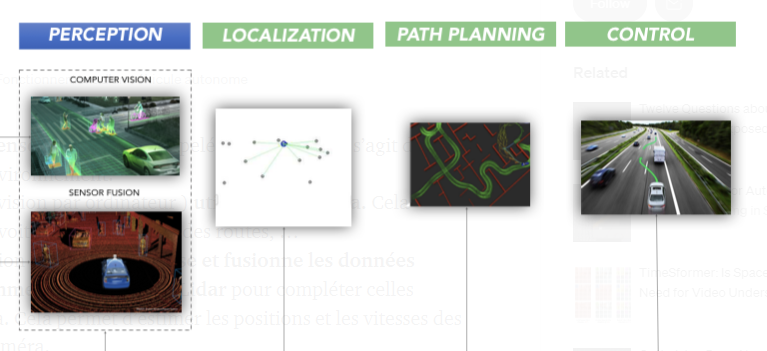
COMPUTER VISION

In the last year, the evolution towards electric vehicles accelerated faster than anticipated. As technology advances, the car industry has used new developments to develop new ways to ease the user (driver). One of them includes using artificial intelligence to make self-driving cars.

The image below shows four main stages in the operation of an autonomous vehicle.



So, a self-driving car (also known as an autonomous car or driverless car) is a vehicle that uses a different number of sensors, radars, cameras, and **artificial intelligence** to travel to destinations without needing a human driver.

These are providing the essential technologies for autonomous cars to improve driving comfort, efficiency and safety.

How??

There are two subsets of artificial intelligence.

**Machine learning**

Machine learning is a set of methods that can automatically detect patterns in data.

These uncovered patterns are then used to predict future data, or to perform other kinds of decisions.

Machine learning techniques offers many algorithm categories that we can choose in our context from:

* regression algorithms.
* pattern recognition.
* cluster algorithms.
* decision matrix algorithms.

Deep Learning is considered an evolution of machine learning. It uses a programmable neural network that enables machines to make accurate decisions without help from humans.

**Deep learning**

In the automotive industry, researchers and developers are actively pushing Deep Learning based approaches for autonomous driving.

Deep learning is a type of machine Learning, which is a subset of artificial intelligence.

Deep Learning employs neural networks and is built to accommodate large volumes of instructure data.

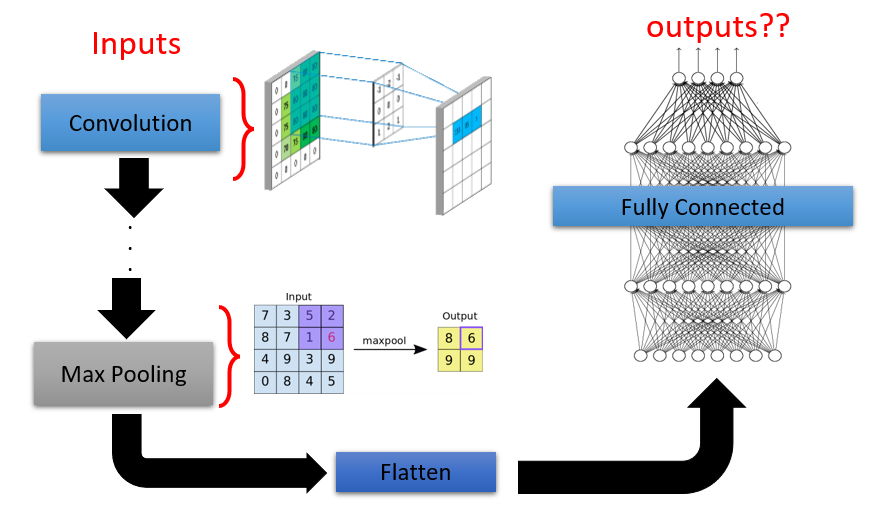
In our project we choose the convolutional neural networks as the algorithm that we will work with in the task of computer vision because of their performance in this field.

Convolutional neural network (CNN) :

We must use algorithms to make estimations about the environment and their location.

In our case, we will choose Convolutional Neural Networks (CNN) , which is one of the most deep neural networks. CNN is an efficient recognition algorithm which is widely used in pattern recognition and image processing.

Convolutional neural networks are neural networks used primarily to classify images (i.e. name what they see), cluster images by similarity (photo search), and perform object recognition within scenes. For example, convolutional neural networks (ConvNets or CNNs) are used to identify faces, individuals, street signs, tumors, platypuses (platypi?) and many other aspects of visual data.



