DRIVES

Development and Research on Innovative Vocational Education Skills

COMPUTER VISION FUNDAMENTALS

U2.E2. STATE-OF-THE-ART VISION TECHNIQUES

Computer Vision Expert

May 2021, Version 1

Co-funded by the Erasmus+ Programme of the European Union

The Development and Research on Innovative Vocational Educational Skills project (DRIVES) is co-funded by the Erasmus+ Programme of the European Union under the agreement 591988-EPP-1-2017-1-CZ-EPPKA2-SSA-B. The European Commission support for the production of this publication does not constitute endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

LEARNING OBJECTIVES

The student is able to …

CVE.U2.E2.PC1	Understand what Computer Vision techniques are and what are their purposes.
CVE.U2.E2.PC2	List and understand the most used state-of-the-art vision techniques.

COMPUTER VISION TECHNIQUES: IMAGE CLASSIFICATION - DEFINITION

DRIVES
Development and Research on Innovative Vocational Education Skills

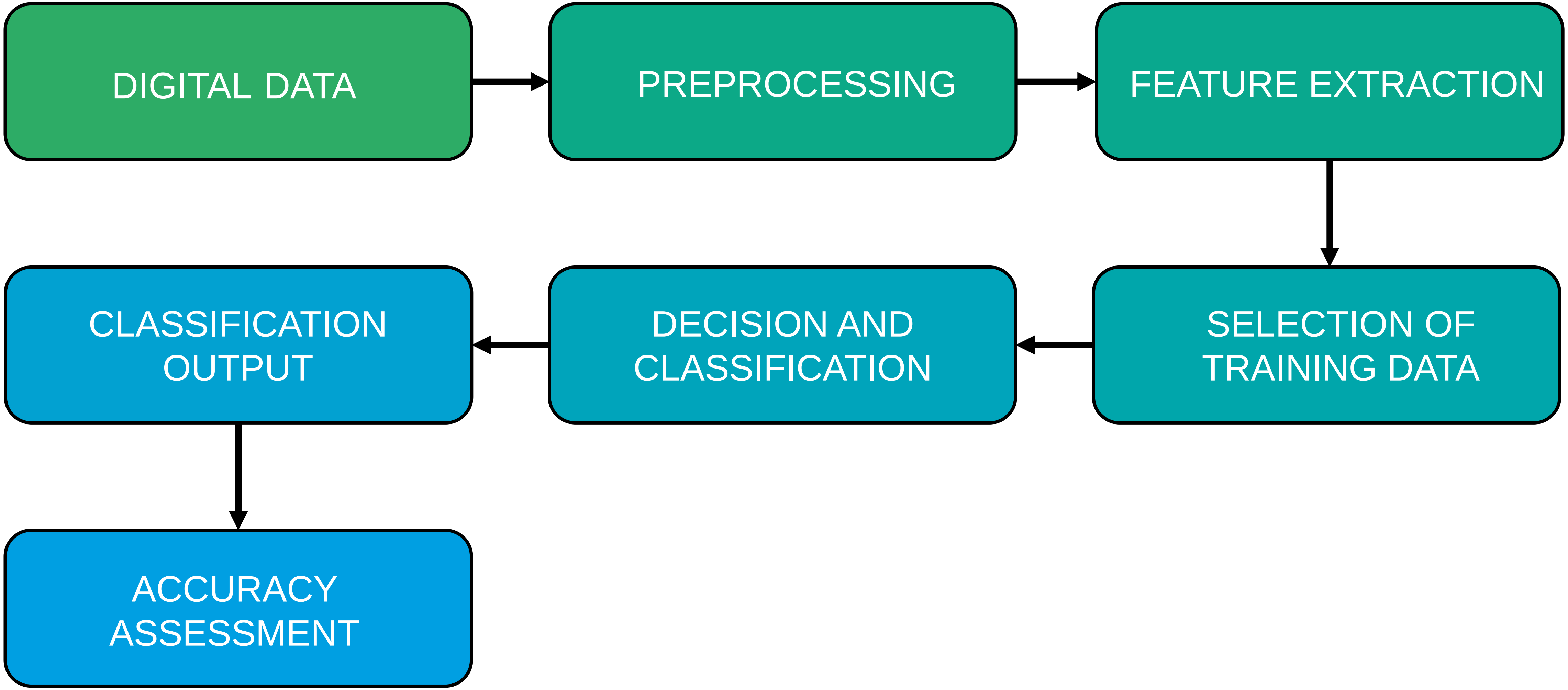
Technique used for detecting, analyzing, and interpreting images.

