Introduction to object detection

Subtitle

Introduction

Computer vision is currently one of the hottest fields of artificial intelligence—and **object detection** played a key role in its rapid development.

This guide will help you understand basic object detection concepts.

Introduction

1. *What's the difference between object detection and object recognition?*
2. *What are bounding boxes?*
3. *Which computer vision technique should I use?*
4. *How should I build an accurate object detection model?*

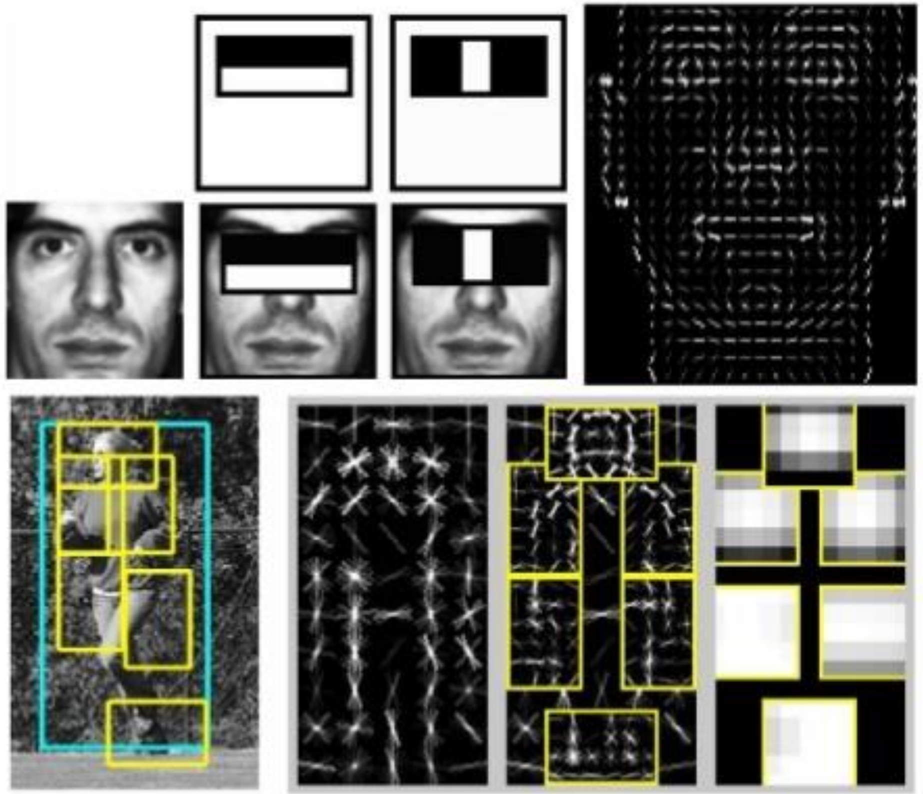
Content

Here’s what we’ll cover:

1. What is object detection?
2. Types and modes of object detection
3. How does object detection work
4. Object detection model architecture
5. Object detection applications
6. Conclusion

Object Detection 2001-2007

- Rapid Object Detection using a Boosted Cascade of Simple Features (2001)
  - Viola & Jones
- Histograms of Oriented Gradients for Human Detection (2005)
  - Dalal & Triggs
- Object Detection with Discriminatively Trained Part Based Models (2010)
  - Felzenszwalb, Girshick, Ramanan
- Fast Feature Pyramids for Object Detection (2014)
  - Dollar



Object Detection 2007-2012

Source: Ross Girshick's CVPR 2017 Tutorial http://deeplearning.csail.mit.edu/instance_ross.pptx

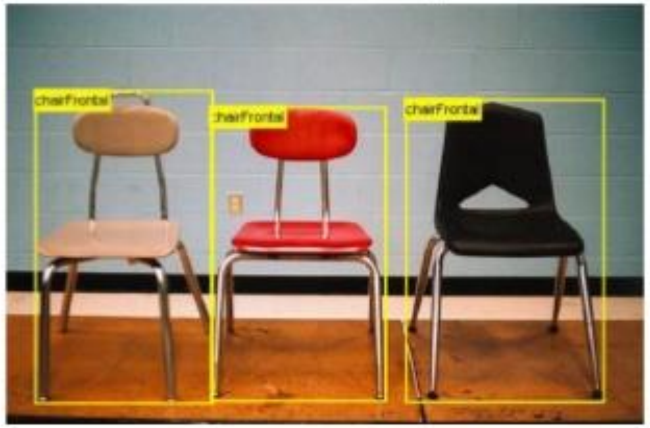
Object Detection Today

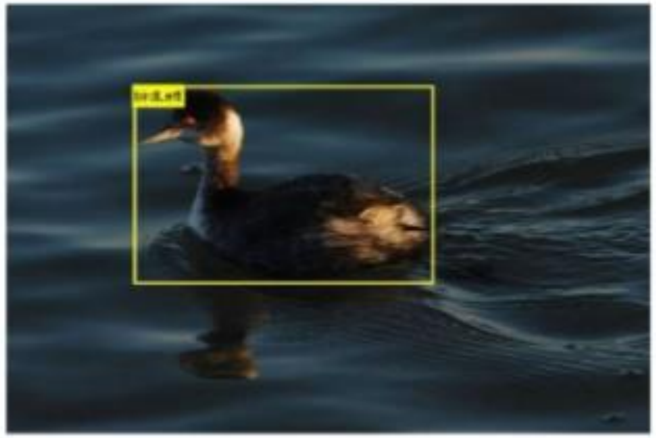
Source: Ross Girshick's CVPR 2017 Tutorial http://deeplearning.csail.mit.edu/instance_ross.pptx

Object Detection: Datasets

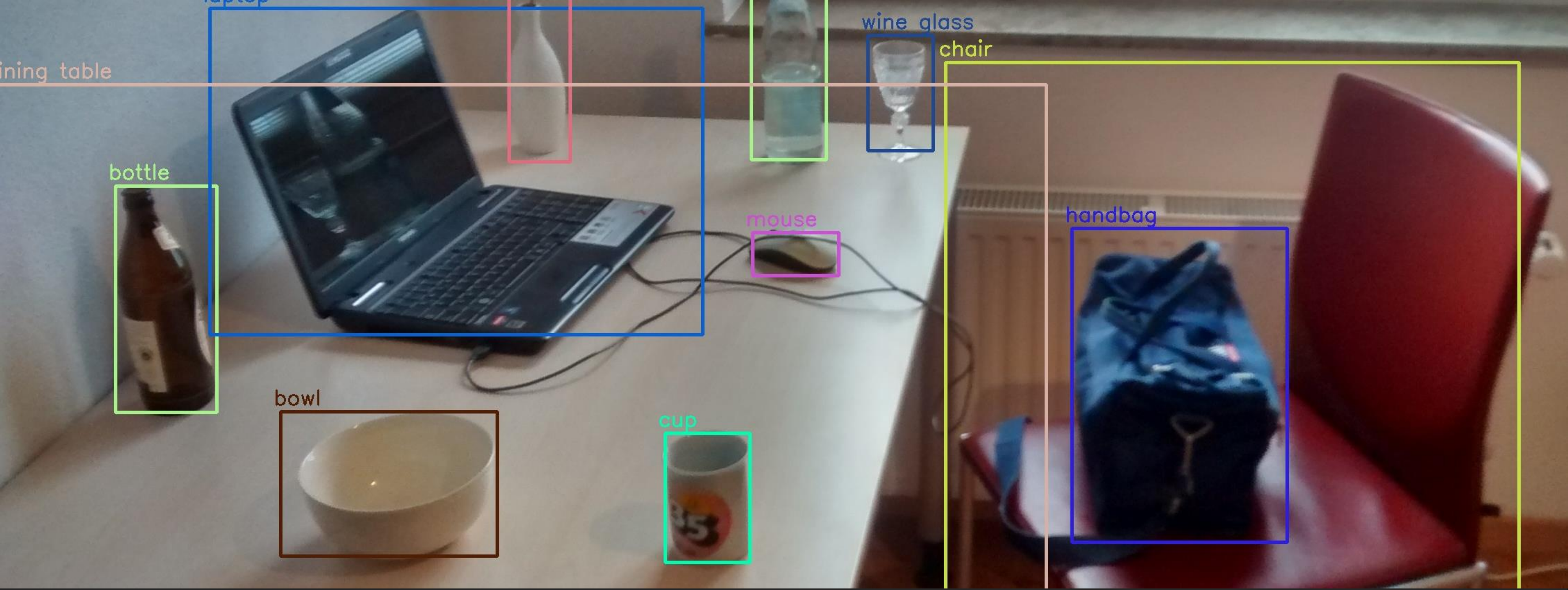
2007	2013	2015
Pascal VOC	**ImageNet ILSVRC**	**MS COCO**
• 20 Classes • 11K Training images • 27K Training objects	• 200 Classes • 476K Training images • 534K Training objects	• 80 Classes • 200K Training images • 1.5M Training objects
Was de-facto standard, currently used as quick benchmark to evaluate new detection algorithms.	Essentially scaled up version of PASCAL VOC, similar object statistics.	More categories and more object instances in every image. Only 10% of images contain a single object category, 60% in Pascal. More small objects than large objects.

Pascal Examples





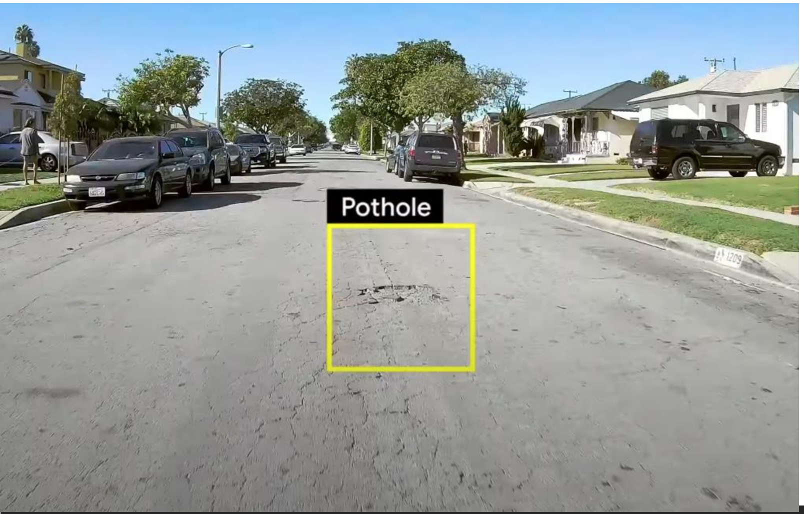
COCO Examples

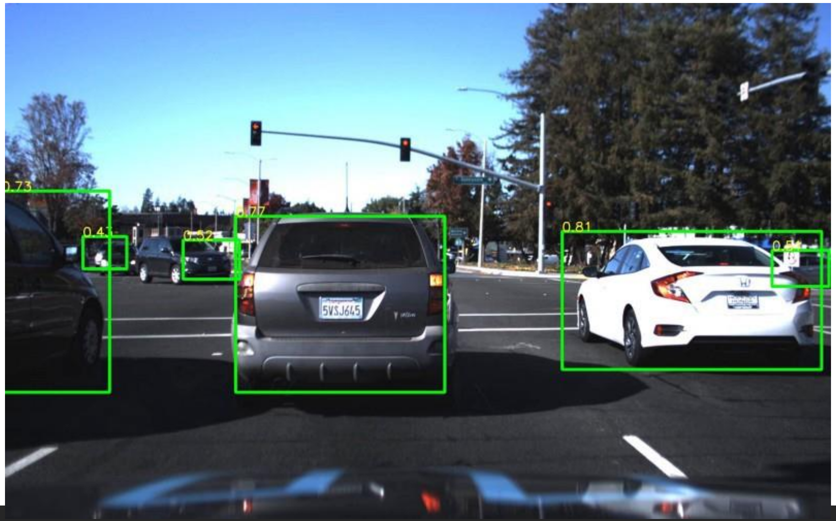


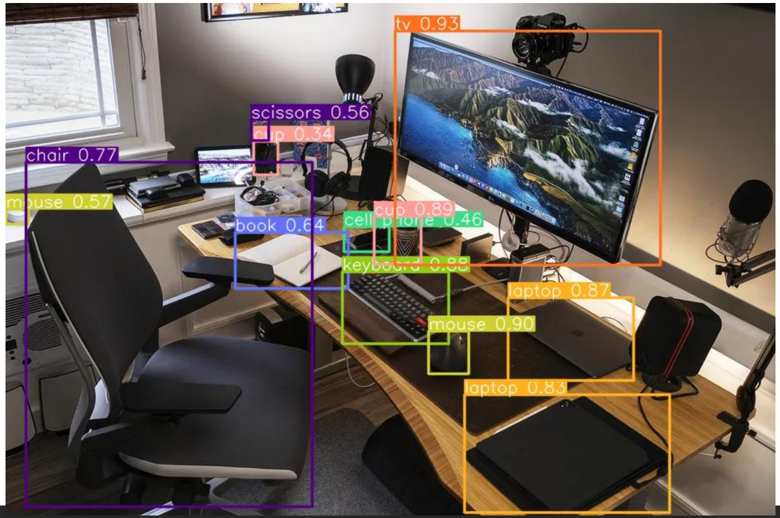
1. What is object detection?

1. What is object detection?

Object detection is the computer vision task that deals with the localization and, most of the time, classification of specific objects in images. This can be done by looking for a single object (left figure), multiple objects of the same class (middle figure) or even multiple objects of multiple classes (right figure).







https://blog.tensorflow.org/2021/06/easier-object-detection-on-mobile-with-tf-lite.html

1. What is object detection?

Object detection is the field of computer vision that deals with the **localization and classification of objects** contained in an image or video.

1. What is object detection?

Object detection is the field of computer vision that deals with the **localization and classification of objects** contained in an image or video.

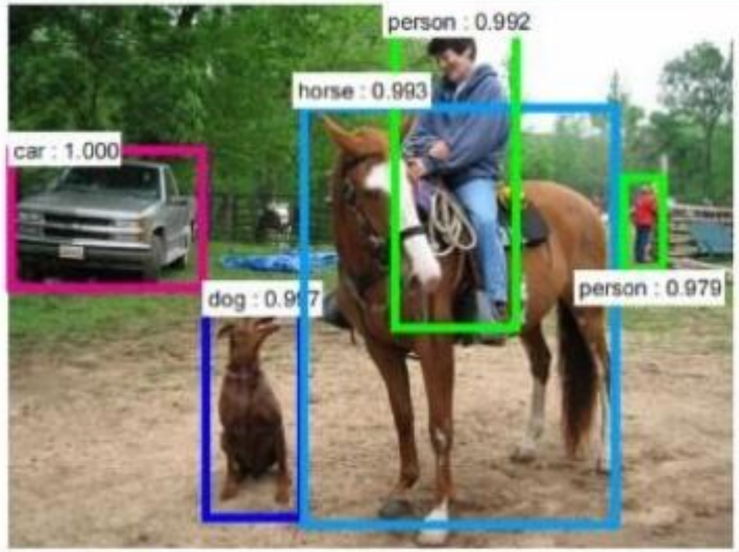
Or—

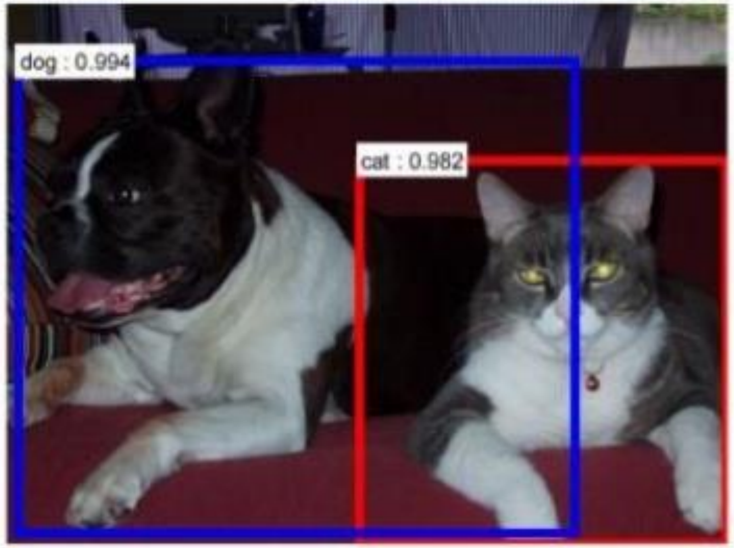
To put it simply: Object detection comes down to drawing bounding boxes around detected objects which allow us to locate them in a given scene (or how they move through it).

