high-efficiency object identification and tracking in a reliable and cost-effective package.[1]

KEY TAKEAWAYS

CV uses machine or deep learning to allow computers to make sense of images.

Deep learning neural networks increase the capabilities of CV technology by automating computer learning.

CV has applications in many fields, including security, medicine, manufacturing, automotive, agriculture, and retail.

CV and Machine Learning

CV is designed to help computers "see."[2] It is powered by machine learning, a basic form of AI, that involves training algorithms on large datasets of labeled data. For example, an algorithm being instructed to differentiate between cats and dogs would need to be trained on thousands of photos of each animal to eventually learn which traits correspond to "dog" and which traits correspond to "cat."

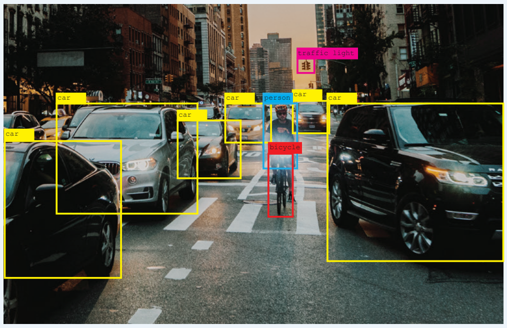


Photo by Juan Di Nella, Unsplash

Deep Learning for Object Detection

Deep learning advances AI to new levels of efficiency. Rather than setting search parameters before each test, developers first calibrate a model using similar images. This enables the model to detect any new data patterns on its own, adjusting search parameters to reflect changes automatically. Deep learning is more reliable than machine learning for inspecting large pools of data and makes use of convolutional neural networks (CNNs).

A common application for this technique is video tracking. Computers can track an object's path of travel from camera footage, with some models even performing in real-time.[3] Within RTI International, the Lab 58 team has applied video tracking for spatial arrangement. Computer data can report movement patterns for specified areas, which are then used to optimize room layouts.

[1] Juan Du, 2018. "Understanding of Object Detection Based on CNN Family and YOLO" Journal of Physics: Conference Series

[2] Atindra Bandi, et al. 2018. "Object Detection Using Google AI Open Images" Towards Data Science

[3] Rohith Gandhi, 2019. "R-CNN, Fast R-CNN, Faster R-CNN, YOLO—Object Detection Algorithms" Towards Data Science

CV Step-by-Step

CV is a complicated process and significant technical knowledge is required to successfully operate each step of the model. The steps involved in CV for object detection are described below:

1. **Train and prepare models.** Each model must be properly calibrated by thousands of images. Before use, developers assign search parameters unique to each test.
2. **Upload imaging.** Test images are provided to the model to be scanned for objects. This occurs in a stepwise fashion.
3. **Recognize and classify objects.** The model processes familiarities within the image until it forms a hypothesis. This is tested against all objects in the training database. If the hypothesis is correct, the image is labeled categorically. The location of the object is not yet known.
4. **Localize objects.** The algorithm then places many potential bounding boxes over the image and analyzes the image inside each box to determine which box is most likely the correct one. The algorithm will then place a bounding box on the section most likely containing the object.
5. **Detect additional objects.** For images with multiple objects, or when the number of objects is unknown, deep learning neural networks are used for object detection. They automatically update search parameters when necessary—a critical step in analyzing video footage because the number of objects may be in flux.
6. **Analyze data recorded.** The outputs of the algorithm are closely inspected by developers to get a sense of the model's accuracy. Because the algorithm is self-learning, it is sometimes difficult to understand the reasoning behind adjustments the algorithm has made during the test.

Popular Methods

Computational networks designed for object detection are computationally expensive. Development of these network families is split between speed and accuracy. Region-based CNN (R-CNN) models offer higher consistency, while the You Only Look Once (YOLO) method is popular for its real-time results.

The R-CNN Family

R-CNNs use image sectioning for object detection. Rather than trying to classify everything in an image at once, each frame is divided into segments. The algorithm tests what object is most likely contained within each region before combining the results, scanning each image in an average of 47 seconds.[4] For this reason, the use of R-CNN is typically limited to smaller datasets. Updates to the R-CNN model, named "Fast R-CNN," and "Faster R-CNN," have engineered ways to streamline the drawing of bounding boxes, which has dramatically increased performance speed. A common application of the R-CNN networks is in forensic analysis of traffic camera footage.