Works through a neural network trained via an annotated dataset.

Accepts the given input images and produces output ‘class’ in which the image falls under.

Occurs when machines can look at an image and assign a (correct) label to it.

It is a key part of computer vision, allowing computers to see the world as we do.


DIGITAL DATA
PREPROCESSING
FEATURE EXTRACTION
SELECTION OF TRAINING DATA
DECISION AND CLASSIFICATION
CLASSIFICATION OUTPUT
ACCURACY ASSESSMENT

1

DIGITAL DATA

Image is captured by using digital camera or any other mobile phone camera

In information theory and information systems, is the discrete, discontinuous representation of information or works. Numbers and letters are commonly used representations.

2

PREPROCESSING

Improvement of the image data.

3

FEATURE EXTRACTION

The process of measuring or calculating or detecting the features from the image samples

The most common two types of feature extraction are:

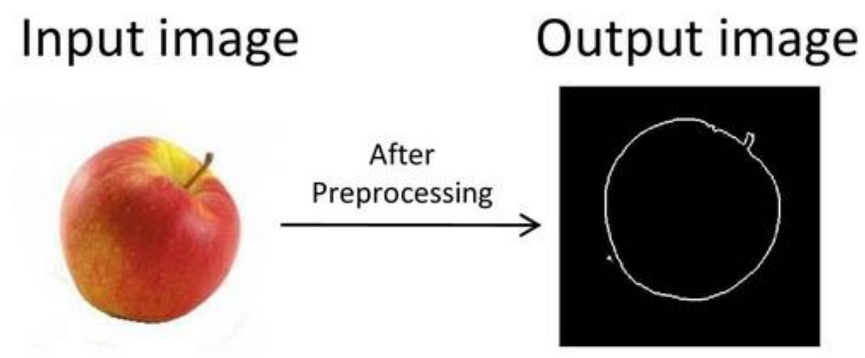
- Geometric feature extraction;
- Color feature extraction

DRIVES
Development and Research on Innovative Vocational Education Skills

4

SELECTION OF TRAINING DATA

Selection of the attributes that best describe the pattern



Training Data Set

NAME	MAJOR	MINOR	AREA	PERIMETER	RED#	YELLOW#
‘Ap.jpg’	175.4774	169.6791	23328	604.9848	1267	621

5

DECISION AND CLASSIFICATION

Categorizes detected objects into predefined classes by using suitable method that compares the image patterns with the target patterns.

6 CLASSIFICATION OUTPUT

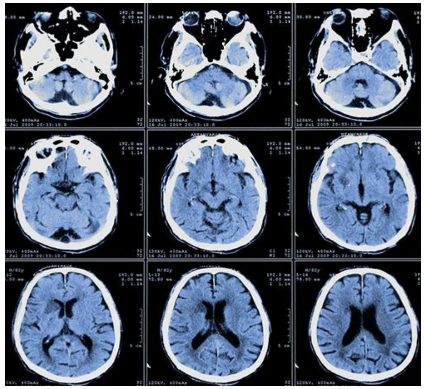
SAMPLE IMAGE						
AREA OR SIZE	1200000	900000	800000	1100000	600000	700000
COLOUR INTENSITY	80	90	100	93	96	85
GRADING	A	B	B	A	C	C

7

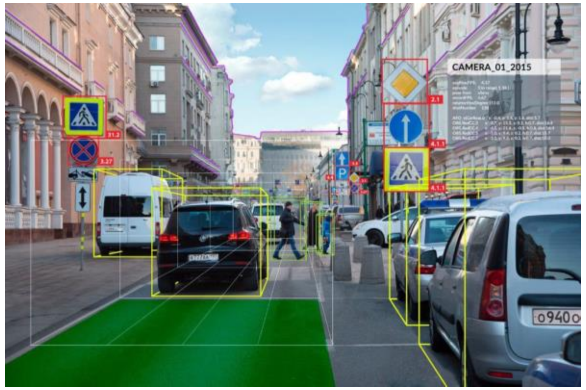
ACCURACY ASSESSMENT

The accuracy assessment is done to identify possible sources of errors as well as indicators used in comparisons.

