

Audience: machine learning/ computer vision amateurs - need to lower the barrier of entry to understanding ML

Length: 7 minutes of reading

Introduction

In this article, we will introduce the concept of Computer Vision and explain some of its use-cases and limitations. Computer Vision has received a lot of attention in recent years especially due to the use-cases in FinTech. We hope to demystify this unfamiliar concept and shed more light on what it can mean for us.

What is Computer Vision? Can computers actually see?

Computers process images and videos by pixels of RGB values, which are all represented in numbers. This helps computers “see” as humans do. By picking out pixels of a certain range of RGB values, we can segment out objects that are of a colour distinct from the background. This technique is called colour thresholding. As shown below, computers can now tell apart red helmets from white helmets.

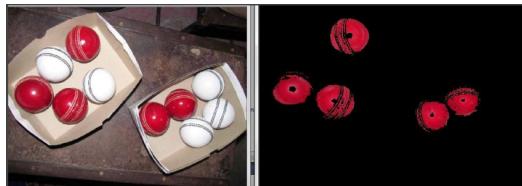


Fig 1. Using colour thresholding to detect red helmets

However, there is a downside to “seeing” with just numbers. While it may seem clear to our eyes that the object below is actually red in colour, computers do not understand the context under which this photo was taken. This is why preprocessing is crucial for Computer Vision. By performing colour balancing and removing some blue from the image, the computer can now detect the object.

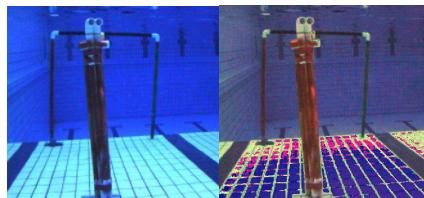


Fig 2. Using colour balancing to detect a red object under water

How can we process the data?

One very useful technique for preprocessing media data before feeding them to a machine learning model is **convolution**.

Imagine that we want to train a model such that it can tell if an image contains a cat or not. For this, we will need thousands of cat photos. But for most of them, cats will be located in the center. Without convolution, a conventional neural network will be misled into making a decision

Commented [1]: maybe we can have a one liner description of what this term means or links back after the example given.

based on the particular position of the image. Thus, when encountering an image like below where a cat is located at the side, it may fail miserably.



Fig 3. Explaining the need for convolution

With a convolutional layer, filters will be used to award scores to parts of the image that are likely to contain the desired feature. In the example of cat detection, filters for whiskers, eyes or ears will give higher scores for areas with those features. The result is a feature map, which can again undergo another convolution to derive higher-level features. For example, a feature map of lower-level features like whiskers and eyes will help us detect a higher-level feature, which is a whole face of a cat.

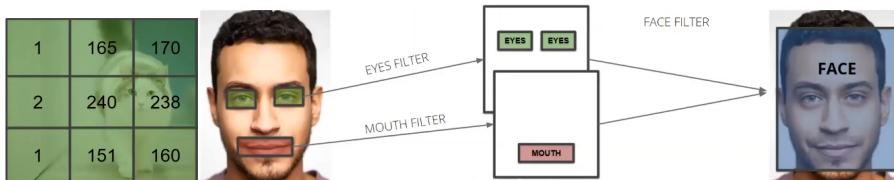


Fig 4. How filters will work to detect features in convolution

You can actually compare the effect of having a convolutional layer on a model. With this conventional neural network (<https://www.cs.ryerson.ca/~aharley/vis/fc/>), try drawing 1 at the side of the canvas. The model will likely make a wrong prediction, whereas the same input will elicit a correct output for a convolutional neural network (<https://www.cs.ryerson.ca/~aharley/vis/conv/>).

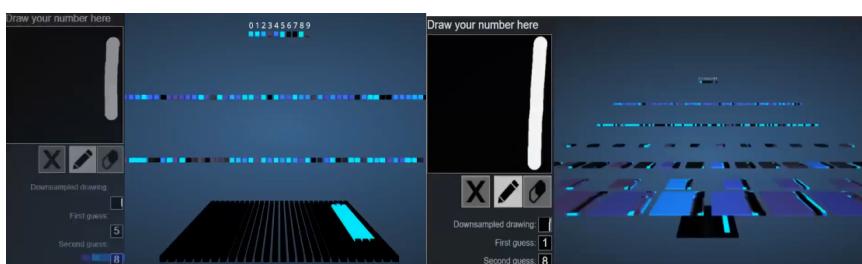


Fig 5. The effect of convolution on a neural network

Use Case #1: Traffic Monitoring

One amazing thing that we can do with Computer Vision is Object Tracking & Counting. Imagine if you could have a real-time people counter for the restaurant, library or gym that you're planning to go to - if the place is full, you would know to either find an alternative or postpone your plans.