Generally the architecture of CNN is like we presented in the previous figure.It consists essentially of some layers of convolution each of our layers is followed by an activation function ,and other layers of Maxpooling able to reduce the size of the image as well as the amount of parameters and calculation. After these layers, we use a network of neurons composed of also some layers named Fully Connected. The first layer and in the end we can find a layer that uses the Softmax function to calculate the distribution probability of classes that we have chosen.

<https://arxiv.org/ftp/arxiv/papers/1506/1506.01195.pdf#:~:text=CNN%20is%20an%20efficient%20recognition,voice%20analysis%20and%20image%20recognition>.

COMPUTER VISION WITH TENSORFLOW &CNN (GENERAL APPROCH)

TensorFlow offers a workflow modeling that we can follow whilst building a CNN for binary image classification

The essential steps are :

* data gathering
* labeling the dataset and getting it ready
* Creating a CNN model (starting with a baseline)
* Fitting a model (getting it to find patterns in our data)
* Evaluating a model
* Improving a model
* Making a prediction with a trained model

[TensorFlow](https://www.tensorflow.org/) is an open source software library for numerical computation using data-flow graphs.

TensorFlow is cross-platform. It runs on nearly everything: GPUs and CPUs—including mobile and embedded platforms—and even tensor processing units ([TPUs](https://www.blog.google/topics/google-cloud/google-cloud-offer-tpus-machine-learning/)).

using the same technique , we will be implementing three different algorithms that we will need in our autonomous robotaxi:

1-LANE LINE DETECTION

Lane line detection is a crucial component for autonomous driving, the objective is to train the model to describe the white line in the ground to avoid drifting to the other way of the street .

In the scientific literature, we note that many have been interested in the subject of Lane Line detection of road markings. Most of these algorithms provide a complete solution through three essential steps:

– Marking extraction

– Parameter estimation

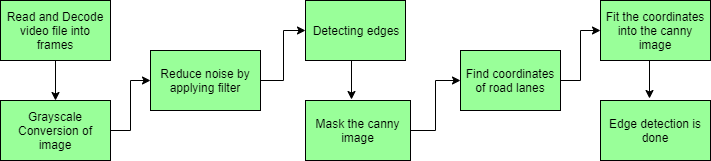
– Monitoring of parameters on image sequence

<https://www.researchgate.net/publication/254200486_Amelioration_de_la_reconnaissance_des_marquages_routiers_par_l'optimisation_d'algorithmes_d'extraction/link/5e577d674585152ce8f294c3/download>

<https://www.researchgate.net/publication/38278988_Nouvel_algorithme_de_detection_des_lignes_de_marquage_au_sol>

In our subject, we will learn how to perform lane detection using videos.

Lane detection involves the following steps:



* Capturing and decoding the video file: We will capture the video using the VideoCapture object and after initializing the capture, each video frame will be decoded (i.e. converting to a sequence of images ).
* Image grayscale conversion: Video images are in RGB format, RGB is converted to grayscale because processing a single-channel image is faster than processing a three-channel colorful image.
* Reduce Noise: Noise can create false edges, so before going any further it is imperative to perform some image smoothing. A Gaussian filter is used to perform this process.
* Canny Edge Detector: It calculates the gradient in all directions of our blurred image and traces the edges with large changes in intensity. For more explanation, please see this article: Canny Edge Detector
* Region of interest: This step consists in taking into account only the region covered by the road. A mask is created here, which is the same dimension as our road image. Additionally, the bitwise AND operation is performed between each pixel of our smart image and this mask. It finally masks the artful image and shows the region of interest traced by the polygonal outline of the mask.
* Hough Line Transform: The Hough Line Transform is a transformation used to detect straight lines. Probabilistic Hough line transformation is used here which gives output as the extremes of detected lines

<https://fr.acervolima.com/opencv-detection-de-voie-routiere-en-temps-reel/>

2-OBJECT DETECTION

object detection goal is to identify all the objects in the path of our vehicule , either to avoid them or to use them as a reference ,

objects can be light, signs, people, other cars ,animals ,etc ..