1. What is object detection?

Object Detection

- Input: Image
- Output: For each object class c and each image i, an algorithm returns predicted detections: $\{(b_{ij}, s_{ij})\}_{j=1}^{M}$ ocations b_{ij}th confidence scores . s_{ij}





1. What is object detection?

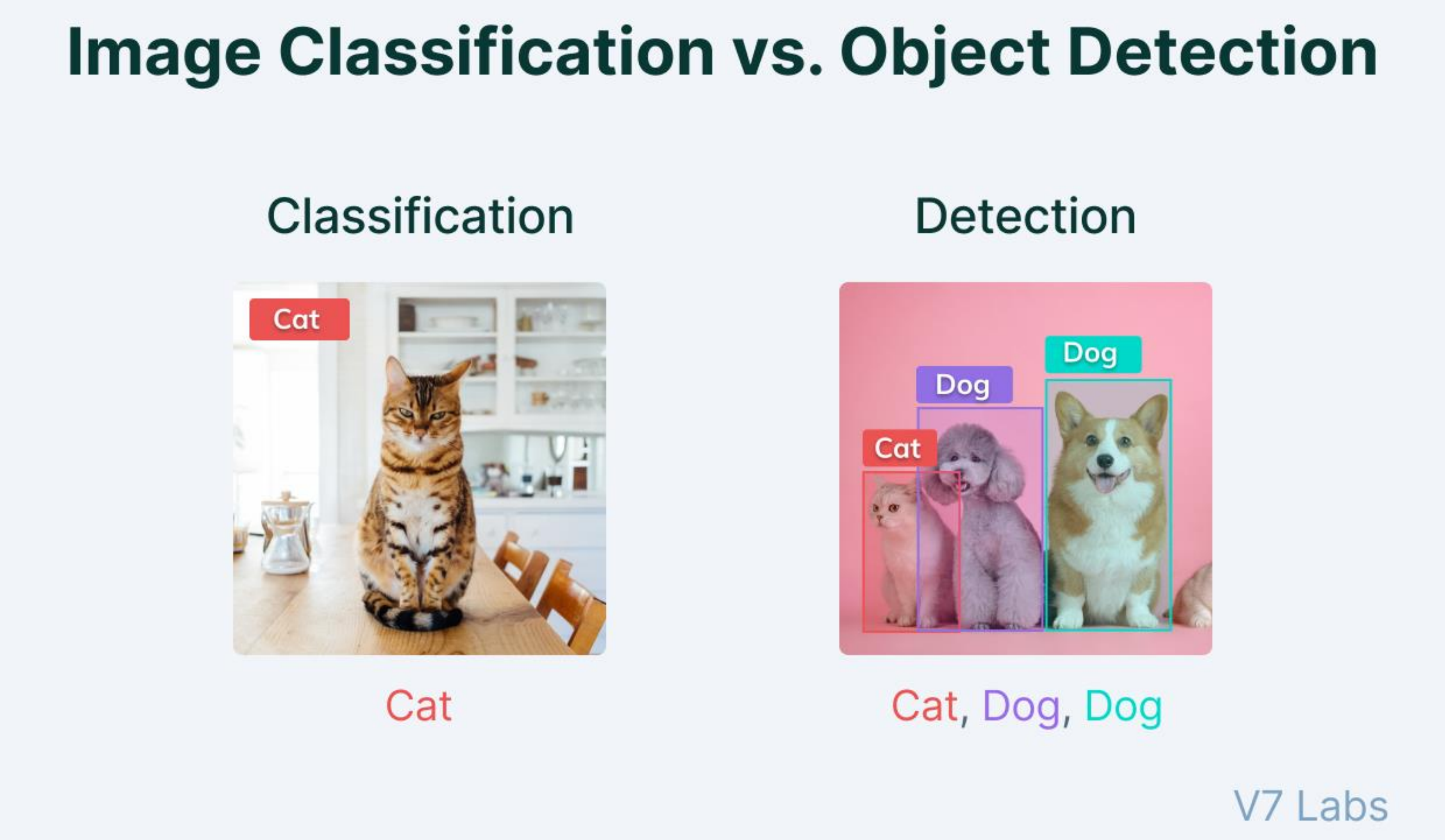
Object detection vs. image classification

Image classification sends a whole image through a classifier (such as a deep neural network) for it to spit out a tag. Classifiers take into consideration the whole image but don’t tell you *where* the tag appears in the image.

Object detection is slightly more advanced, as it creates a bounding box around the classified object.

1. What is object detection?

Object detection vs. image classification



1. What is object detection?

Object detection vs image segmentation

Image segmentation is the process of defining *which pixels* of an object class are found in an image.

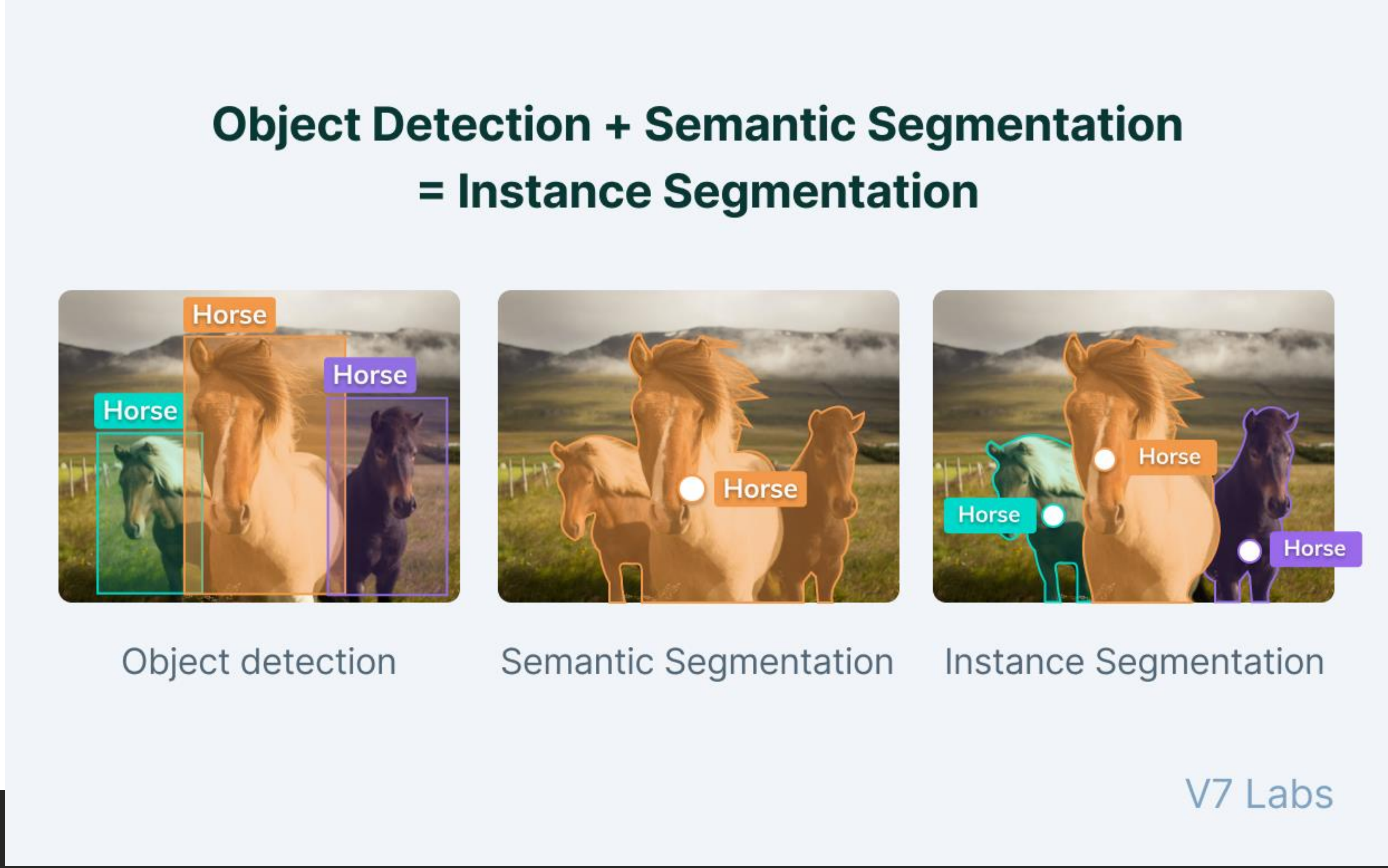
Semantic image segmentation will mark all pixels belonging to that tag, but won't define the boundaries of each object.

Object detection instead will not segment the object, but will clearly define the location of each individual object instance with a box.

Object detection vs image segmentation

Combining semantic segmentation with object detection leads to **instance segmentation,** which first detects the object instances, and then segments each within the detected boxes (known in this case as regions of interest).

Object detection vs image segmentation



1. What is object detection?

Pros and cons of object detection

Object detection is very good at:

- Detecting objects that take up between 2% and 60% of an image's area.
- Detecting objects with clear boundaries.
- Detecting clusters of objects as 1 item.
- Localizing objects at high speed (>15fps)

Pros and cons of object detection

However, it is outclassed by other methods in other scenarios.

You have to always ask yourself: *Do these scenarios apply to my problem?*

Either way, here's a cheat sheet you can use when choosing the right computer vision techniques for your needs.

1. What is object detection?

The right computer vision techniques:

Objects that are elongated—Use Instance Segmentation.

→ Long and thin items such as a pencil will occupy less than 10% of a box's area when detected. This biases model towards background pixels rather than the object itself.

1. What is object detection?

The right computer vision techniques:

Objects that have no physical presence—Use classification

→ Things in an image such as the tag “sunny”, “bright”, or “skewed” are best identified by image classification techniques—letting a network take the image and figure out which feature correlate to these tags.

The right computer vision techniques:

Objects that have no clear boundaries at different angles—Use semantic segmentation

→ The sky, ground, or vegetation in aerial images don't really have a defined set of boundaries. Semantic segmentation is more efficient at "painting" pixels that belong to these classes. Object detection will still pick up the "sky" as an object, but it will struggle far more with such objects.

The right computer vision techniques:

Objects that are often occluded—Use Instance Segmentation if possible

→ Occlusion is handled far better in two-stage detection networks than one-shot approaches. Within this branch of detectors, instance segmentation models will do a better job at understanding and segmenting occluded objects than mere bounding-box detectors.

2. Types and modes of object detection

Computer vision has made enormous progress in the last couple of decades, and object detection is not the exception.

Mainly it can be divided into two different "eras": Before and After Deep Learning (BDL and ADL).

2. Types and modes of object detection: BDL

Before using **deep learning on object detection (**took off in 2013), the methods were based on **hand-crafted features** and **classical machine learning techniques** (logistic regression, color histograms, or random forests).

These features come from various **algorithms** with information that can be obtained directly from the image.

The methods are sometimes labeled as **Traditional object detectors**.

There are 3 representative examples of this era:

1. Haar-like features: These were implemented in object-detection research by Viola and Jones. They detected faces based on 3 basic types: edge, line, and four-rectangle features.

ACCEPTED CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION 2001

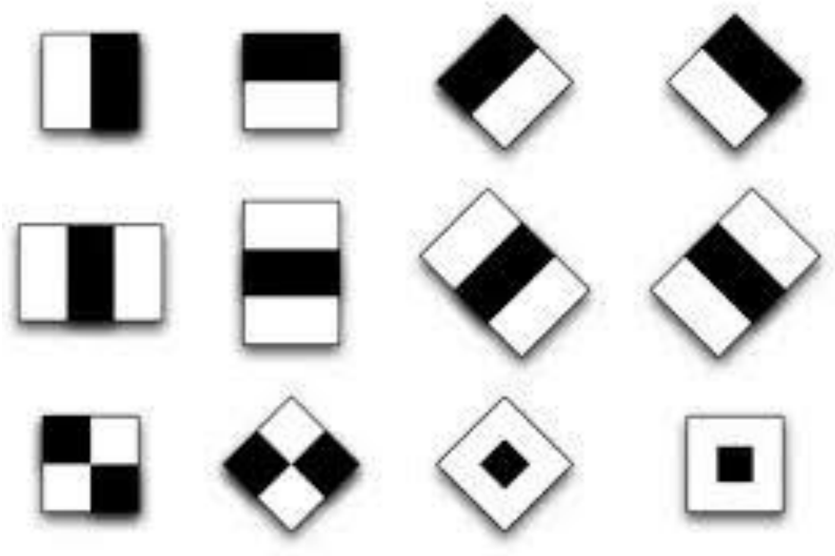
Rapid Object Detection using a Boosted Cascade of Simple Features

Paul Viola
viola@merl.com
Mitsubishi Electric Research Labs
201 Broadway, 8th FL
Cambridge, MA 02139

Michael Jones
mjones@crl.dec.com
Compaq CRL
One Cambridge Center
Cambridge, MA 02142

2. Types and modes of object detection: BDL

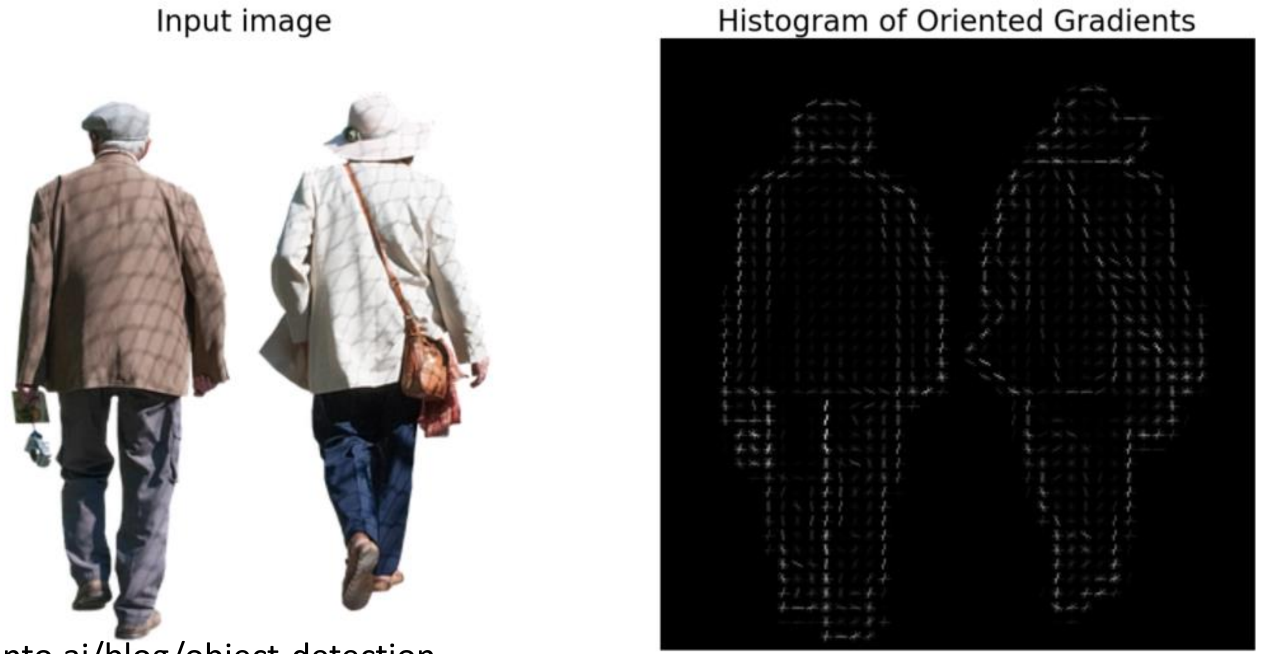
1. Haar-like features: These were implemented in object-detection research by Viola and Jones. They detected faces based on 3 basic types: edge, line, and four-rectangle features.



Nguồn: https://www.pento.ai/blog/object-detection

2. Types and modes of object detection: BDL

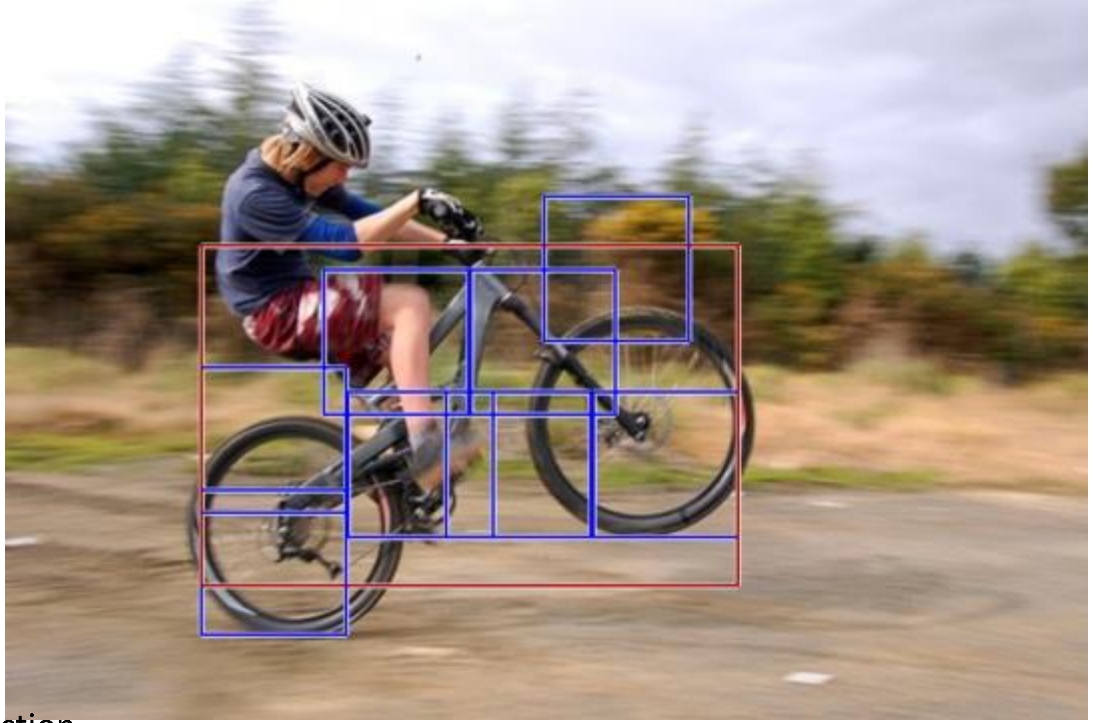
2. HOG Detector: Histograms of Oriented Gradients gained popularity after the Conference on CVPR held in 2005. This method counts how many times a gradient orientation appears in a certain portion of an image



Nguồn: https://www.pento.ai/blog/object-detection

2. Types and modes of object detection: BDL

3. DPM: Deformable Parts Model consists of a group of templates arranged in a deformable configuration. It has one global template and many part templates.