Computer Vision Model: YOLOv5

Using Computer Vision, we can build our own traffic counter using Deep Learning. We will be using [YOLOv5](#) (You Only Look Once), an open-source object detection system that utilizes Deep Learning.

```
### INSERT AS CODE ###
git clone https://github.com/ultralytics/yolov5  # clone
cd yolov5
pip install -r requirements.txt  # install
#####
```

We will be selecting a pre-trained model to start training from. For this example, we will select [YOLOv5s](#).

```
### INSERT AS CODE ###
!wget
https://github.com/ultralytics/yolov5/releases/download/v5.0/yol
ov5s.pt
#####
```

Detection

Next, we get some training image data of traffic and run it through the model to test it. We will be using [this](#) image repository of highway surveillance cameras. To run the test, we can run the detect command.

```
### INSERT AS CODE ###
!python detect.py --weights yolov5s.pt --img 640 --conf 0.5 --
source ../DLTrafficCounter/data/bbox_highway/test
#####
```

This is an output of what our model is able to do.

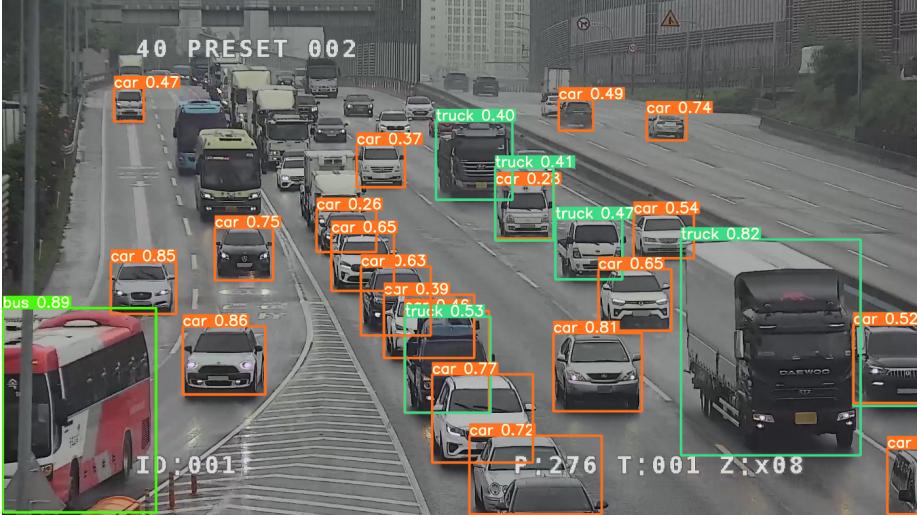


Fig 6. Detecting vehicles using YOLOv5 Model

As you can see, even based on a pre-trained model, we were already able to correctly and accurately detect many vehicles. There are some vehicles at the back that were not accurately captured but it can be remedied by training the model to identify them better.

Use Case #2: Orbital Insight

At this point, we can't help but wonder how else we can use Computer Vision. Can computer vision shine a light on financial analysis and sales predictions? Even better, how about using location analytics for humanitarian efforts? In the hands of Orbital Insight, it sure can.

Founded in 2013, Orbital Insights is pushing the boundaries of geospatial analytics, the field of computer vision technology and machine learning. Synthesizing multiple sources of geospatial data—including satellite images, mobile location, connected cars, and Internet of Things (IoT) devices—Orbital Insights provides insight into the events happening around the globe (Business Wire, 2022).

Orbital Insight is developing unique processing applications for the data they capture, creating new use cases for businesses and governments alike.

Predicting Sales Patterns:

In the brick-and-mortar space, for example, the Palo Alto-based company is able to analyze satellite imagery of hundreds of thousands of parking lots to study vehicle presence, which helps clients make sales predictions based on traffic patterns at retailers ranging from Chipotle to JC Penney to Whole Foods (Fast Company, 2022).

Humanitarian Efforts:

In one of its more compelling use cases, Orbital Insight worked with the World Bank to use satellite imagery to map poverty levels in places such as Sri Lanka and Mexico. This collaboration aims to increase awareness and response accuracy for aid organizations and NGOs while pushing local governments toward more timely action (Fast Company, 2022).

The Technology behind Orbital Insight

Computer vision, particularly multiclass object detection, is the foundational technology that can power these insights.

- Training Data - Planet's Skysat and Airbus' Pleiades clear, high-resolution global satellite imagery. Used the red, green, and blue bands from their ortho visual product, which is pan-sharpened and orthorectified.
- Data collection process - ensure that data is collected uniformly across all regions worldwide and accounted for all seasonal variations, haze, cloudy, choppy waters, and snow regions.
- Training Model - Used TensorFlow to train the model, and used Mask-RCNN network to output the masks over objects. Overcame challenges such as clustered objects, weather obstacles and low dataset diversity.