The YOLO Method

The YOLO method is an alternative method used in real-time video analysis. True to its name, the method wraps the entire image in a square grid and uses bounding boxes to label subjects based on probability matches after viewing the image only once.[5] This supports rapid object detection when details are less of a concern than speed. YOLO networks are most notably used in self-driving cars, allowing fast identification of pedestrians, road signs, obstructions, and other cars.

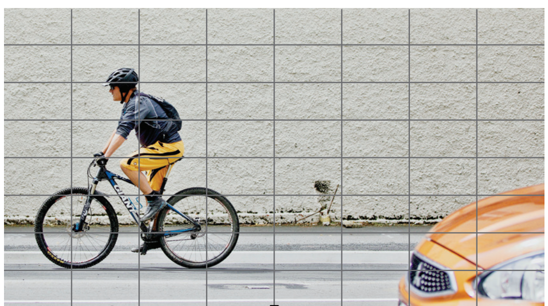


Figure 1: The YOLO algorithm initially uses a grid to divide an image into smaller parts. YOLO networks scan each box separately instead of the entire image at once.

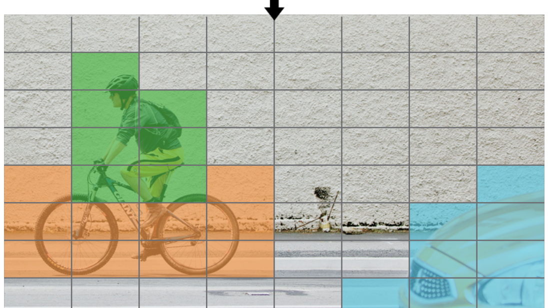


Figure 2: The contents of each section are cross-referenced with the memory bank of objects used for training. When the highest probability matches are determined, boxes with similar contents are then grouped together.

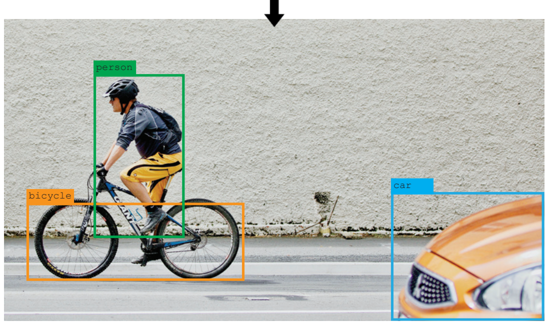


Figure 3: Each grouped region is classified as a single object. The narrowed focus of the analysis allows more efficient use of processing power, allowing the YOLO method to work in real-time.

Photo by Yolanda Sun, Unsplash

4 Li, Johnson, and Yeung, 2018. "Detection and Segmentation" Stanford University

5 Ayoosh Kathuria, 2018. "What's New in YOLO v3?" Towards Data Science

Object Tracking with Kalman Filtering

It is an important task to reliably detect and track multiple moving objects for video surveillance and other monitoring tasks. Object tracking is made difficult when the object is partially or temporarily obscured from vision. One method to make algorithms more reliable when objects become obstructed is called Kalman filtering.

Kalman filtering makes predictions about where each identified object may be in subsequent frames. The process continually gathers features and information about past movements to make better predictions about future movement for as long as an object is identified.

An additional algorithm, called the Kuhn-Munkres algorithm, allows the program to determine whether a bounding box is drawn around the same object or a new object from frame to frame. When working together, Kalman filtering and the Kuhn-Munkres algorithm allow CV applications to learn about and more accurately track multiple objects as they move through a video.[6]

Operational Limitations

When implementing video tracking in any use case, some technical requirements and recommendations will help achieve the greatest accuracy and success. Important operational limitations to consider are as follows:

Efficiency and accuracy are not always ensured

- Most algorithms are new and open-source, so properly tweaking parameters and training the AI are critical steps in preventing inconsistencies.

CV can have mixed results

- Despite meticulous calibration, self-learning AI is still unpredictable. It is difficult to anticipate the behavior of autonomous models.

Experimentation is required

- Testing and experimentation are required after calibration. This ensures model reliability across large-scale datasets.

Key Considerations for Successful Applications

Research continues to improve object recognition software on all fronts. Currently, the technology is mainly limited by model efficiency and quality of footage captured. Development continues to expand the capabilities of object detection and CV, but businesses looking to implement this technology should consider the following:

Object recognition software is complex and requires sufficient processing power

- Adequate processing power is a necessity to access the software. Navigation of these models is often complicated and can quickly lead to inefficiencies.