Nguồn: https://www.pento.ai/blog/object-detection

2. Types and modes of object detection: BDL

Progress got stuck around 2010 until AlexNet came up in 2012, starting the new **DL** era. This project implemented **CNN**, combined with **data augmentation**, and achieved the lowest error rates to that date.

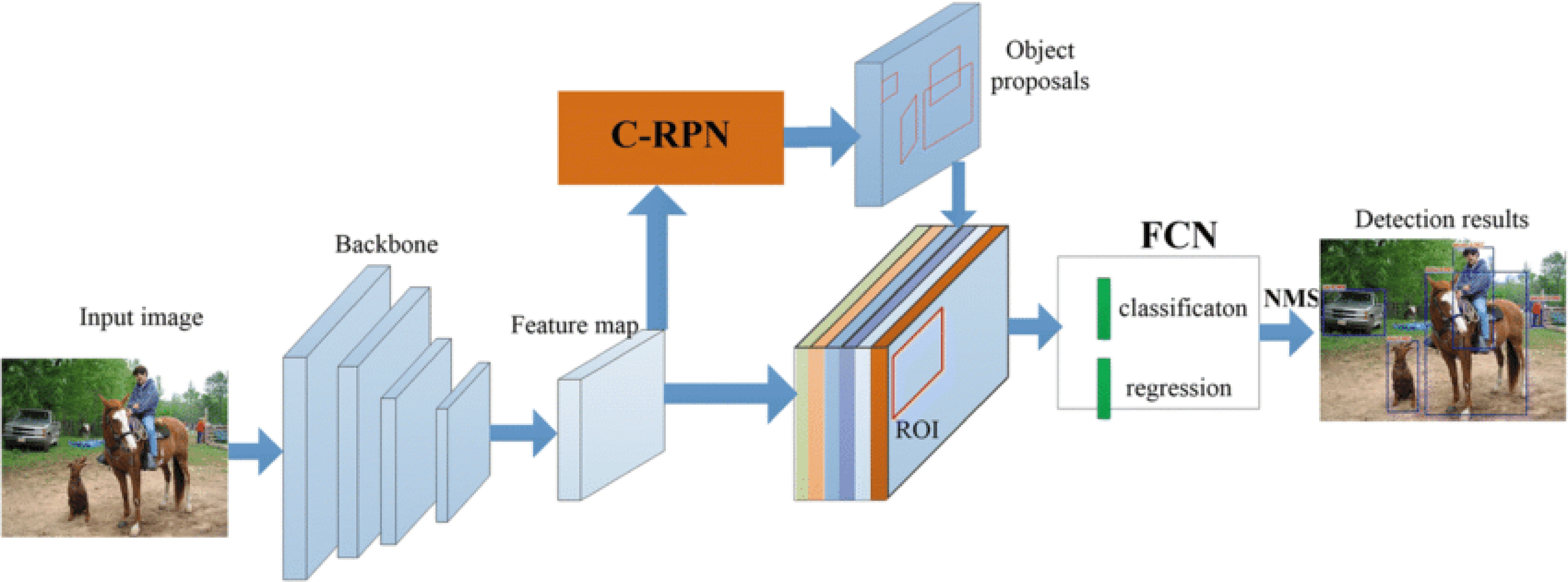
CNNs had been applied to **handwritten recognition**, but there were computational limitations and not large enough databases to scale to **object detection** in a wider range of images. AlexNet tackled this problem.

Nguồn: https://www.pento.ai/blog/object-detection

2. Types and modes of object detection

Today's deep learning-based techniques vastly outperform these.

Deep learning-based approaches use neural network architectures like RetinaNet, YOLO (You Only Look Once), CenterNet, SSD (Single Shot Multibox detector), Region proposals (R-CNN, Fast-RCNN, Faster RCNN, Cascade R-CNN) for feature detection of the object, and then identification into labels.



3. How does object detection work

3. How does object detection work

All the different models proposed in the last decade can be sorted out into 2 main categories:

1. Two-stage methods: Deriving from **R-CNN**, the method consists of a first stage where a model is used to localize possible object regions. Then its results are used as input for a second model which classifies the objects. Most of them build on previous methods and research, focusing on a specific drawback that the previous methods had. Namely SPP-Net, Fast R-CNN, Faster R-CNN, FPN, Mask R-CNN, and Cascade R-CNN.

Nguồn: https://www.pento.ai/blog/object-detection

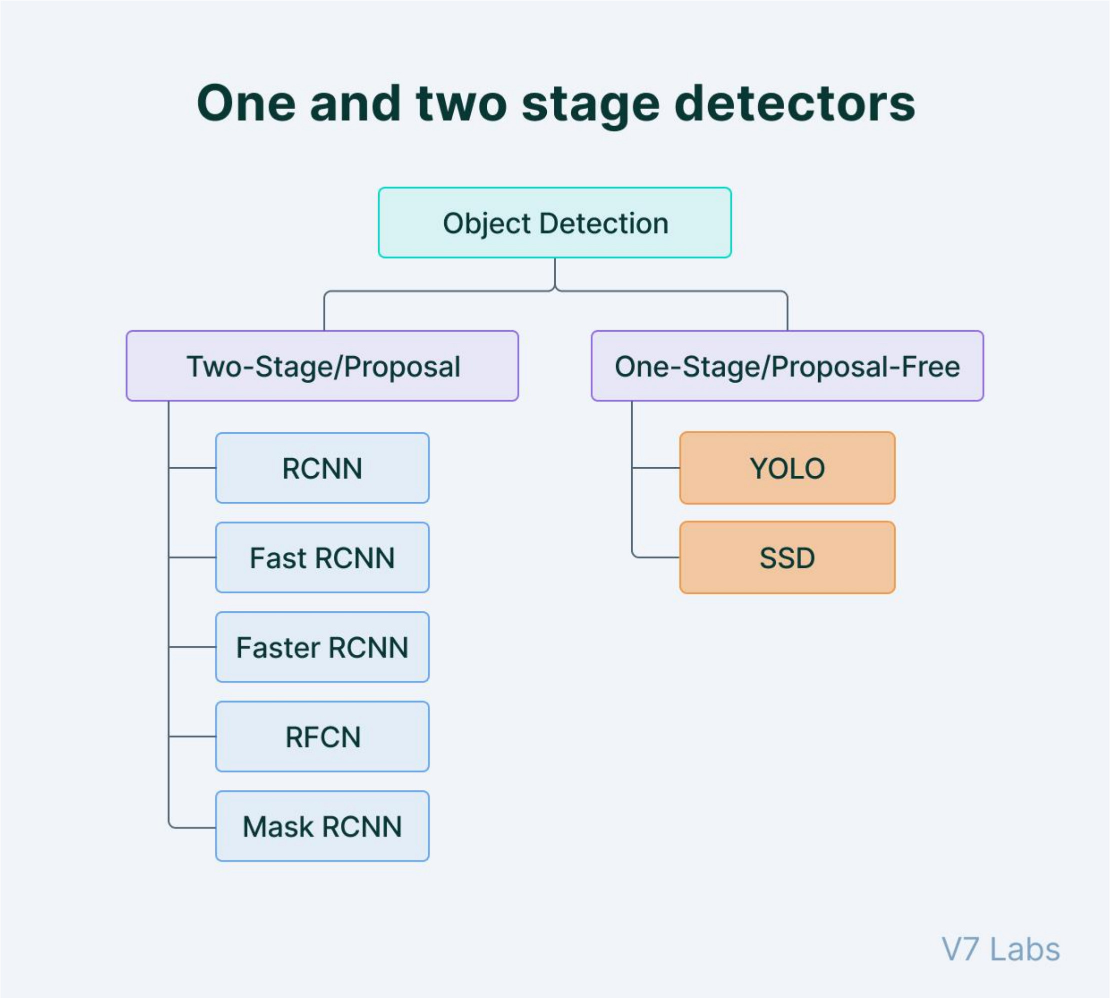
3. How does object detection work

All the different models proposed in the last decade can be sorted out into 2 main categories:

2. One-stage methods: This method directly predicts an **object's bounding boxes** for an image. As an upgrade from the previous two-stage method we discussed, this is faster and simpler but sometimes not as flexible. Examples of this are **SSD** and **YOLO**, which we will add a little demo for you to try on your own.

Nguồn: https://www.pento.ai/blog/object-detection

3. How does object detection work



Nguồn: https://www.v7labs.com/blog/object-detection-guide

3. How does object detection work

State of the art object detection architectures consists of 2 stage architectures, many of which have been pre-trained on the **COCO dataset**.

COCO is an image dataset composed of 90 different classes of objects (cars, persons, sport balls, bicycles, dogs, cats, horses e.t.c).

3. How does object detection work

The dataset was gathered to solve common object detection problems. Nowadays it is becoming outdated as its images were captured mostly in the early 2,000’s making them much smaller, grainier, and with different objects than today’s images. Newer datasets like **OpenImages** are taking its spot as the de-facto pre-training dataset.

Single-stage object detectors

A single-stage detector removes the RoI extraction process and directly classifies and regresses the candidate anchor boxes. Examples are: YOLO family (YOLOv2, YOLOv3, YOLOv4, and YOLOv5) CornerNet, CenterNet, and others. For instance, let's take a look at how YOLO Works.

Two-stage object detectors

Two-stage detectors divide the object detection task into two stages: extract RoIs (Region of interest), then classify and regress the RoIs. Examples of object detection architectures that are 2 stage oriented include R-CNN, Fast-RCNN, Faster-RCNN, Mask-RCNN and others. Let’s take a look at the Mask R-CNN for instance.

R-CNN Model Family

The R-CNN Model family includes the following:

R-CNN—This utilizes a selective search method to locate RoIs in the input images and uses a DCN (Deep Convolutional Neural Network)-based region wise classifier to classify the RoIs independently.

SPPNet and Fast R-CNN—This is an improved version of R-CNN that deals with the extraction of the RoIs from the feature maps. This was found to be much faster than the conventional R-CNN architecture.

R-CNN Model Family

The R-CNN Model family includes the following:

Faster R-CNN—This is an improved version of Fast R-CNN that was trained end to end by introducing RPN (region proposal network). An RPN is a network utilized in generating RoIs by regressing the anchor boxes. Hence, the anchor boxes are then used in the object detection task.

Mask R-CNN adds a mask prediction branch on the Faster R-CNN, which can detect objects and predict their masks at the same time.

R-CNN Model Family

The R-CNN Model family includes the following:

R-FCN replaces the fully connected layers with the position-sensitive score maps for better detecting objects.

Cascade R-CNN addresses the problem of overfitting at training and quality mismatch at inference by training a sequence of detectors with increasing IoU thresholds.

YOLO Model Family

The YOLO family model includes the following:

YOLO uses fewer anchor boxes (divide the input image into an $S \times S$ grid) to do regression and classification. This was built using darknet neural networks.

YOLOv2 improves the performance by using more anchor boxes and a new bounding box regression method..

YOLO Model Family

The YOLO family model includes the following:

YOLOv3 is an enhanced version of the v2 variant with a deeper feature detector network and minor representational changes. YOLOv3 has relatively speedy inference times with it taking roughly 30ms per inference.

YOLOv4 (YOLOv3 upgrade) works by breaking the object detection task into two pieces, regression to identify object positioning via bounding boxes and classification to determine the object's class. YOLO V4 and its successors are technically the product of a different set of researchers than versions 1-3.

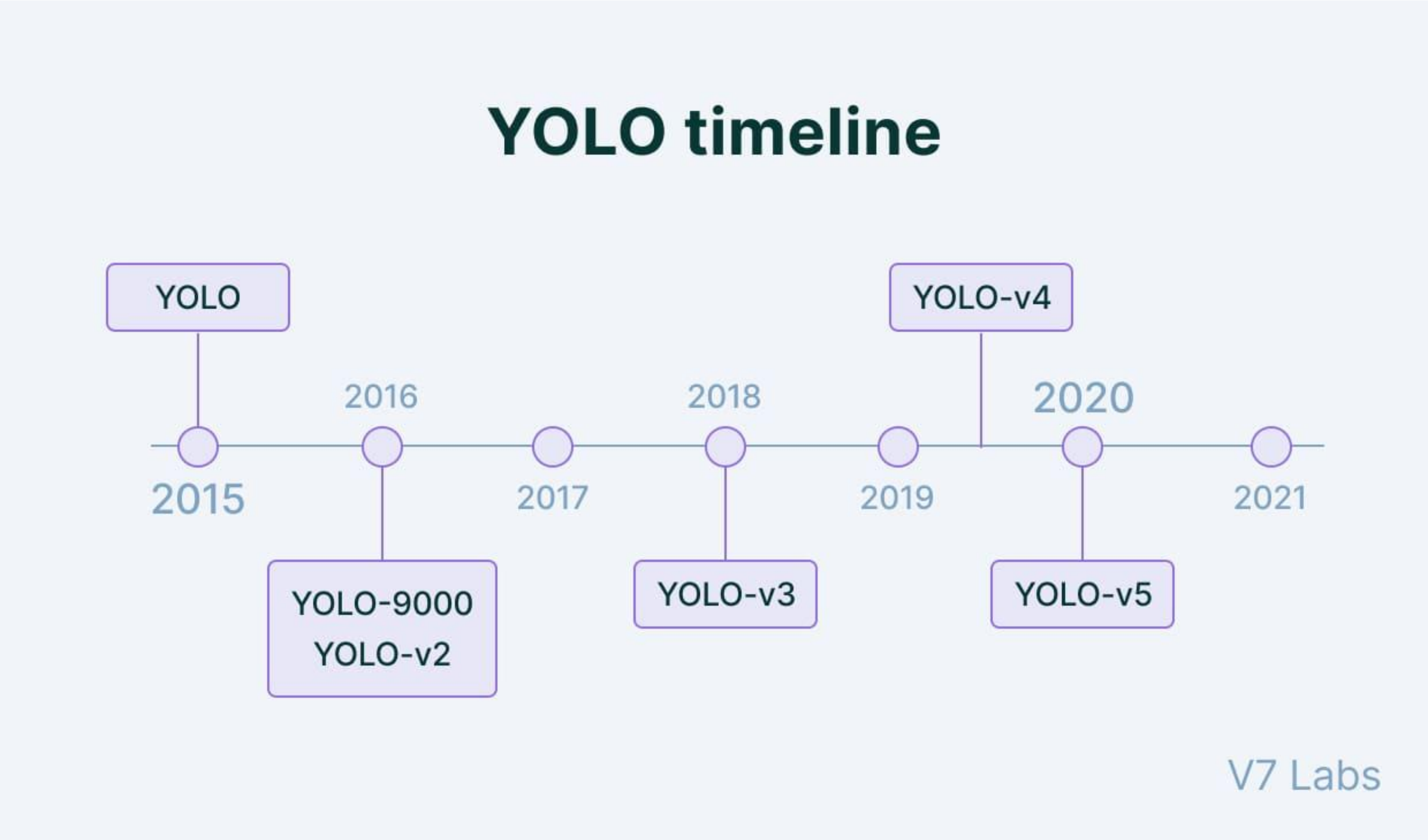
YOLO Model Family

The YOLO family model includes the following:

YOLOv5 is an improved version of YOLOv4 with a mosaic augmentation technique for increasing the general performance of YOLOv4.

YOLOv6 – 7 – X ,…

YOLO Model Family



https://www.v7labs.com/blog/yolo-object-detection

4. Object detection model architecture

CenterNet Family

The CenterNet family model includes the following:

- SSD places anchor boxes densely over an input image and uses features from different convolutional layers to regress and classify the anchor boxes.
- DSSD introduces a deconvolution module into SSD to combine low level and high-level features. While R-SSD uses pooling and deconvolution operations in different feature layers to combine low-level and high-level features.
- RON proposes a reverse connection and an objectness prior to extracting multiscale features effectively.
- RefineDet refines the locations and sizes of the anchor boxes for two times, which inherits the merits of both one-stage and two-stage approaches.
- CornerNet is another keypoint-based approach, which directly detects an object using a pair of corners. Although CornerNet achieves high performance, it still has more room to improve.
- CenterNet explores the visual patterns within each bounding box. For detecting an object, this uses a triplet, rather than a pair, of keypoints. CenterNet evaluates objects as single points by predicting the x and y coordinate of the object’s center and it’s area of coverage (width and height). It is a unique technique that has proven to out-perform variants like the SSD and R-CNN family.

CenterNet Family

The CenterNet family model includes the following:

SSD places anchor boxes densely over an input image and uses features from different convolutional layers to regress and classify the anchor boxes.

DSSD introduces a deconvolution module into SSD to combine low level and high-level features. While R-SSD uses pooling and deconvolution operations in different feature layers to combine low-level and high-level features.

CenterNet Family

The CenterNet family model includes the following:

RON proposes a reverse connection and an objectness prior to extracting multiscale features effectively.

RefineDet refines the locations and sizes of the anchor boxes for two times, which inherits the merits of both one-stage and two-stage approaches.

CornerNet is another keypoint-based approach, which directly detects an object using a pair of corners. Although CornerNet achieves high performance, it still has more room to improve.

CenterNet Family

The CenterNet family model includes the following:

CenterNet explores the visual patterns within each bounding box. For detecting an object, this uses a triplet, rather than a pair, of keypoints. CenterNet evaluates objects as single points by predicting the x and y coordinate of the object's center and it's area of coverage (width and height). It is a unique technique that has proven to out-perform variants like the SSD and R-CNN family.

Face and person detection

Most face recognition systems are powered by object detection. It can be used to detect faces, classify emotions or expressions, and feed the resulting box to an image-retrieval system to identify a specific person out of a group.