COMPUTER VISION TECHNIQUES: IMAGE CLASSIFICATION – APPLICATIONS

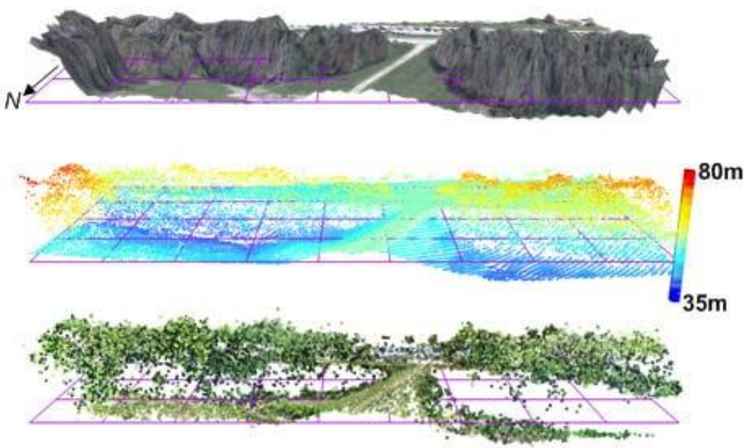


BIOMEDICAL IMAGING



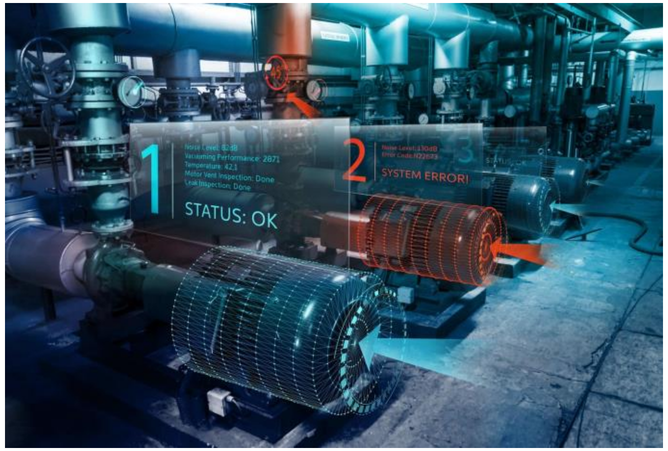
VEHICAL NAVIGATION

BIOMETRY



REMOTE SENSING

ROBOT NAVIGATION



INDUSTRIAL VISUAL INSPECTION

Object detection can be explained as the technique that identifies and locates objects within an image or video. Locate the presence of objects with a bounding box and types or classes of the located objects in an image.

- ***Input*:** An image with one or more objects, such as a photograph.
- ***Output*:** One or more bounding boxes (e.g. defined by a point, width, and height), and a class label for each bounding box.

Algorithms produce a list of object categories present in the image along with an axis-aligned bounding box indicating the position and scale of every instance of each object category.

⬇

Generally these algorithms fall into **machine-learning** approaches or **deep-learning** approaches.

Machine Learning approaches:

- Viola–Jones object detection framework based on Haar features
- Scale-invariant feature transform (SIFT)
- Histogram of oriented gradients (HOG) features

Deep Learning approaches:

- Region Proposals (R-CNN, Fast R-CNN, Faster R-CNN, cascade R-CNN)
- Single Shot MultiBox Detector (SSD)
- You Only Look Once (YOLO)
- Single-Shot Refinement Neural Network for Object Detection (RefineDet)
- Retina-Net
- Deformable convolutional networks

Why is it important?

Object detection is linked to image recognition and image segmentation, which are similar to computer vision techniques. These techniques help us to understand and to analyze scenes in images or video.

But there are some differences!

Image recognition only outputs a class label for an identified object, and image segmentation creates a pixel-level understanding of a scene’s elements. What separates object detection from these other tasks is its ability to *locate objects within an image or video*. This allows us to count and then track those objects.