Consistent and reliable footage is difficult to capture

- The specifications of the capture camera, as well as its distance from desired subjects, all play a significant part in analytical accuracy. Computational models are most accurate when analyzing clear and well-lit images, and subjects that change form or figure are hard to detect through current CV methods.

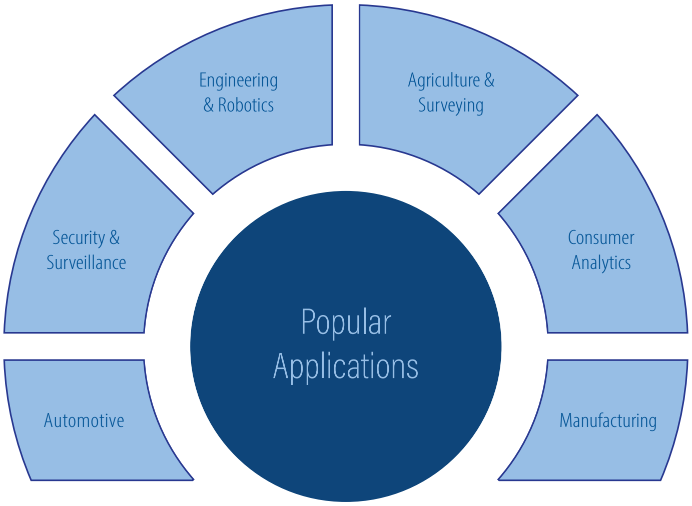
Reliable analysis requires heavy model calibration

- To avoid messy data collection, algorithms must be calibrated to detect the difference between key objects and background clutter. As a result, both training time and test speed play a role when choosing the correct model.

[6] Jeremy Cohen, 2019. "Computer Vision for Tracking" Towards Data Science

How Is CV Used in Business Today?

Examples of object detection can be seen everywhere. Within Lab 58, we have found video tracking to be particularly useful to research consumer movement patterns. This is especially useful for optimizing public spaces, such as within museums and airports. A widely publicized use of CV is within Tesla's array of self-driving vehicles, which use the YOLO method for autonomous driving. Another example of object detection includes global terrestrial mapping, a process used by the Google Earth client to identify objects from satellite footage.[7] Deep learning CV models can offer additional applications in customer service, facial recognition, emotion tracking, public transportation, and much more.[8]



Future Applications

Contemporary efforts to further this technology are focused around combining CV models with other emerging technologies. Object detection can be used alongside thermal imaging, for instance, to aid virus contact tracing. We will soon be able to track travelers with fever symptoms through airports, mapping their path of contact along the way.

CV-enhanced security systems could help prevent mass harm. Low-cost drones flying over fields could collect camera footage that, when analyzed with CV, would allow small farmers to gain insights about their crops that previously would have been prohibitively expensive to gather. CV can help in areas of traffic safety, autonomous vehicles, quality control in manufacturing scenarios, and more. Several of these applications are already being used by big businesses, but the lowering cost and barrier to entry for CV will create a proliferation wave of the technology across many fields.

[7] Mahony et al, 2019. "Deep Learning vs. Traditional Computer Vision" Cornell University

[8] Athanasios Voulodimos, 2017. "Deep Learning for Computer Vision" Computational Intelligence and Neuroscience

Ethical Concerns

Model training requires huge datasets

- Data required to train these models can be private or sensitive. Training these algorithms for uses such as facial recognition demands copious amounts of human facial data.

Results can surpass human ability

- In some cases, CV can better categorize image data than industry professionals. This raises questions about job security and machine reliability.

Applied CV could enable incredibly large surveillance networks

- CV could open the door for large-scale surveillance networks that could easily become overly invasive. When designing CV applications, it is important to consider how to collect and store only the minimal amount of data to achieve the intended results and ensure individual privacy and data security.

Learn More About Lab 58

Thanks for your interest in our work! We invite you to reach out to us for a helping hand as you explore opportunities to work with CV object detection.

Please reach out by emailing us at Lab58@rti.org. We will set up a 30-minute, one-on-one chat to discuss opportunities and answer any questions. We are interested in finding a solution that meets your needs.

For more information, contact

Lab58@rti.org