Face detection is one of the most popular object detection use cases, and you are probably already using it whenever you unlock your phone with your face.

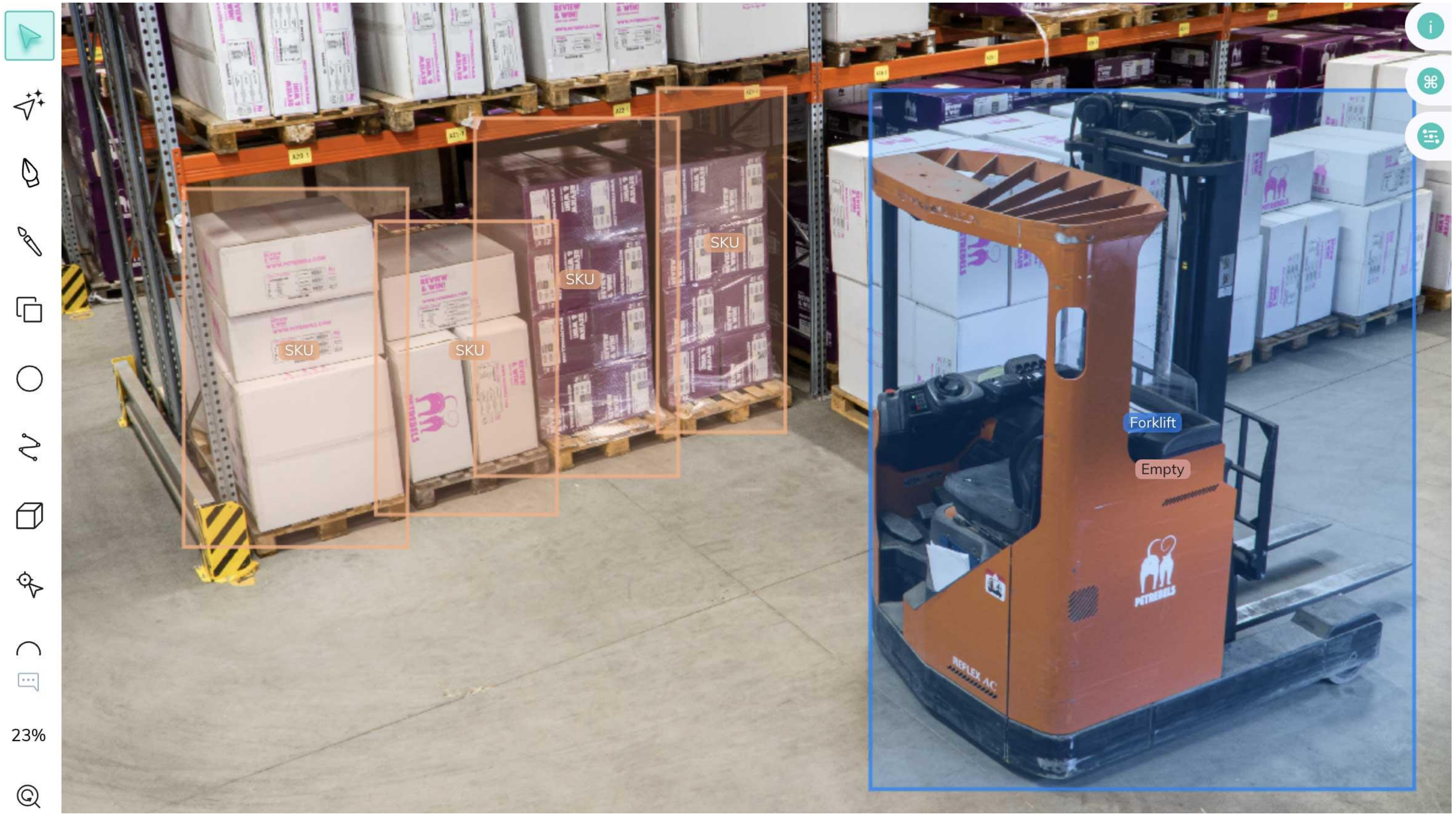
Person detection is also commonly used to count the number of people in retail stores or ensure social distancing metrics.

Face and person detection

Intelligent video analytics

Object detection is used in intelligent video analytics (IVA) anywhere CCTV cameras are present in retail venues to understand how shoppers are interacting with products. These video streams pass through an anonymizaion pipeline to blur out people's faces and de-identify individuals. Some IVA use cases preserve privacy by only looking at people's shoes, by placing cameras below knee level and ensuring the system captures the presence of a person, without having to directly look at their identifiable features. IVA is often used in factories, airports and transport hubs to track queue lengths and access to restricted areas.

Intelligent video analytics



Autonomous vehicles

Self-driving cars use object detection to spot pedestrians, other cars, and obstacles on the road in order to move around safely. Autonomous vehicles equipped with LIDAR will sometimes use 3D object detection, which applies cuboids around objects.

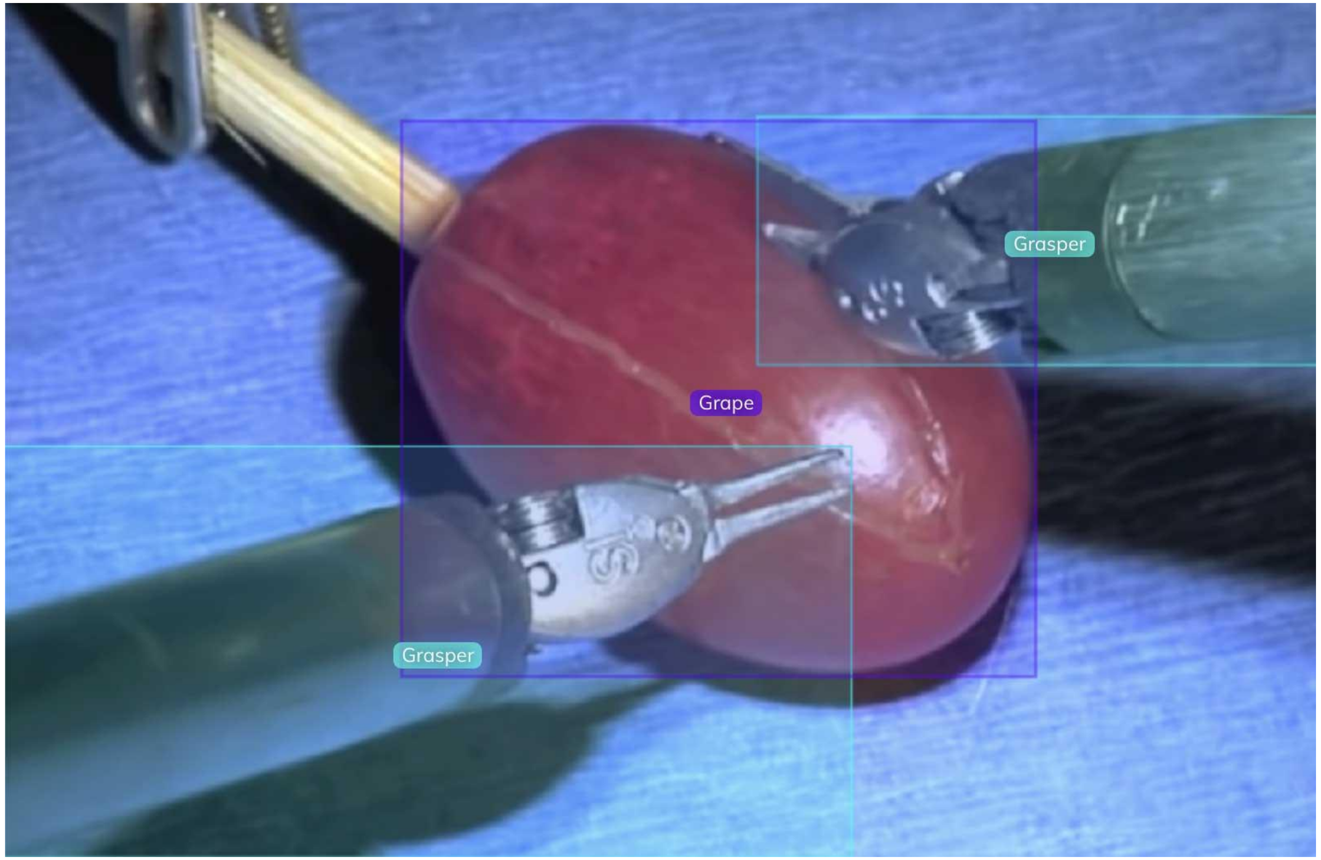
Autonomous vehicles

5. Object detection applications

Intelligence video surgery

Surgical video is very noisy data that is taken from endoscopes during crucial operations. Object detection can be used to spot hard-to-see items such as polyps or lesions that require a surgeon's immediate attention. It's also being used to inform hospital staff of the status of the operation.

Intelligence video surgery



Defect Inspection

Manufacturing companies can use object detection to spot defects in the production line. Neural networks can be trained to detect minute defects, from folds in fabric to dents or flashes in injection molded plastics.

Unlike traditional machine learning approaches, deep learning-based object detection can also spot defects in heavily varying objects, such as food.

Defect Inspection

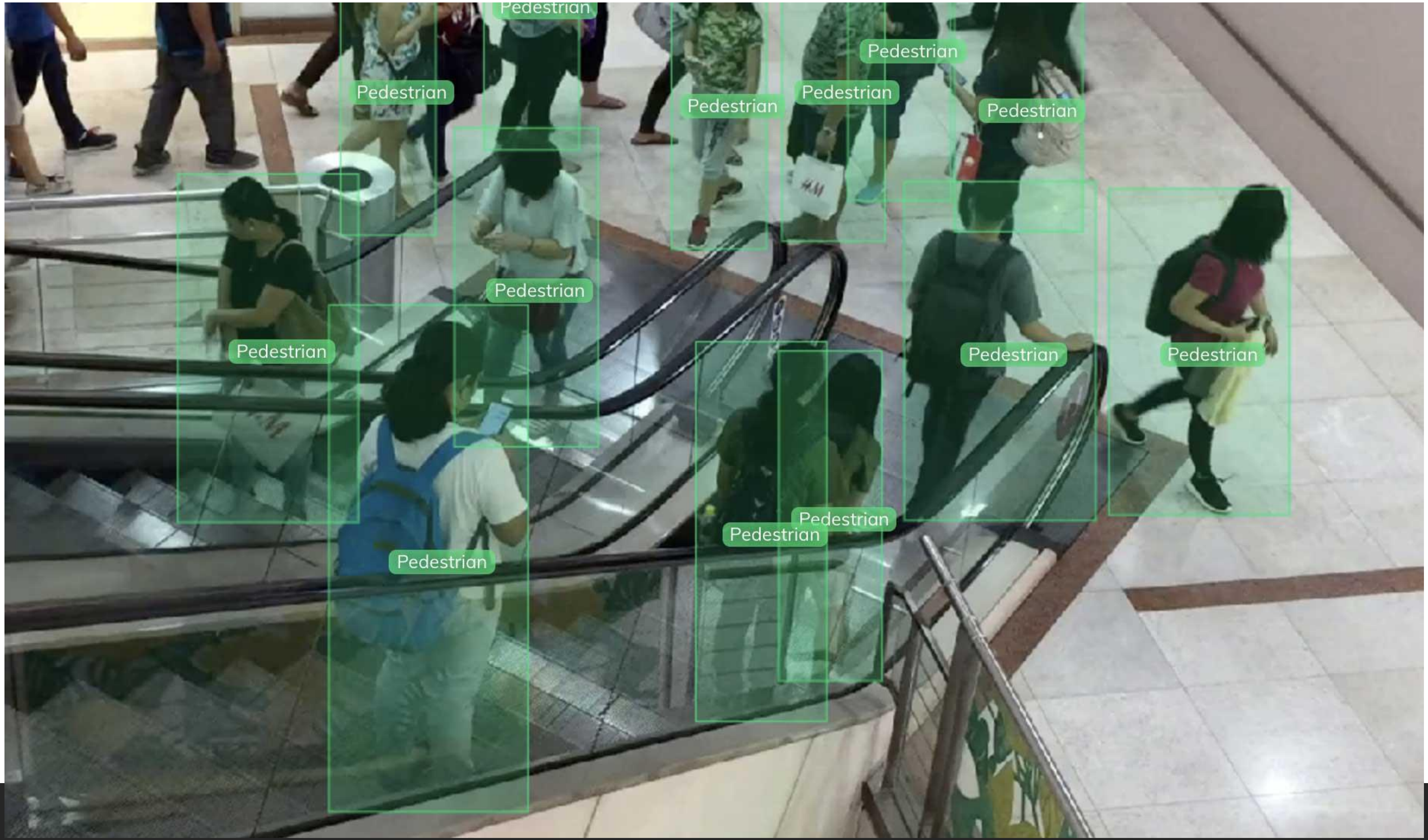
Pedestrian detection

It is one of the most essential computer vision tasks that is applied in robotics, video surveillance, and automotive safety. Pedestrian detection plays a key role in object detection research as it provides the fundamental information for the semantic understanding of video footages.

However—

Despite its relatively high performance, this technology still faces challenges such as various styles of clothing in appearance or the presence of occluding accessories that decrease the accuracy of the existing detectors.

Pedestrian detection

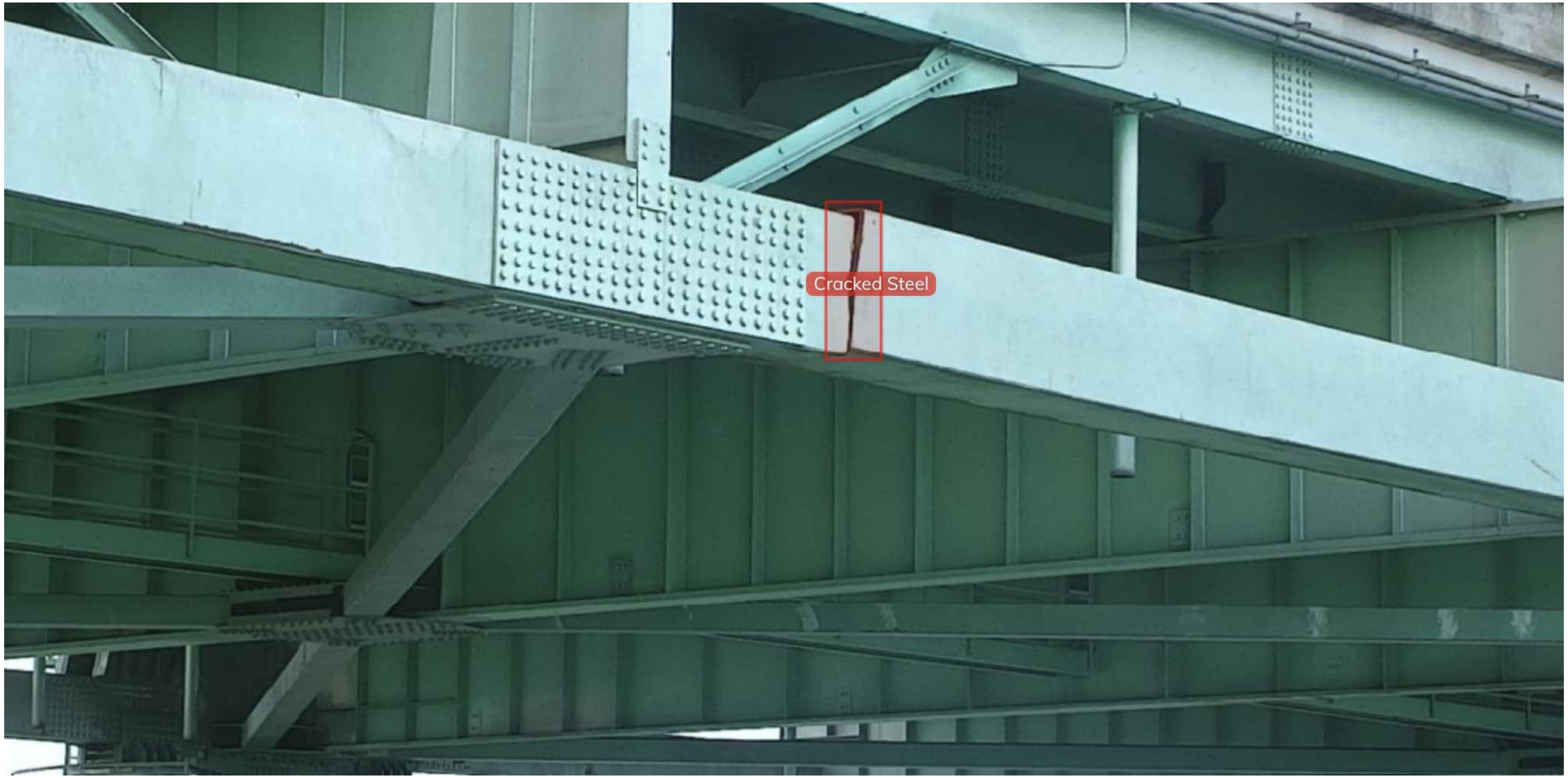


AI Drone Navigation

Drones sport incredible cameras nowadays and can leverage models hosted in the cloud to assess any object they encounter.

For example, they can be used to inspect hard-to-reach areas in bridges for cracks and other structural damage or to inspect power lines, replacing dangerous routine helicopter operations.

AI Drone Navigation



Tài liệu tham khảo

Nguồn: https://www.pento.ai/blog/object-detection

https://www.v7labs.com/blog/object-detection-guide

https://www.tensorflow.org/hub/tutorials/object_detection

https://www.datacamp.com/tutorial/object-detection-guide

https://blog.tensorflow.org/2021/06/easier-object-detection-on-mobile-with-tf-lite.html

Introduction to object detection

Subtitle

Introduction

Computer vision is currently one of the hottest fields of artificial intelligence—and **object detection** played a key role in its rapid development.