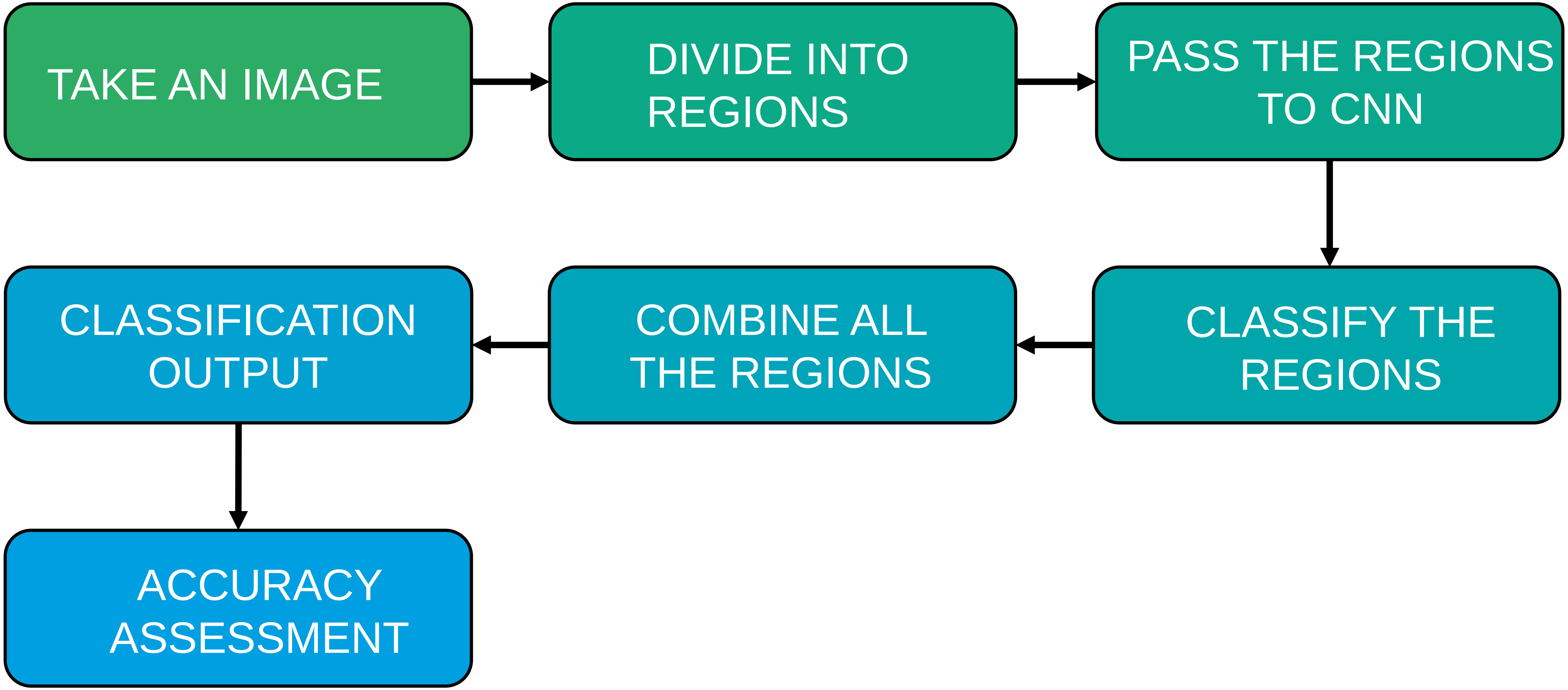
COMPUTER VISION TECHNIQUES: OBJECT DETECTION – WHY IT IS IMPORTANT?

DRIVES
Development and Research on Innovative Vocational Education Skills

Given these key distinctions and object detection's unique capabilities, we can see how it can be applied in a number of ways:

- Crowd counting
- Self-driving cars
- Video surveillance
- Face detection
- Anomaly detection

COMPUTER VISION TECHNIQUES: OBJECT DETECTION– STEP BY STEP



Object tracking aims to track objects as they move across a series of video frames. Generally the objects are people, but may also be animals, vehicles or other objects of interest, such as the ball in a game of soccer.

This can be applied in many areas, such as surveillance, medical imaging, traffic flow analysis, self driving cars, people counting, audience flow analysis, and human-computer interaction.

This technique starts with object detection, where the objects are identified in an image and assign them bounding boxes. The algorithm is responsible to assign an ID to each object identified in the image, and in subsequent frames tries to carry across this ID and identify the new position of the same object.

There are two main types of object tracking:

- **Offline object tracking**—object tracking on a recorded video where all the frames, including future activity, are known in advance.

- **Online object tracking**—object tracking done on a live video stream, such as, a surveillance camera. This is more challenging because the algorithm must work fast, and it is not possible to take future frames and combine them into the analysis.