(Orbital Insight, 2021)

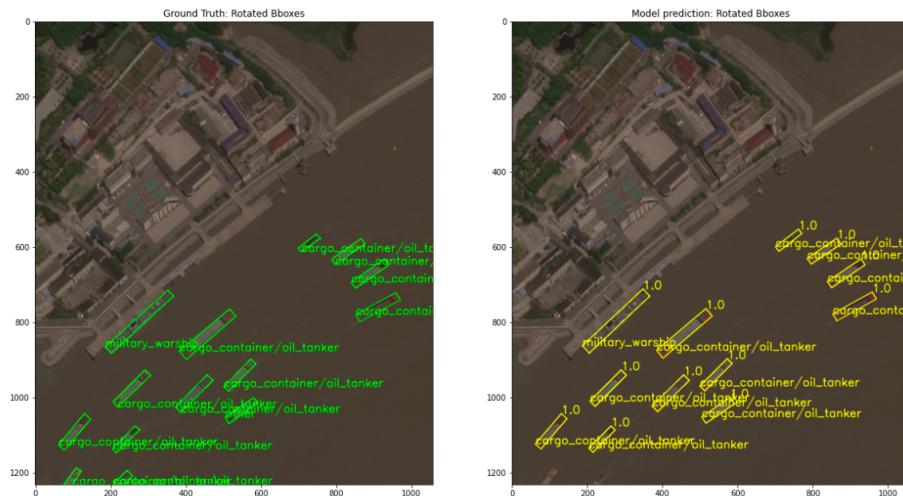


Fig 7. Container vessels carrying goods through the Huangpu River, Shanghai, China. Left is the ground truth, and the right image is the model predictions. The model can detect container vessels with high confidence.

Challenges/Limitations

We have seen the possible applications of computer vision in our daily lives and also explained the basic concepts of computer vision. On that note, people would usually begin to question why has computer vision not seen more extensive use beyond just object detection, counting and recognition. There exists a multitude of possible reasons for the hindrance in the progression of computer vision. For this, we will be generally exploring some of these challenges:

Limitation of Data:

Firstly, there is the limitation of data. More specifically, computer vision models demand images and data that is high in volume and excellent in quality.

Volume of Data

The task of gathering valid images for a computer vision model can often be very gruelling. Firstly, there is the problem of access to valid sample images. Most image public datasets available are very noisy in nature and require a lot of manual man-hours to filter, prune and label. The collection of public image datasets can sometimes take up to weeks or months depending on the scope of the project. On the other hand, trying to obtain more mature and quality datasets usually involves paying companies to obtain their proprietary datasets. Additionally, there is usually a concern about privacy, especially for use cases that involve more personal data (for example, computer vision in detecting tumours through CT Scans and X-rays).

As such, it is usually hard to come by a large volume of pre-processed perfect images for training a computer vision model.

Quality of Data

Commonly known as Garbage-in-garbage-out (GIGO), the quality of computer vision models is only as good as the quality of the input data or images that are provided.

In order to limit the presence of GIGO, the selection of images must be diverse, well-labelled and clear in quality. Take for example, a common computer vision recognition problem, Chihuahua or Muffin?



Fig.8. Sample dataset for Chihuahua and muffins classification

In this case, when scraping for images of Chihuahuas and muffins, we preferably would want to look for images where a muffin might look like a Chihuahua or vice versa. This is to provide the computer vision model with data that is similar but different. Hence, finding similar images would allow the computer vision model to scrutinize and recognise small differences between similar images. Thus, this would produce a model that is able to more accurately recognise characteristics belonging to Chihuahua or a Muffin specifically.

Limitations in hardware:

Unfortunately, some advances in computer vision and the current use of computer vision is limited by the hardware available. This includes sensory hardware used to collect data itself to hardware for processing these datasets.

Sensory hardware

On a more related note to the quality of data, the hardware used for collecting data must be clear and precise. This implies having cameras that are properly positioned, angled and cover all required areas for training. Other aspects to consider is the quality of video or image captured and the frames captured per second by the camera. Any compromises in these aspects would severely affect the quality of data collected and subsequently lead to GIGO.

Hardware for processing

Delving more into the processing and training of computer vision models, having the appropriate hardware for training these machine learning models is crucial as well.