This guide will help you understand basic object detection concepts.

Introduction

1. *What's the difference between object detection and object recognition?*
2. *What are bounding boxes?*
3. *Which computer vision technique should I use?*
4. *How should I build an accurate object detection model?*

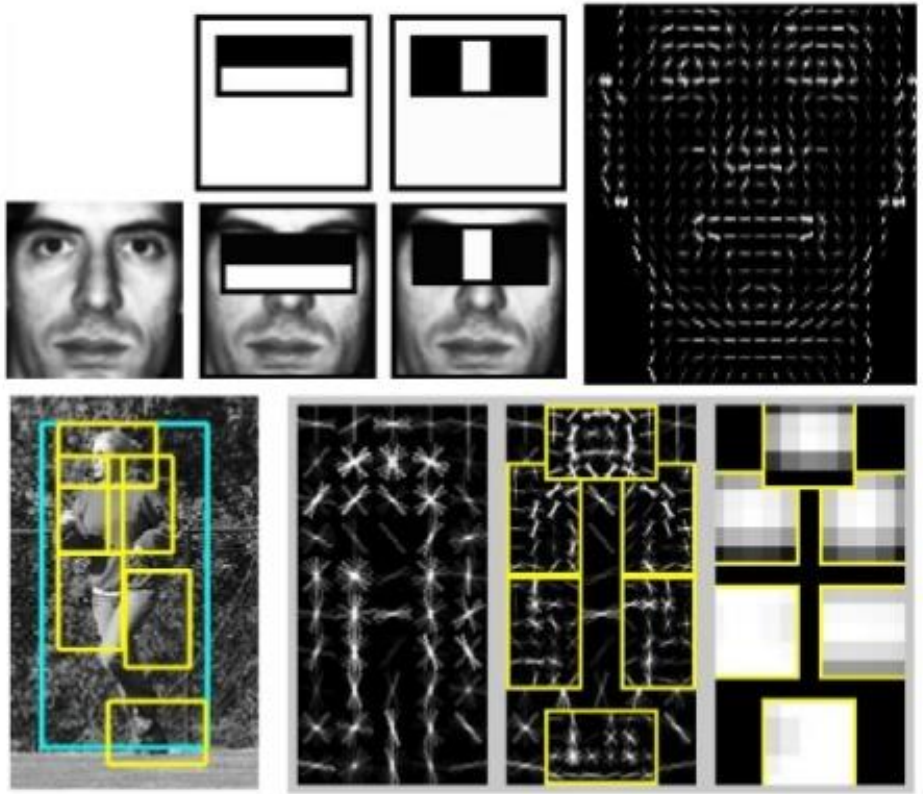
Content

Here's what we'll cover:

1. What is object detection?
2. Types and modes of object detection
3. How does object detection work
4. Object detection model architecture
5. Object detection applications
6. Conclusion

Object Detection 2001-2007

- Rapid Object Detection using a Boosted Cascade of Simple Features (2001)
  - Viola & Jones
- Histograms of Oriented Gradients for Human Detection (2005)
  - Dalal & Triggs
- Object Detection with Discriminatively Trained Part Based Models (2010)
  - Felzenszwalb, Girshick, Ramanan
- Fast Feature Pyramids for Object Detection (2014)
  - Dollar



Object Detection 2007-2012

Source: Ross Girshick's CVPR 2017 Tutorial http://deeplearning.csail.mit.edu/instance_ross.pptx

Object Detection Today

Source: Ross Girshick's CVPR 2017 Tutorial http://deeplearning.csail.mit.edu/instance_ross.pptx

Object Detection: Datasets

2007 — 2013 — 2015

Pascal VOC (2007)

- 20 Classes
- 11K Training images
- 27K Training objects

Was de-facto standard, currently used as quick benchmark to evaluate new detection algorithms.

ImageNet ILSVRC (2013)

- 200 Classes
- 476K Training images
- 534K Training objects

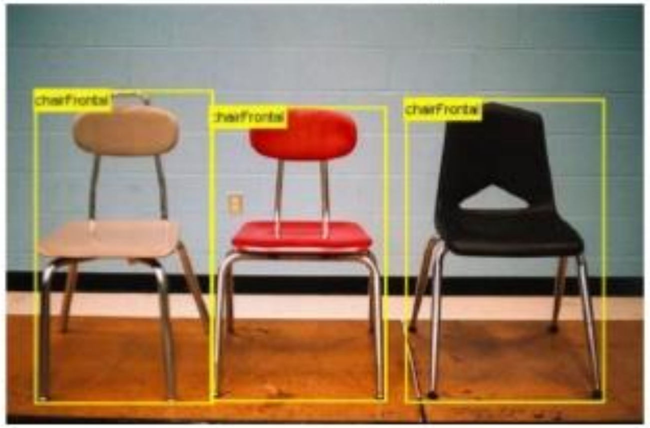
Essentially scaled up version of PASCAL VOC, similar object statistics.

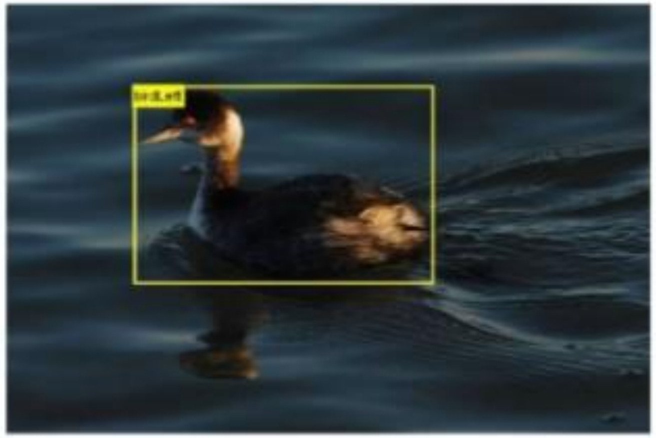
MS COCO (2015)

- 80 Classes
- 200K Training images
- 1.5M Training objects

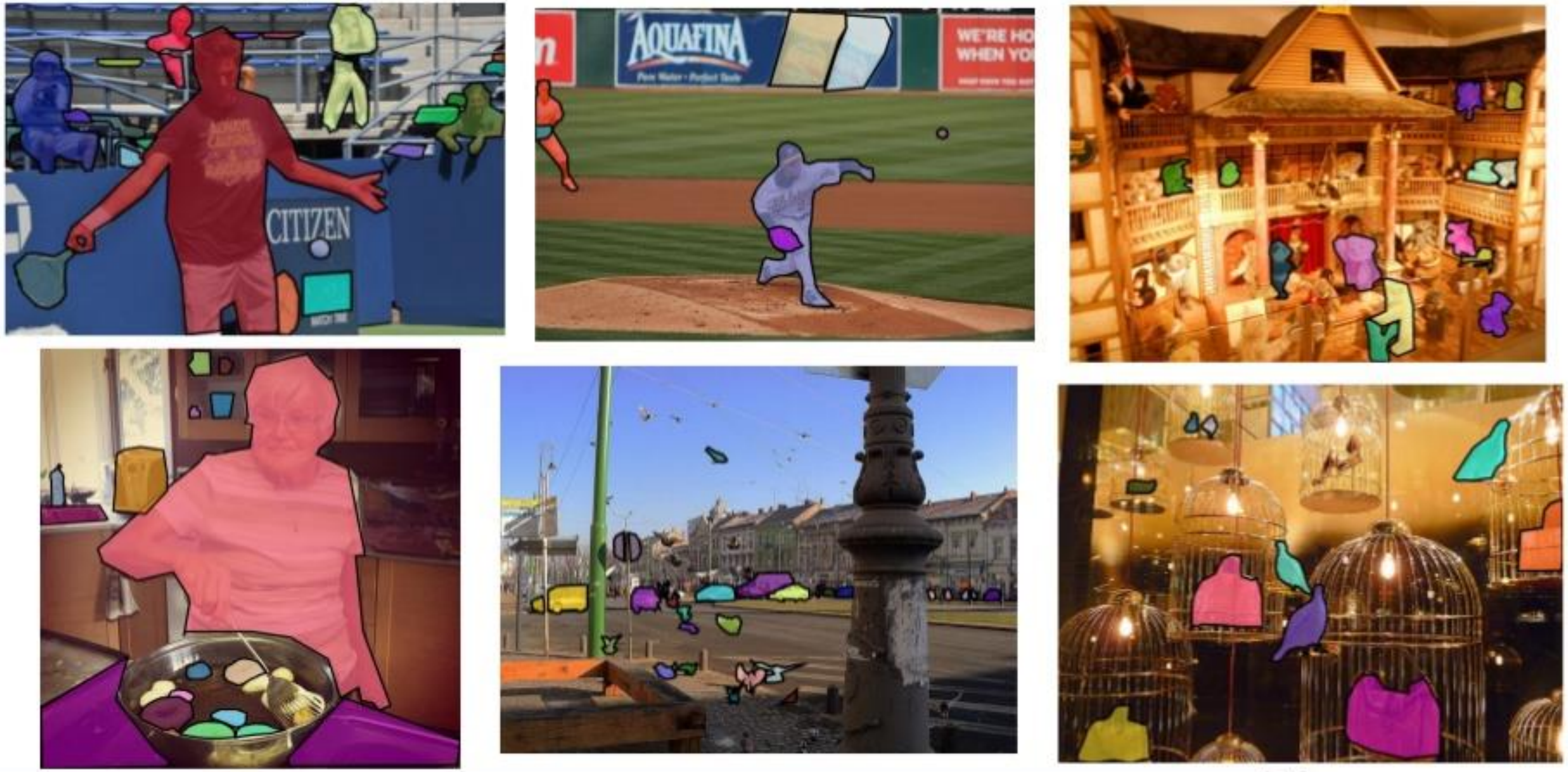
More categories and more object instances in every image. Only 10% of images contain a single object category, 60% in Pascal. More small objects than large objects.

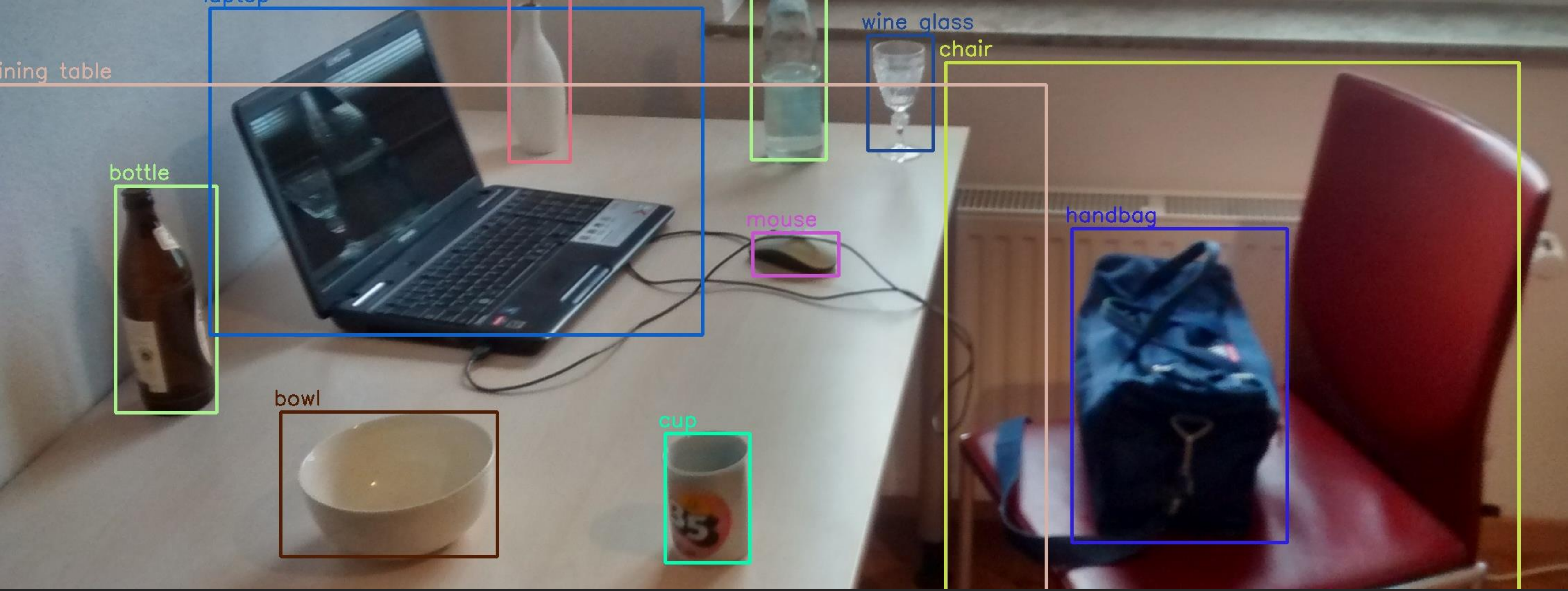
Pascal Examples





COCO Examples

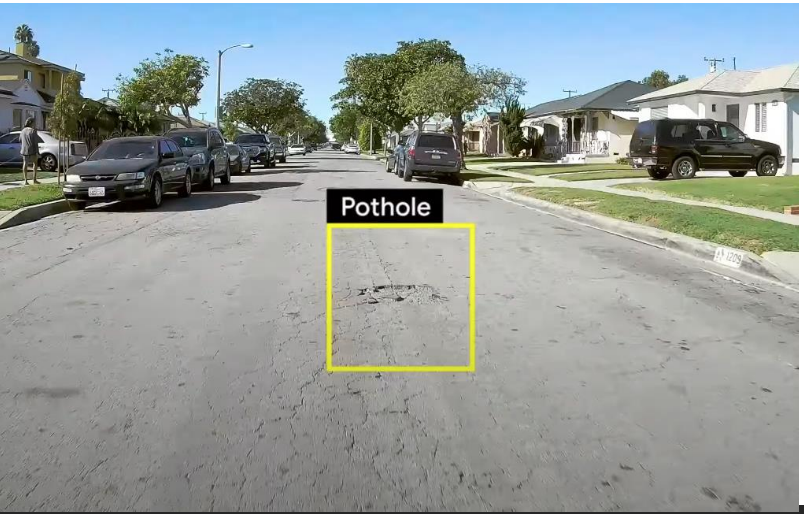


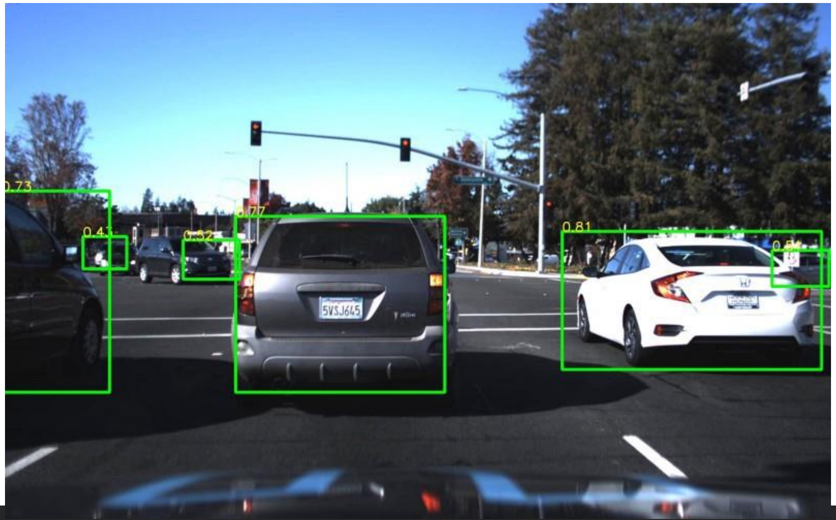


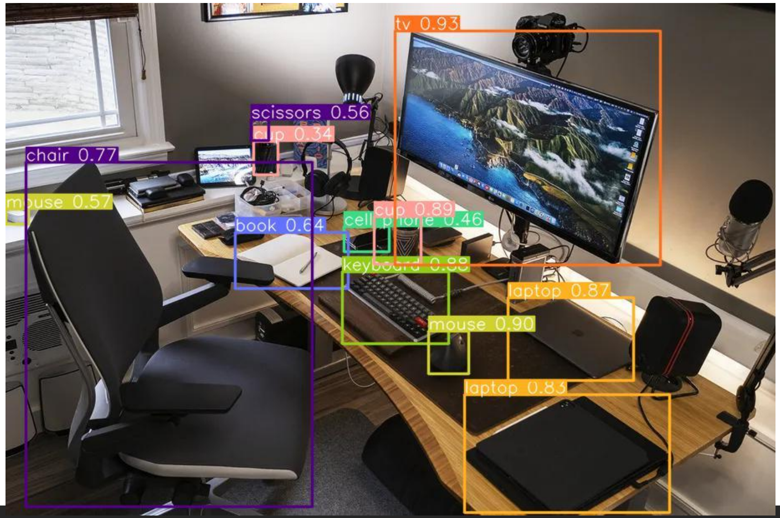
1. What is object detection?

1. What is object detection?

Object detection is the computer vision task that deals with the localization and, most of the time, classification of specific objects in images. This can be done by looking for a single object (left figure), multiple objects of the same class (middle figure) or even multiple objects of multiple classes (right figure).







https://blog.tensorflow.org/2021/06/easier-object-detection-on-mobile-with-tf-lite.html

1. What is object detection?

Object detection is the field of computer vision that deals with the **localization and classification of objects** contained in an image or video.