Challenges of object tracking compared to static object detection:

- **Re-identification:** connecting an object in one frame to the same object in the subsequent frames.
- **Appearance and disappearance:** objects can move into or out of the frame unpredictably and we need to connect them to objects previously seen in the video.
- **Occlusion:** objects are partially or completely occluded in some frames, as other objects appear in front of them and cover them up.
- **Identity switches:** when two objects cross each other, we need to discern which one is which.

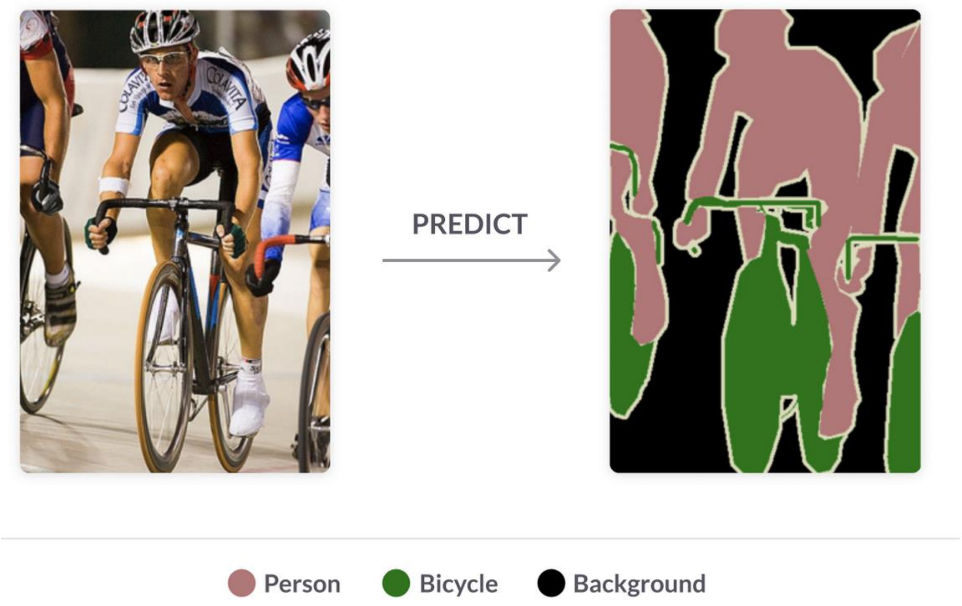
- **Motion blur:** objects may look different due to their own motion or camera motion.
- **View points:** objects may look very different from different viewpoints, and we have to consistently identify the same object from all perspectives.
- **Scale change:** objects in a video can change scale dramatically due to camera zoom for example.
- **Illumination:** lighting changes in a video can have a big effect on how objects look and can make it harder to consistently detect them.

Simple Online and Real-Time Tracking (SORT), this type of algorithm relies mainly on the analysis of an underlying object detection engine. It can plug into any object detection algorithm. The algorithm tracks multiple objects in real time, associating the objects in each frame with those detected in previous frames using simple heuristics.

Generic Object Tracking Using Regression Network (GOTURN) is trained by comparing pairs of cropped frames from thousands of video sequences. In the first frame (“previous frame”), the location of the object is known, and the frame is cropped to twice the size of the bounding box around the object, with the object centered.

Multi-Domain Network (MDNet) is a CNN architecture that won the VOT2015 challenge. The objective of MDNet is to speed up training in order to provide real-time results. The strategy is to split the network into two parts. The first part acts as a generic feature extractor that trains over multiple training sets and learns to distinguish objects from their background. The second part is trained on a specific training set and learns to identify objects within video frames.

This technique classifies all the pixels of an image into meaningful classes of objects. These classes are “semantically interpretable” and correspond to real-world categories. For instance, you could isolate all the pixels associated with a cat and color them green. This is also known as dense prediction because it predicts the meaning of each pixel.



This technique identifies each instance of each object in an image. It is distinct from semantic segmentation in that it does not classify every pixel. If an image contains three cars, semantic segmentation classifies all of them as one instance, whereas instance segmentation identifies each individual car.

Image Segmentation Methods

Other image segmentation techniques that were popular in the past are less efficient than deep learning counterparts because they use rigid algorithms and require human intervention and expertise. These are some examples:

- **Thresholding;**
- **K- means clustering;**
- **Histogram-based image segmentation;**
- **Edge detection.**

- **Thresholding:** divides an image into a foreground and background. A specified threshold value separatespixels into one of two levels to isolate objects. Thresholding converts grayscale images into binary images or distinguishes the lighter and darker pixels of a color image.