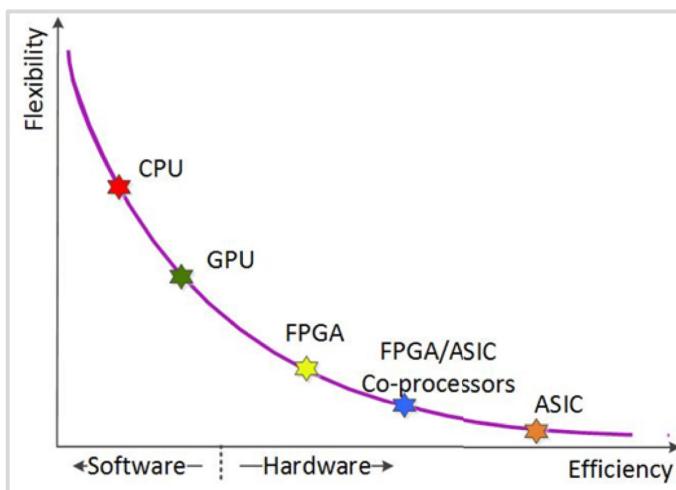


Fig.9. Computational flexibility versus computational efficiency

In the diagram above, we may easily see that typically CPUs and GPUs offer larger flexibility in computation. Put it simply, computer vision models make use of complex Deep learning and neural network techniques. These techniques require a diverse range of calculations that have to be made. As such, your CPUs and GPUs that offer this flexibility are able to meet these needs.

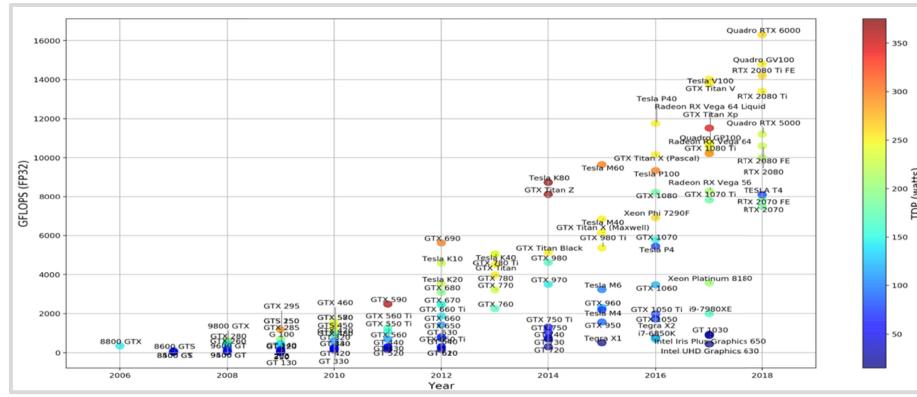


Fig.10. Computational Performance versus years of release of GPUs and power consumption.

Additionally, looking into the specifics of GPU computation in training computer vision models, we can see that more recent, expensive and advanced models of GPU have a higher GFLOPS score. GFLOPS refers to the capability of performing one billion floating-point operations per second and is usually a reference to gauge computational performance. In essence, having a newer, more advanced model of GPU, would improve the models generated when training our computer vision model.

Conclusion: What does this mean for us?

Moving into the future, we can see the immense value that Computer Vision has in our daily lives. Although not explicitly mentioned in this article, Computer Vision has made drastic impacts in FinTech. Not only can it improve our daily commute through traffic counting, it can help us better understand what is going on in different realms and help us form intelligent financial insights from it.

Computer Vision is truly a nascent technology that has yet to be fully utilized and its use-cases are seemingly endless. While there are existing limitations that pose a challenge for Computer Vision technology, we believe that in the near future, technology will improve drastically and allow Computer Vision to be commonplace in our lives.

Sources:

Introduction

Training slides from NUS Bumblebee CCA

https://www.youtube.com/watch?v=pul_qe-FKVA

<https://www.cs.ryerson.ca/~aharley/vis/fc/>

<https://www.cs.ryerson.ca/~aharley/vis/conv/>

#1 Use-Case: Traffic Monitoring

<https://github.com/ultralytics/yolov5>

<https://github.com/changsin/DLTrafficCounter>

<https://changsin.medium.com/deep-learning-for-traffic-counting-1821079d5871>

#2 Use-Case: Orbital Insight

<https://www.fastcompany.com/company/orbital-insight>

<https://via.tt.se/pressmeddelande/orbital-insight-integrates-with-esris-arcgis-platform-to-streamline-satellite-and-sensor-imagery-analysis?publisherId=259167&releaseId=3312194>

<https://orbitalinsight.com/blog/using-satellite-imagery-and-computer-vision-for-automated-ship-detection>

<https://www.prnewswire.com/news-releases/orbital-insight-unveils-multiclass-object-detection-algorithms-for-ships-aircraft-and-vehicles-301444405.html>

<https://via.tt.se/pressmeddelande/orbital-insight-integrates-with-esris-arcgis-platform-to-streamline-satellite-and-sensor-imagery-analysis?publisherId=259167&releaseId=3312194>

Limitations and challenges

General Reference:

<https://www.infopulse.com/blog/challenges-implementing-computer-vision>

Data Challenges:

<https://www.freecodecamp.org/news/chihuahua-or-muffin-my-search-for-the-best-computer-vision-api-cbda4d6b425d/>

Hardware Challenges:

<https://www.sciencedirect.com/science/article/pii/S0167926019301762#sec3>