1. What is object detection?

Object detection is the field of computer vision that deals with the **localization and classification of objects** contained in an image or video.

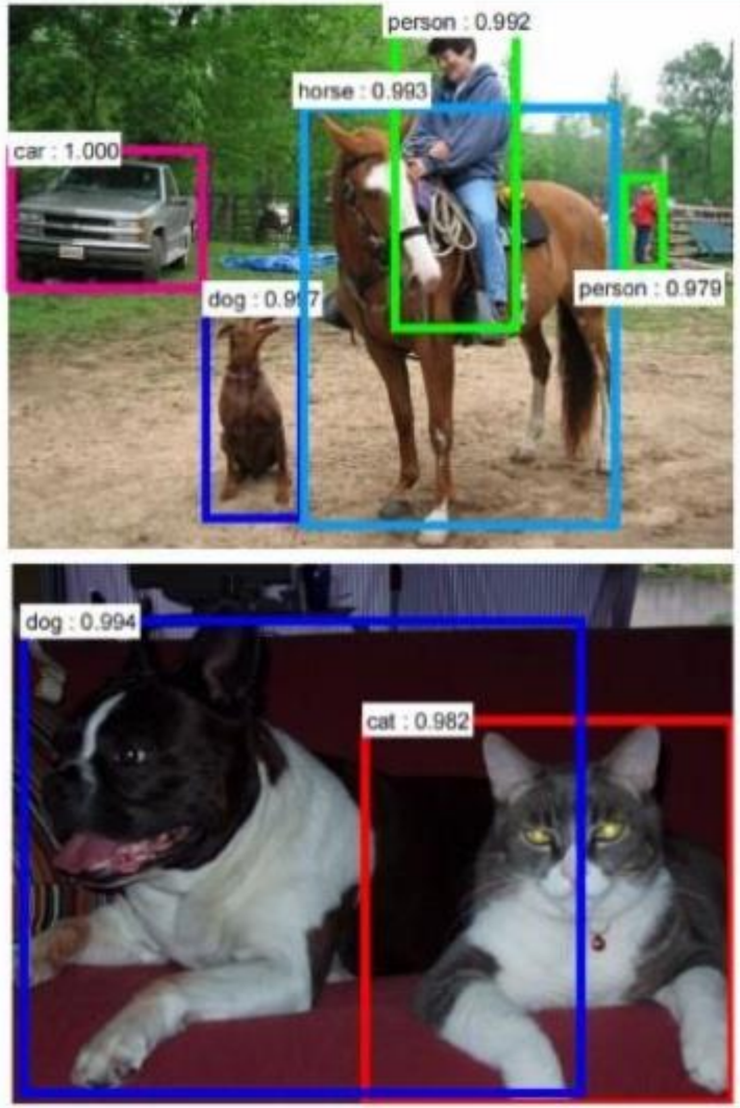
Or—

To put it simply: Object detection comes down to drawing bounding boxes around detected objects which allow us to locate them in a given scene (or how they move through it).

1. What is object detection?

Object Detection

- Input: Image
- Output: For each object class c and each image i, an algorithm returns predicted detections: $\{(b_{ij}, s_{ij})\}_{j=1}^{M}$ ocations b_{ij}th confidence scores . s_{ij}



Object detection vs. image classification

Image classification sends a whole image through a classifier (such as a deep neural network) for it to spit out a tag. Classifiers take into consideration the whole image but don't tell you *where* the tag appears in the image.

Object detection is slightly more advanced, as it creates a bounding box around the classified object.

Object detection vs. image classification

1. What is object detection?

Object detection vs image segmentation

Image segmentation is the process of defining *which pixels* of an object class are found in an image.

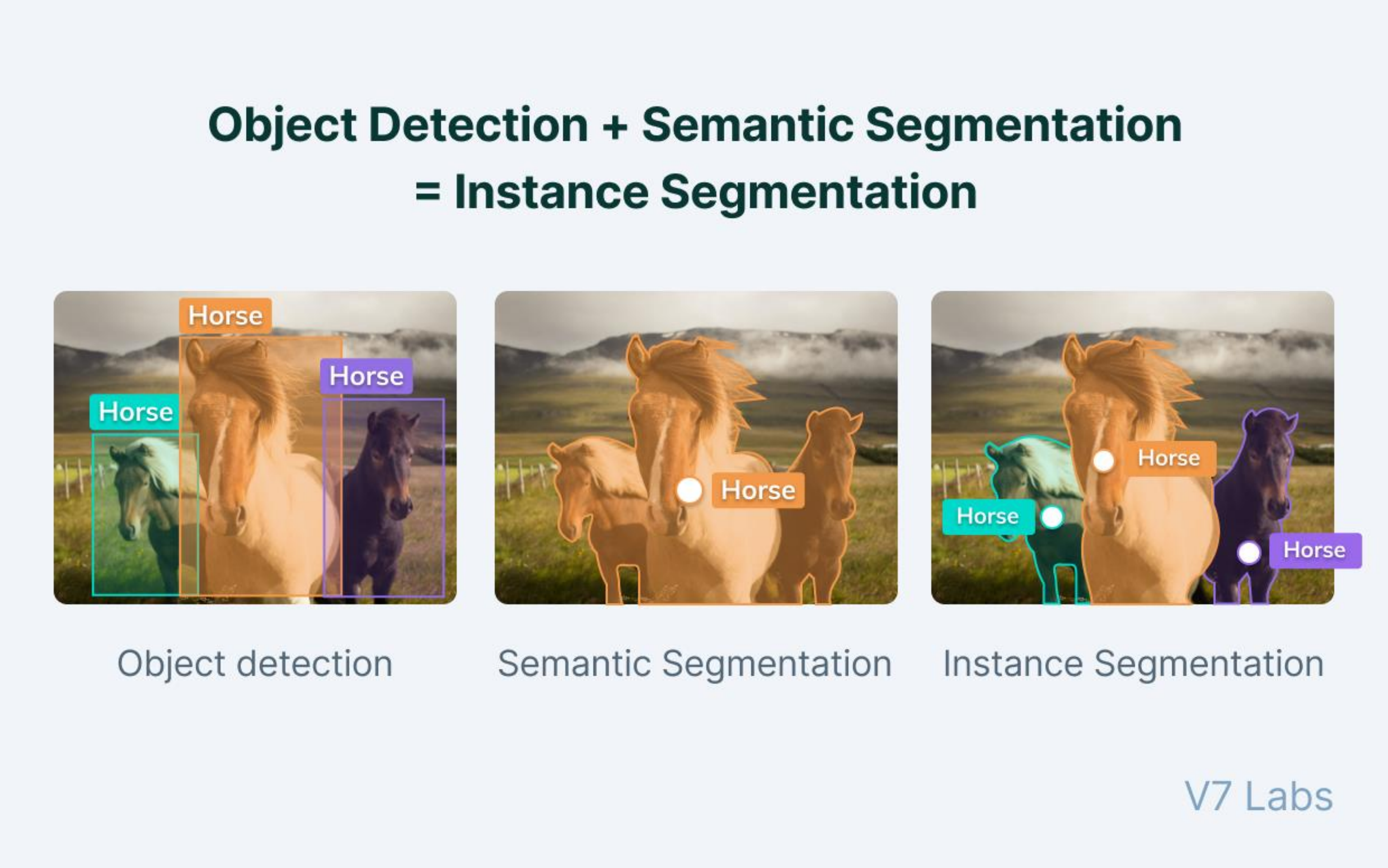
Semantic image segmentation will mark all pixels belonging to that tag, but won't define the boundaries of each object.

Object detection instead will not segment the object, but will clearly define the location of each individual object instance with a box.

Object detection vs image segmentation

Combining semantic segmentation with object detection leads to **instance segmentation,** which first detects the object instances, and then segments each within the detected boxes (known in this case as regions of interest).

Object detection vs image segmentation



Pros and cons of object detection

Object detection is very good at:

- Detecting objects that take up between 2% and 60% of an image's area.
- Detecting objects with clear boundaries.
- Detecting clusters of objects as 1 item.
- Localizing objects at high speed (>15fps)

Pros and cons of object detection

However, it is outclassed by other methods in other scenarios.

You have to always ask yourself: *Do these scenarios apply to my problem?*

Either way, here's a cheat sheet you can use when choosing the right computer vision techniques for your needs.

The right computer vision techniques:

Objects that are elongated—Use Instance Segmentation.

→ Long and thin items such as a pencil will occupy less than 10% of a box's area when detected. This biases model towards background pixels rather than the object itself.

1. What is object detection?

The right computer vision techniques:

Objects that have no physical presence—Use classification

→ Things in an image such as the tag “sunny”, “bright”, or “skewed” are best identified by image classification techniques—letting a network take the image and figure out which feature correlate to these tags.

1. What is object detection?

The right computer vision techniques:

Objects that have no clear boundaries at different angles—Use semantic segmentation

→ The sky, ground, or vegetation in aerial images don’t really have a defined set of boundaries. Semantic segmentation is more efficient at “painting” pixels that belong to these classes. Object detection will still pick up the “sky” as an object, but it will struggle far more with such objects.

The right computer vision techniques:

Objects that are often occluded—Use Instance Segmentation if possible

→ Occlusion is handled far better in two-stage detection networks than one-shot approaches. Within this branch of detectors, instance segmentation models will do a better job at understanding and segmenting occluded objects than mere bounding-box detectors.

2. Types and modes of object detection

Computer vision has made enormous progress in the last couple of decades, and object detection is not the exception.

Mainly it can be divided into two different "eras": Before and After Deep Learning (BDL and ADL).

2. Types and modes of object detection: BDL

Before using **deep learning on object detection (**took off in 2013), the methods were based on **hand-crafted features** and **classical machine learning techniques** (logistic regression, color histograms, or random forests).

These features come from various **algorithms** with information that can be obtained directly from the image.

The methods are sometimes labeled as **Traditional object detectors**.

There are 3 representative examples of this era:

Nguồn: https://www.pento.ai/blog/object-detection

2. Types and modes of object detection: BDL

1. Haar-like features: These were implemented in object-detection research by Viola and Jones. They detected faces based on 3 basic types: edge, line, and four-rectangle features.

ACCEPTED CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION 2001

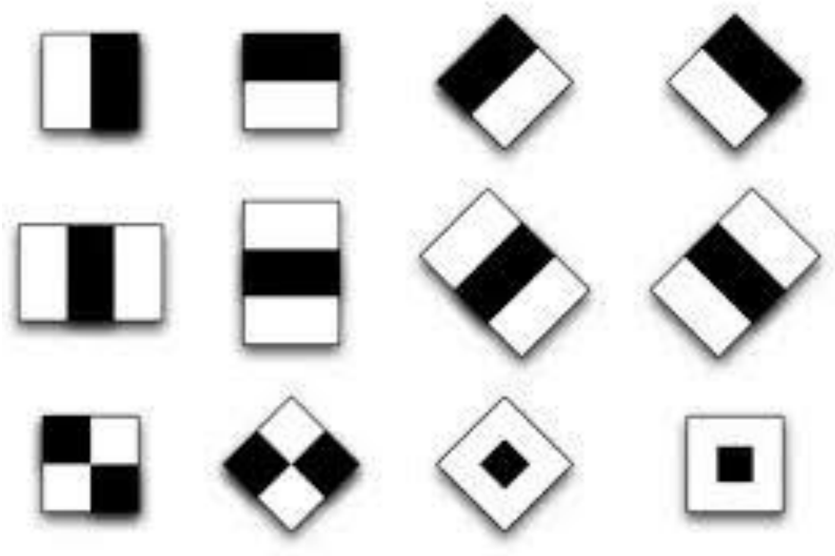
Rapid Object Detection using a Boosted Cascade of Simple Features

Paul Viola
viola@merl.com
Mitsubishi Electric Research Labs
201 Broadway, 8th FL
Cambridge, MA 02139

Michael Jones
mjones@crl.dec.com
Compaq CRL
One Cambridge Center
Cambridge, MA 02142

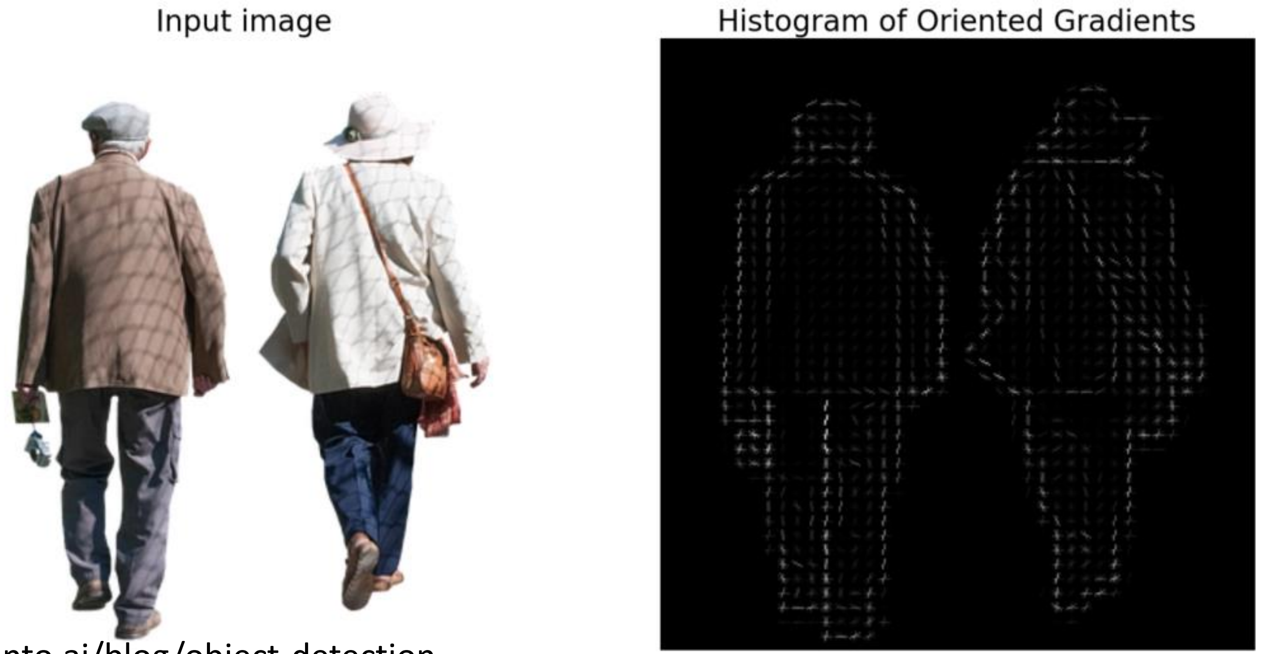
2. Types and modes of object detection: BDL

1. Haar-like features: These were implemented in object-detection research by Viola and Jones. They detected faces based on 3 basic types: edge, line, and four-rectangle features.



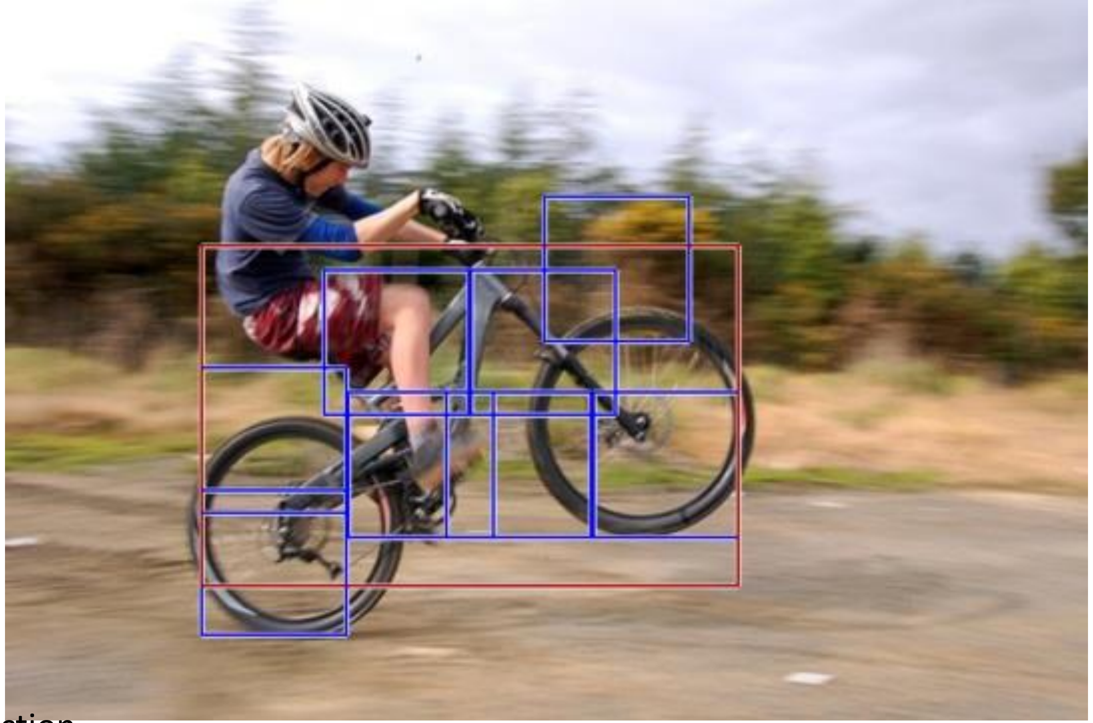
Nguồn: https://www.pento.ai/blog/object-detection

2. HOG Detector: Histograms of Oriented Gradients gained popularity after the Conference on CVPR held in 2005. This method counts how many times a gradient orientation appears in a certain portion of an image



Nguồn: https://www.pento.ai/blog/object-detection

3. DPM: Deformable Parts Model consists of a group of templates arranged in a deformable configuration. It has one global template and many part templates.