- **K- means clustering:** An algorithm identifies groups in the data, with the variable K representing the number of groups. The algorithm assigns each data point (or pixel) to one of the groups based on feature similarity. Rather than analyzing predefined groups, clustering works iteratively to form groups organically.

- **Histogram-based image segmentation:** Uses a histogram to group pixels based on “gray levels”. Simple images consist of an object and a background. The background is usually one gray level and is the larger entity. Thus, a large peak represents the background gray level in the histogram. A smaller peak represents the object, which is another gray level.

- **Edge detection:** Identifies sharp changes or discontinuities in brightness. Edge detection usually involves arranging points of discontinuit into curved line segments, or edges. For example, the border between a block of red and a block of blue.

COMPUTER VISION TECHNIQUES: OTHERS....

Other vision techniques are:

- Style transfer
- Colorization
- Action recognition
- 3D objects
- Human pose estimation

SUMMARY

- Image classification is a technique used for detecting, analyzing, and interpreting images.
- Image classification has 7 steps.
- Object detection can be explained as the technique that identify and locate objects within na image or video.
- The two main types of object tracking are Offline object tracking and Online object tracking.
- Object tracking algorithms are Simple Online and Real-Time Tracking (SORT), Generic Object Tracking Using Regression Network (GOTURN) and Multi-Domain Network (MDNet).
- Semantic Segmentation classifies all the pixels of an image into meaningful classes of objects.
- Image Segmentation Methods are Thresholding, K- means clustering, Histogram based image segmentation, and Edge detection.
- Other vision techniques are Style transfer, Colorization, Action recognition, 3D objects, and Human pose estimation.

Biomedical imaging. (n.d.). Retrieved from https://ingenieriabiologicaymedica.uc.cl/en/research/research-areas/biomedical-imaging

Biometrics and Computer Vision Merge to Create Limitless Possibilities |. (n.d.). Retrieved from https://www.analyticsinsight.net/biometrics-and-computer-vision-merge-to-create-limitless-possibilities/

Computer Vision applications in Self-Driving Cars | by Jeremy Cohen | Becoming Human: Artificial Intelligence Magazine. (n.d.). Retrieved from https://becominghuman.ai/computer-vision-applications-in-self-driving-cars-610561e14118

Dandois, J. P., & Ellis, E. C. (2010). Remote sensing of vegetation structure using computer vision. *Remote Sensing*, *2*(4), 1157–1176. https://doi.org/10.3390/rs2041157

Everything you need to know about Visual Inspection with AI. (n.d.). Retrieved from https://nanonets.com/blog/ai-visual-inspection/

Robot Navigation using Stereo Vision. (n.d.). Retrieved from https://boredomprojects.net/index.php/projects/robot-navigation-using-stereo-vision

REFERENCE TO AUTHORS

Regina Sousa

- PhD student in Biomedical Engineering
- Research Collaborator of the Algoritmi Research Center

0000-0002-2988-196X

Diana Ferreira

- PhD student in Biomedical Engineering
- Research Collaborator of the Algoritmi Research Center

0000-0003-2326-2153

José Machado

- Associate Professor with Habilitation at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

0000-0003-4121-6169

António Abelha

- Assistant Professor at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

0000-0001-6457-0756

Victor Alves

- Assistant Professor at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

0000-0003-1819-7051

REFERENCE TO AUTHORS

This Training Material has been certified according to the rules of **ECQA – European Certification and Qualification Association.**

The Training Material was developed within the international job role committee **“Computer Vision Expert”:**

UMINHO – University of Minho (https://www.uminho.pt/PT)

The development of the training material was partly funded by the EU under Blueprint Project DRIVES.

DRIVES

Development and Research on Innovative Vocational Education Skills

Thank you for your attention

DRIVES project is project under **The Blueprint for Sectoral Cooperation on Skills in Automotive Sector**, as part of New Skills Agenda.

The aim of the Blueprint is **to support an overall sectoral strategy and to develop concrete actions to address short and medium term skills needs.**

Follow DRIVES project at:

More information at:

www.project-drives.eu

Co-funded by the Erasmus+ Programme of the European Union

The Development and Research on Innovative Vocational Educational Skills project (DRIVES) is co-funded by the Erasmus+ Programme of the European Union under the agreement 591988-EPP-1-2017-1-CZ-EPPKA2-SSA-B. The European Commission support for the production of this publication does not constitute endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.