Nguồn: https://www.pento.ai/blog/object-detection

2. Types and modes of object detection: BDL

Progress got stuck around 2010 until AlexNet came up in 2012, starting the new **DL** era. This project implemented **CNN**, combined with **data augmentation**, and achieved the lowest error rates to that date.

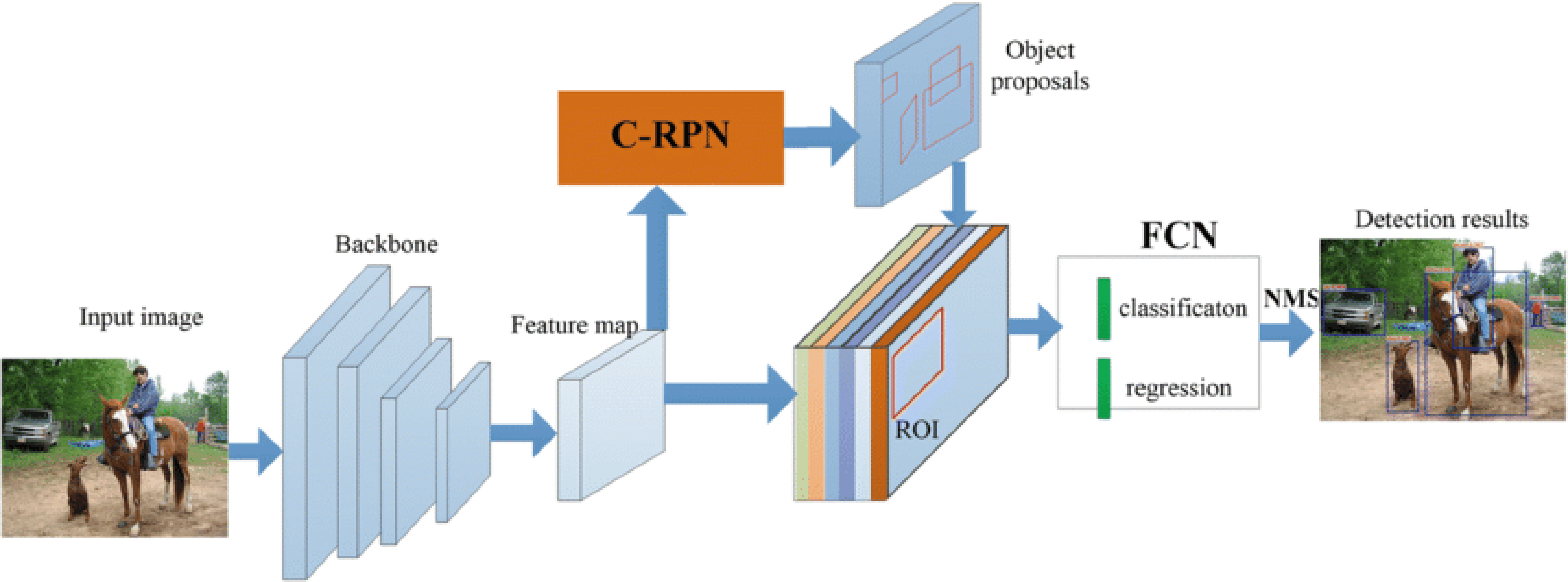
CNNs had been applied to **handwritten recognition**, but there were computational limitations and not large enough databases to scale to **object detection** in a wider range of images. AlexNet tackled this problem.

Nguồn: https://www.pento.ai/blog/object-detection

2. Types and modes of object detection

Today’s deep learning-based techniques vastly outperform these.

Deep learning-based approaches use neural network architectures like RetinaNet, YOLO (You Only Look Once), CenterNet, SSD (Single Shot Multibox detector), Region proposals (R-CNN, Fast-RCNN, Faster RCNN, Cascade R-CNN) for feature detection of the object, and then identification into labels.



3. How does object detection work

3. How does object detection work

All the different models proposed in the last decade can be sorted out into 2 main categories:

1. **Two-stage methods:** Deriving from **R-CNN**, the method consists of a first stage where a model is used to localize possible object regions. Then its results are used as input for a second model which classifies the objects. Most of them build on previous methods and research, focusing on a specific drawback that the previous methods had. Namely SPP-Net, Fast R-CNN, Faster R-CNN, FPN, Mask R-CNN, and Cascade R-CNN.

Nguồn: https://www.pento.ai/blog/object-detection

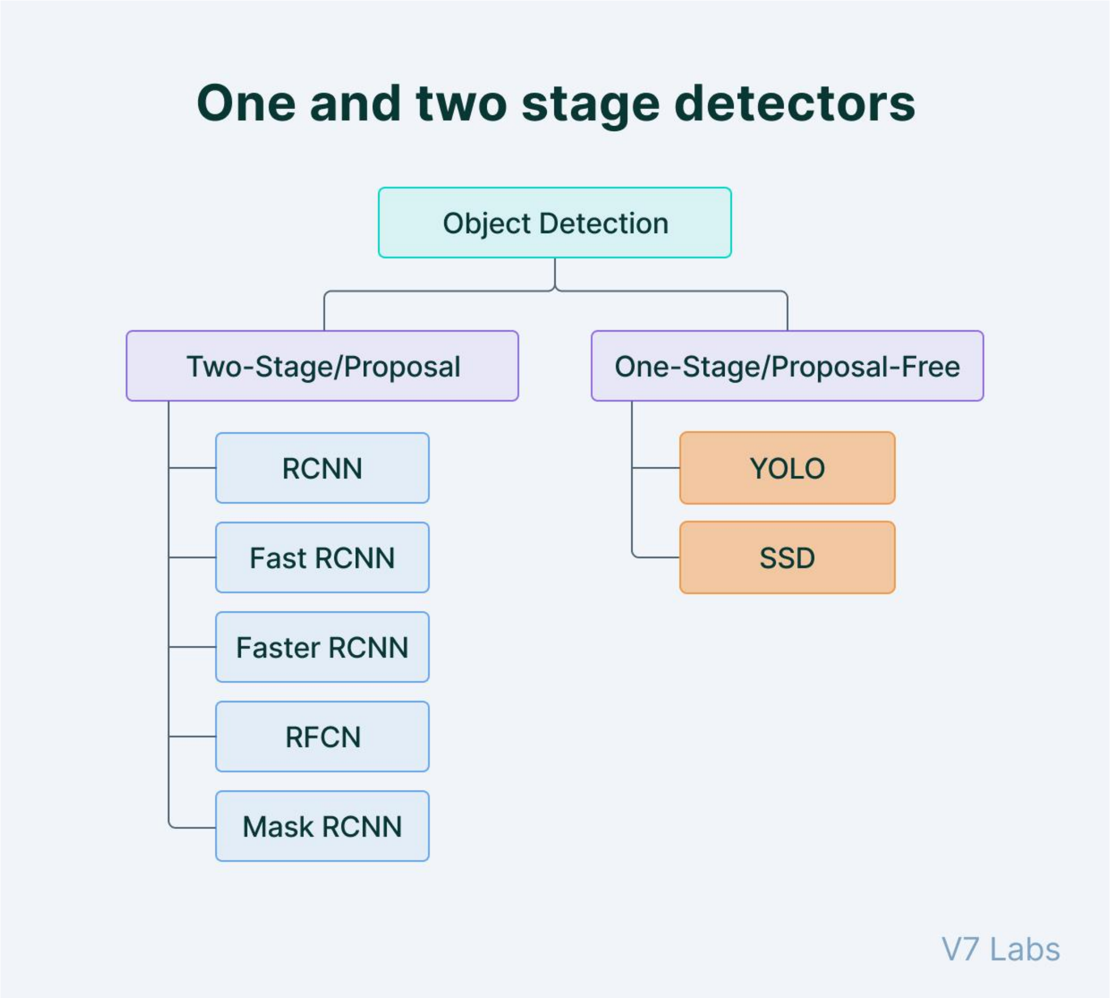
3. How does object detection work

All the different models proposed in the last decade can be sorted out into 2 main categories:

2. One-stage methods: This method directly predicts an **object's bounding boxes** for an image. As an upgrade from the previous two-stage method we discussed, this is faster and simpler but sometimes not as flexible. Examples of this are **SSD** and **YOLO**, which we will add a little demo for you to try on your own.

Nguồn: https://www.pento.ai/blog/object-detection

3. How does object detection work



Nguồn: https://www.v7labs.com/blog/object-detection-guide

3. How does object detection work

State of the art object detection architectures consists of 2 stage architectures, many of which have been pre-trained on the **COCO dataset**.

COCO is an image dataset composed of 90 different classes of objects (cars, persons, sport balls, bicycles, dogs, cats, horses e.t.c).

3. How does object detection work

The dataset was gathered to solve common object detection problems. Nowadays it is becoming outdated as its images were captured mostly in the early 2,000’s making them much smaller, grainier, and with different objects than today’s images. Newer datasets like **OpenImages** are taking its spot as the de-facto pre-training dataset.

Single-stage object detectors

A single-stage detector removes the RoI extraction process and directly classifies and regresses the candidate anchor boxes. Examples are: YOLO family (YOLOv2, YOLOv3, YOLOv4, and YOLOv5, …. YOLOv10) CornerNet, CenterNet, and others. For instance, let’s take a look at how YOLO Works.

YOLOv10 - Ultralytics YOLO Docs

docs.ultralytics.com/models/yolov10/#citations-and-acknowledgements

ultralytics YOLOv10 Search ultralytics/ultralytics v8.2.0 24.7k 4.9k

Home Quickstart Modes Tasks Models Datasets NEW Solutions Guides Integrations HUB Reference Help

Models

- YOLOv3
- YOLOv4
- YOLOv5
- YOLOv6
- YOLOv7
- YOLOv8
- YOLOv9
- YOLOv10
- SAM (Segment Anything Model)
- MobileSAM (Mobile Segment Anything Model)
- FastSAM (Fast Segment Anything Model)
- YOLO-NAS (Neural Architecture Search)
- RT-DETR (Realtime Detection Transformer)
- YOLO-World (Real-Time Open-Vocabulary Object Detection)

Citations and Acknowledgements

Back to top

We would like to acknowledge the YOLOv10 authors from Tsinghua University for their extensive research and significant contributions to the Ultralytics framework:

BibTeX

```
@article{THU-MIGyolov10,
  title={YOLOv10: Real-Time End-to-End Object Detection},
  author={Ao Wang, Hui Chen, Lihao Liu, et al.},
  journal={arXiv preprint arXiv:2405.14458},
  year={2024},
  institution={Tsinghua University},
  license = {AGPL-3.0}
}
```

For detailed implementation, architectural innovations, and experimental results, please refer to the YOLOv10 research paper and GitHub repository by the Tsinghua University team.

Created 2024-05-25, Updated 2024-06-02

Authors: glenn-jocher (2), RizwanMunawar (2)

Tweet Share

Table of contents

EN

Two-stage object detectors

Two-stage detectors divide the object detection task into two stages: extract RoIs (Region of interest), then classify and regress the RoIs. Examples of object detection architectures that are 2 stage oriented include R-CNN, Fast-RCNN, Faster-RCNN, Mask-RCNN and others. Let’s take a look at the Mask R-CNN for instance.

R-CNN Model Family

The R-CNN Model family includes the following:

R-CNN—This utilizes a selective search method to locate RoIs in the input images and uses a DCN (Deep Convolutional Neural Network)-based region wise classifier to classify the RoIs independently.

SPPNet and Fast R-CNN—This is an improved version of R-CNN that deals with the extraction of the RoIs from the feature maps. This was found to be much faster than the conventional R-CNN architecture.

R-CNN Model Family

The R-CNN Model family includes the following:

Faster R-CNN—This is an improved version of Fast R-CNN that was trained end to end by introducing RPN (region proposal network). An RPN is a network utilized in generating RoIs by regressing the anchor boxes. Hence, the anchor boxes are then used in the object detection task.

Mask R-CNN adds a mask prediction branch on the Faster R-CNN, which can detect objects and predict their masks at the same time.

R-CNN Model Family

The R-CNN Model family includes the following:

R-FCN replaces the fully connected layers with the position-sensitive score maps for better detecting objects.

Cascade R-CNN addresses the problem of overfitting at training and quality mismatch at inference by training a sequence of detectors with increasing IoU thresholds.

YOLO Model Family

The YOLO family model includes the following:

YOLO uses fewer anchor boxes (divide the input image into an $S \times S$ grid) to do regression and classification. This was built using darknet neural networks.

YOLOv2 improves the performance by using more anchor boxes and a new bounding box regression method..

YOLO Model Family

The YOLO family model includes the following:

YOLOv3 is an enhanced version of the v2 variant with a deeper feature detector network and minor representational changes. YOLOv3 has relatively speedy inference times with it taking roughly 30ms per inference.

YOLOv4 (YOLOv3 upgrade) works by breaking the object detection task into two pieces, regression to identify object positioning via bounding boxes and classification to determine the object's class. YOLO V4 and its successors are technically the product of a different set of researchers than versions 1-3.

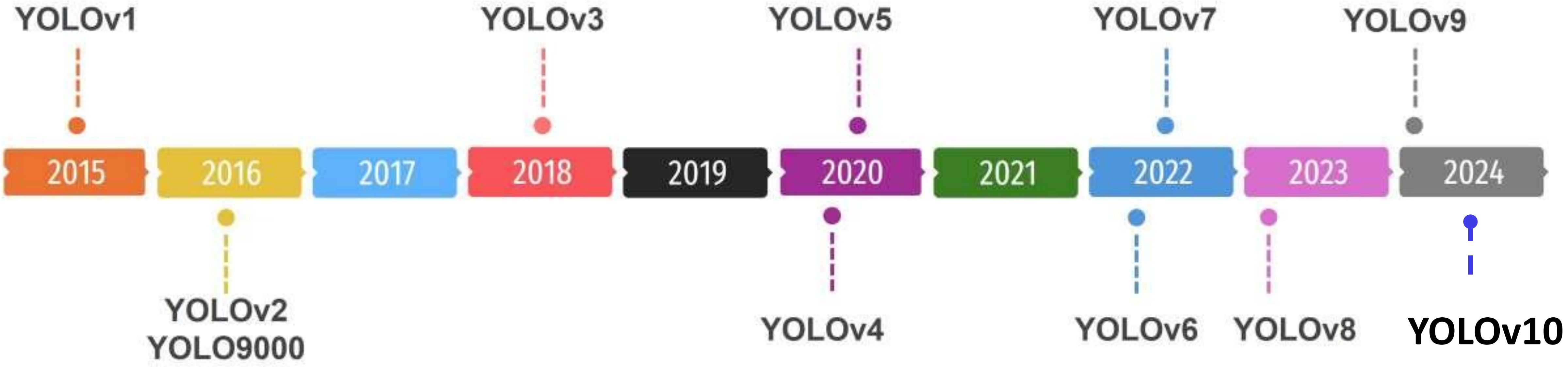
YOLO Model Family

The YOLO family model includes the following:

YOLOv5 is an improved version of YOLOv4 with a mosaic augmentation technique for increasing the general performance of YOLOv4.

YOLOv6 – 7 – X ,…

YOLO Model Family



https://www.v7labs.com/blog/yolo-object-detection

CenterNet Family

The CenterNet family model includes the following:

- SSD places anchor boxes densely over an input image and uses features from different convolutional layers to regress and classify the anchor boxes.
- DSSD introduces a deconvolution module into SSD to combine low level and high-level features. While R-SSD uses pooling and deconvolution operations in different feature layers to combine low-level and high-level features.
- RON proposes a reverse connection and an objectness prior to extracting multiscale features effectively.
- RefineDet refines the locations and sizes of the anchor boxes for two times, which inherits the merits of both one-stage and two-stage approaches.
- CornerNet is another keypoint-based approach, which directly detects an object using a pair of corners. Although CornerNet achieves high performance, it still has more room to improve.
- CenterNet explores the visual patterns within each bounding box. For detecting an object, this uses a triplet, rather than a pair, of keypoints. CenterNet evaluates objects as single points by predicting the x and y coordinate of the object’s center and it’s area of coverage (width and height). It is a unique technique that has proven to out-perform variants like the SSD and R-CNN family.

CenterNet Family

The CenterNet family model includes the following:

SSD places anchor boxes densely over an input image and uses features from different convolutional layers to regress and classify the anchor boxes.

DSSD introduces a deconvolution module into SSD to combine low level and high-level features. While R-SSD uses pooling and deconvolution operations in different feature layers to combine low-level and high-level features.

CenterNet Family

The CenterNet family model includes the following:

RON proposes a reverse connection and an objectness prior to extracting multiscale features effectively.

RefineDet refines the locations and sizes of the anchor boxes for two times, which inherits the merits of both one-stage and two-stage approaches.

CornerNet is another keypoint-based approach, which directly detects an object using a pair of corners. Although CornerNet achieves high performance, it still has more room to improve.

CenterNet Family

The CenterNet family model includes the following:

CenterNet explores the visual patterns within each bounding box. For detecting an object, this uses a triplet, rather than a pair, of keypoints. CenterNet evaluates objects as single points by predicting the x and y coordinate of the object’s center and it’s area of coverage (width and height). It is a unique technique that has proven to out-perform variants like the SSD and R-CNN family.

Face and person detection

Most face recognition systems are powered by object detection. It can be used to detect faces, classify emotions or expressions, and feed the resulting box to an image-retrieval system to identify a specific person out of a group.

Face detection is one of the most popular object detection use cases, and you are probably already using it whenever you unlock your phone with your face.

Person detection is also commonly used to count the number of people in retail stores or ensure social distancing metrics.

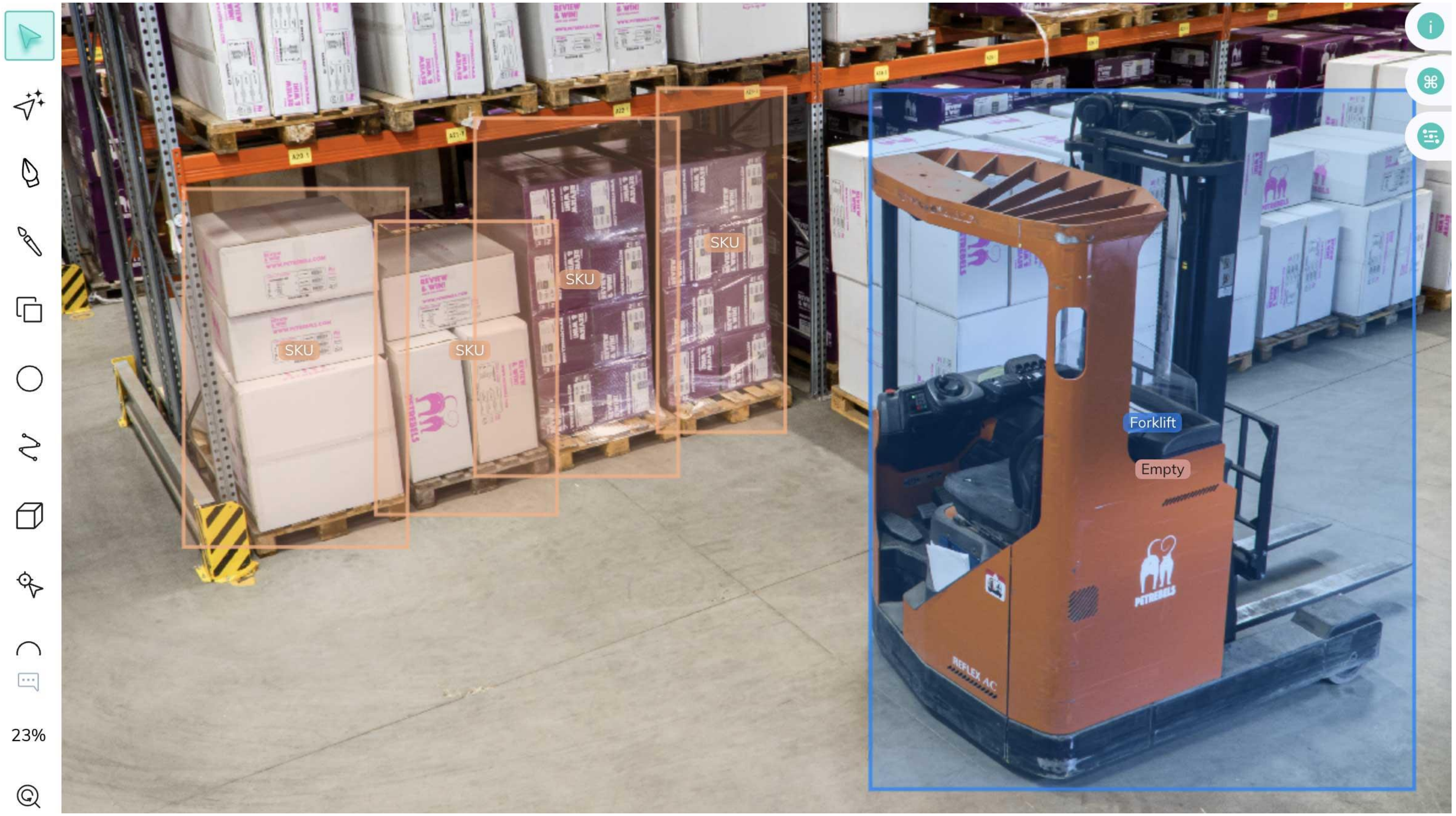
Face and person detection

5. Object detection applications

Intelligent video analytics

Object detection is used in intelligent video analytics (IVA) anywhere CCTV cameras are present in retail venues to understand how shoppers are interacting with products. These video streams pass through an anonymizaion pipeline to blur out people's faces and de-identify individuals. Some IVA use cases preserve privacy by only looking at people's shoes, by placing cameras below knee level and ensuring the system captures the presence of a person, without having to directly look at their identifiable features. IVA is often used in factories, airports and transport hubs to track queue lengths and access to restricted areas.

Intelligent video analytics



Autonomous vehicles

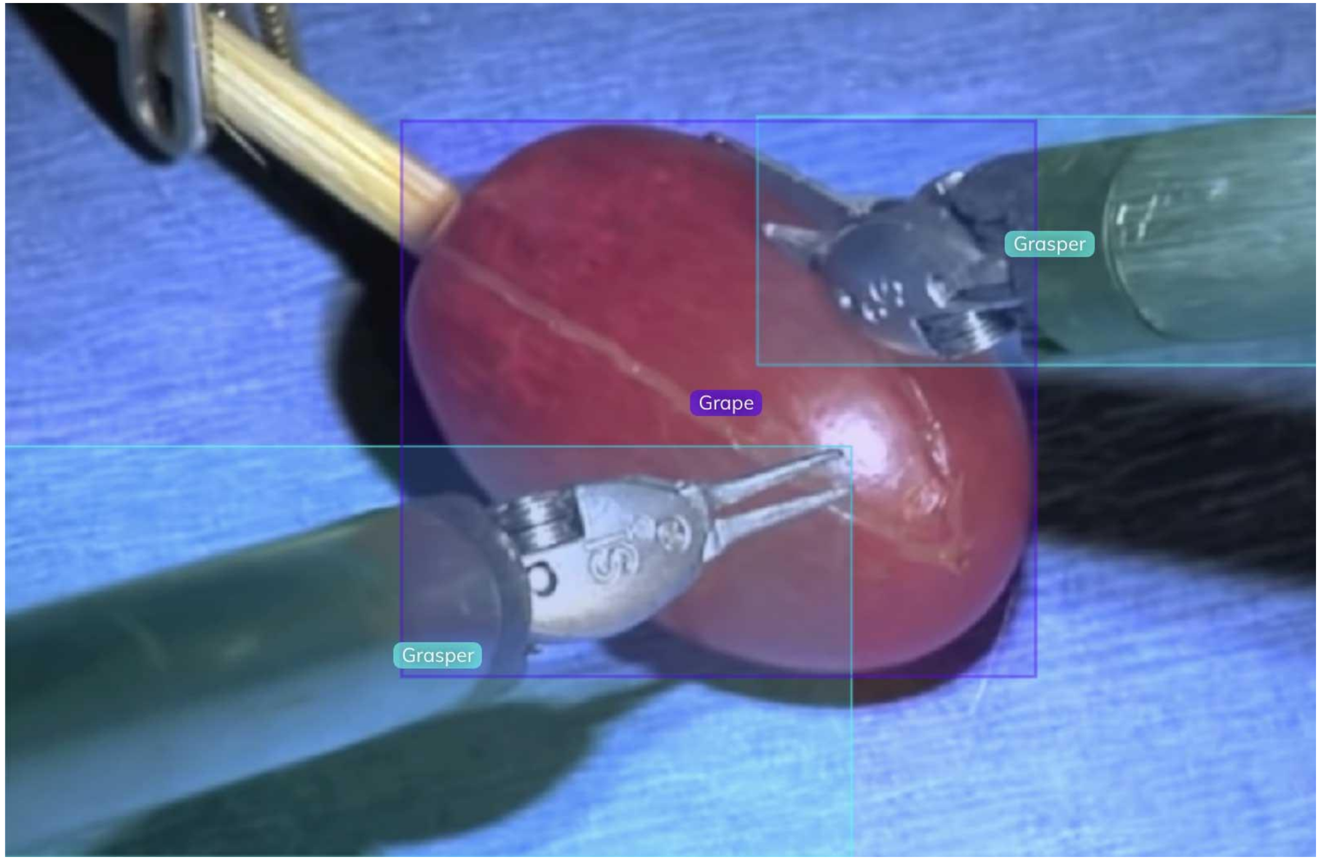
Self-driving cars use object detection to spot pedestrians, other cars, and obstacles on the road in order to move around safely. Autonomous vehicles equipped with LIDAR will sometimes use 3D object detection, which applies cuboids around objects.

Autonomous vehicles

Intelligence video surgery

Surgical video is very noisy data that is taken from endoscopes during crucial operations. Object detection can be used to spot hard-to-see items such as polyps or lesions that require a surgeon's immediate attention. It's also being used to inform hospital staff of the status of the operation.

Intelligence video surgery



Defect Inspection

Manufacturing companies can use object detection to spot defects in the production line. Neural networks can be trained to detect minute defects, from folds in fabric to dents or flashes in injection molded plastics.

Unlike traditional machine learning approaches, deep learning-based object detection can also spot defects in heavily varying objects, such as food.

5. Object detection applications

Defect Inspection

5. Object detection applications

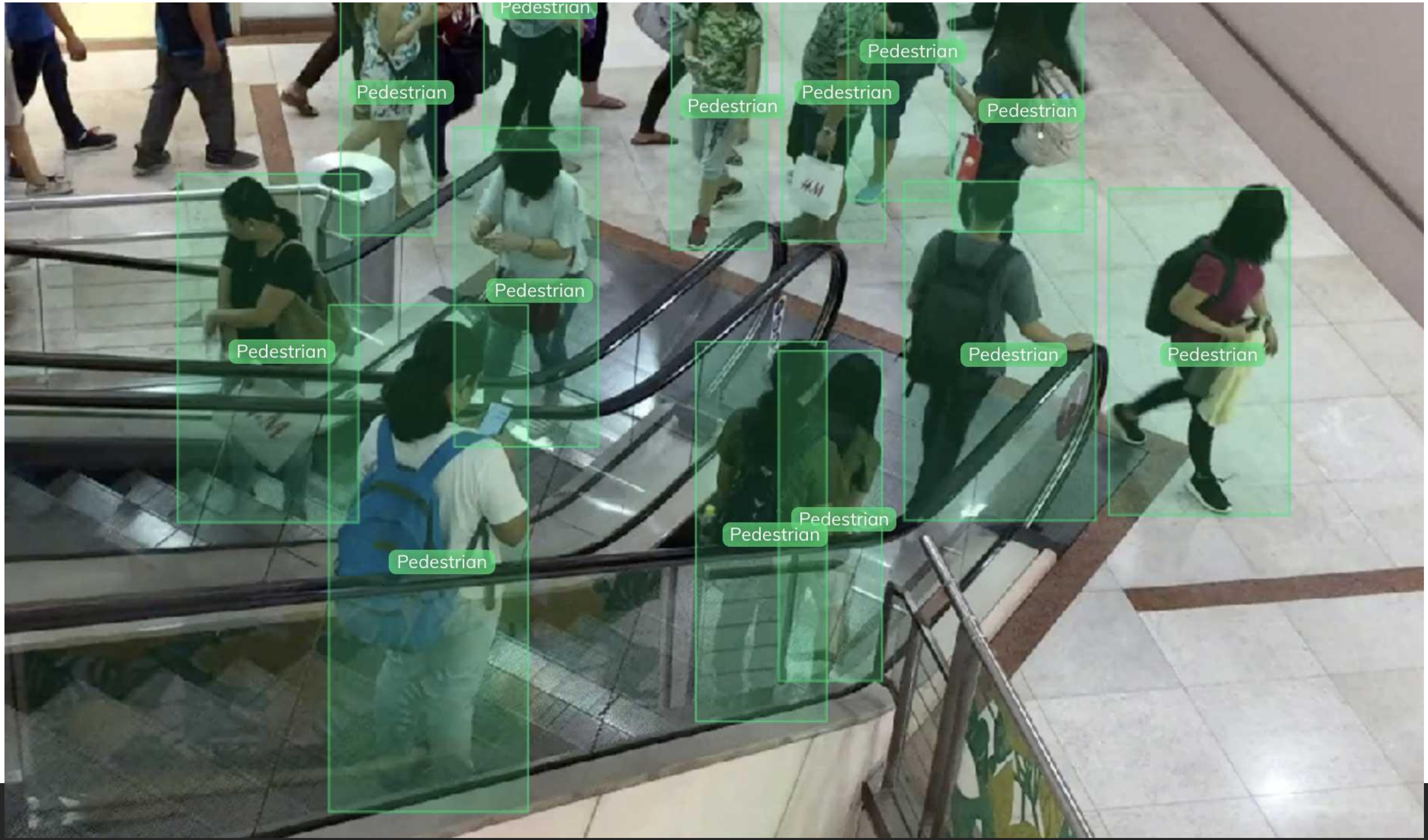
Pedestrian detection

It is one of the most essential computer vision tasks that is applied in robotics, video surveillance, and automotive safety. Pedestrian detection plays a key role in object detection research as it provides the fundamental information for the semantic understanding of video footages.

However—

Despite its relatively high performance, this technology still faces challenges such as various styles of clothing in appearance or the presence of occluding accessories that decrease the accuracy of the existing detectors.

Pedestrian detection

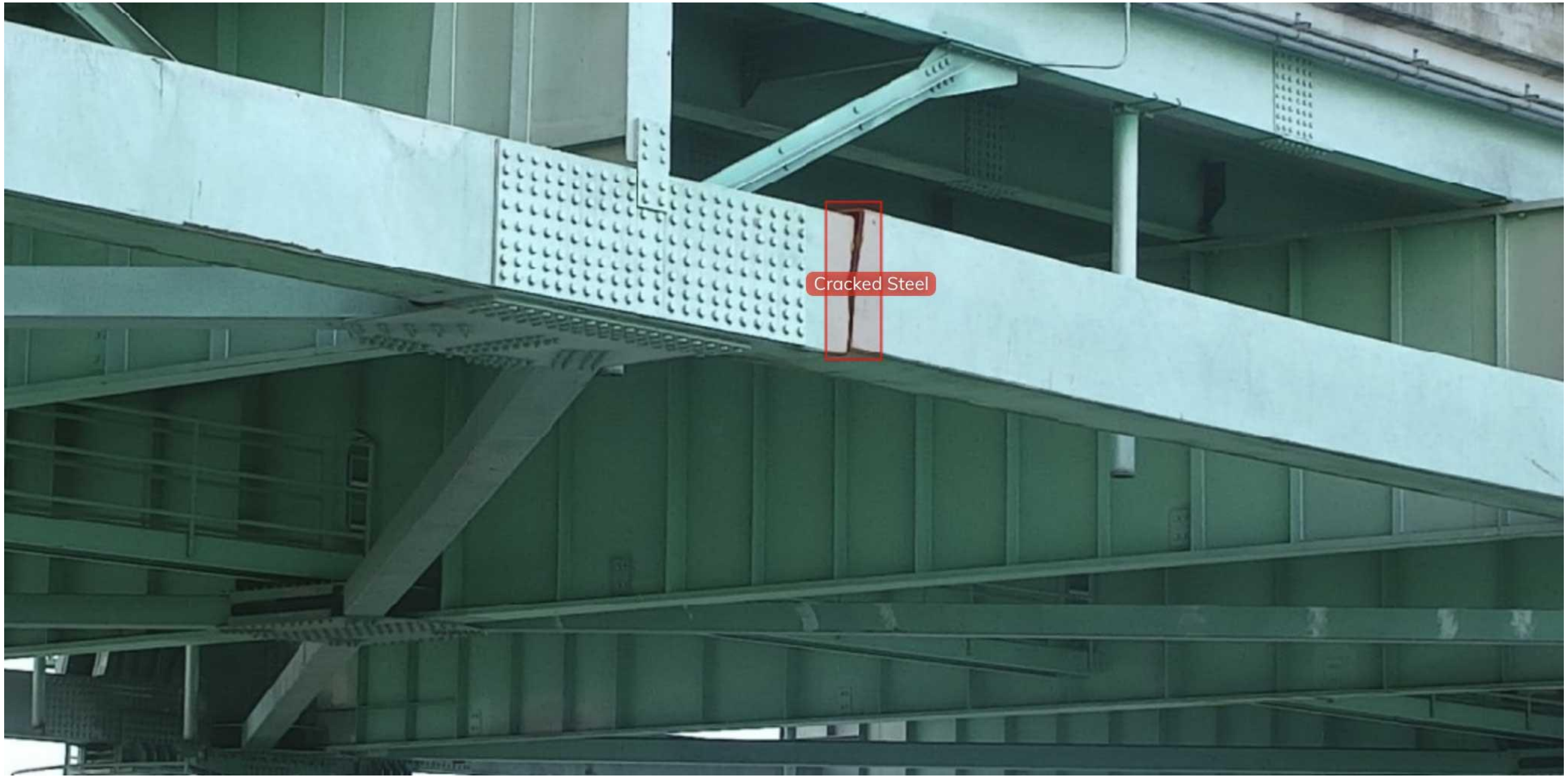


AI Drone Navigation

Drones sport incredible cameras nowadays and can leverage models hosted in the cloud to assess any object they encounter.

For example, they can be used to inspect hard-to-reach areas in bridges for cracks and other structural damage or to inspect power lines, replacing dangerous routine helicopter operations.

AI Drone Navigation



Tài liệu tham khảo

Nguồn: https://www.pento.ai/blog/object-detection

https://www.v7labs.com/blog/object-detection-guide

https://www.tensorflow.org/hub/tutorials/object_detection

https://www.datacamp.com/tutorial/object-detection-guide

https://blog.tensorflow.org/2021/06/easier-object-detection-on-mobile-with-tf-lite.html