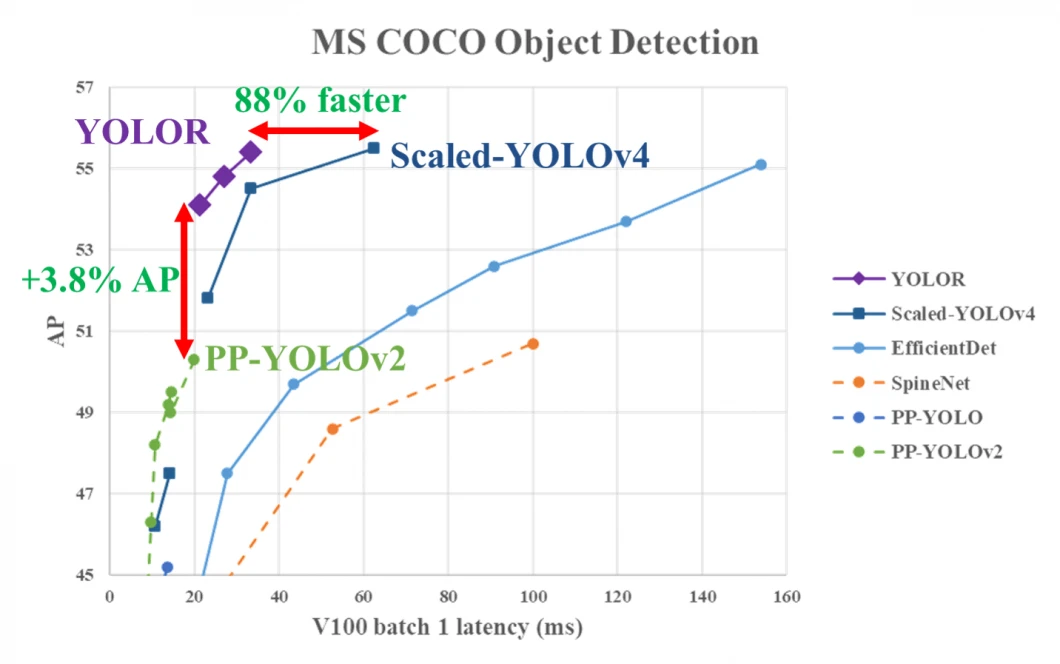
Use-cases based on object detection models have been developed a lot recently: object counting for road traffic analysis, detection of road signs for the autonomous car, detection of faults on infrastructure… The objective here is no longer to classify an image, but to detect the objects within it, by drawing a rectangle (we speak of a bounding box) surrounding the objects present as precisely as possible.

Many algorithms have already emerged: YOLO, R-CNN, Fast R-CNN, Faster R-CNN, SSD, RetinaNet… In this project, we focus on the last version of the YOLO model Knowing that their layers are based on CNN architecture.<https://www.aquiladata.fr/insights/classification-dimages-et-detection-dobjets-par-cnn/>

The last version of YOLO named YOLOR the abbreviation of “You Only Learn One Representation” . It is a new object detection algorithm, rather than other machine learning use cases such as object identification or analysis.

In combination with state-of-the-art methods, the YOLOR achieved comparable object detection accuracy as the Scaled YOLOv4, It is 88% faster and 3.8% more accurate when compared to the YOLO model variants.

### **How YOLOR works??**



Humans are able to learn and understand the physical world based on vision, hearing, tactile (explicit knowledge) – but also based on past experience (implicit knowledge).

Based on this idea YOLOR is proposed as a “unified network to encode implicit knowledge and explicit knowledge together”.

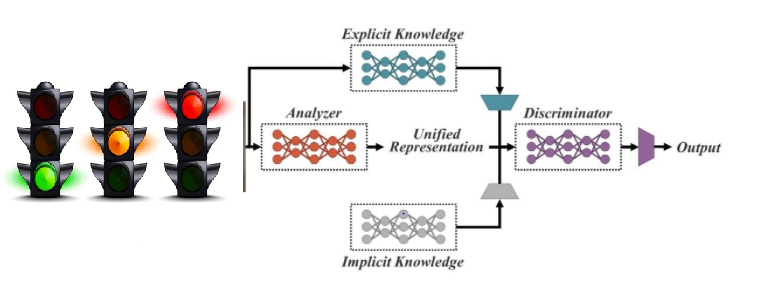


fig: YOLOR concept with implicit and explicit knowledge-based multi-task learning

[**https://viso.ai/deep-learning/yolor/?fbclid=IwAR2aHtoBH7UMRuQoY6JnePtX6Oxm1bmp38gXel-3T\_cFXxu2FmTRWRM4XnI**](https://viso.ai/deep-learning/yolor/?fbclid=IwAR2aHtoBH7UMRuQoY6JnePtX6Oxm1bmp38gXel-3T_cFXxu2FmTRWRM4XnI)

**Data collection**

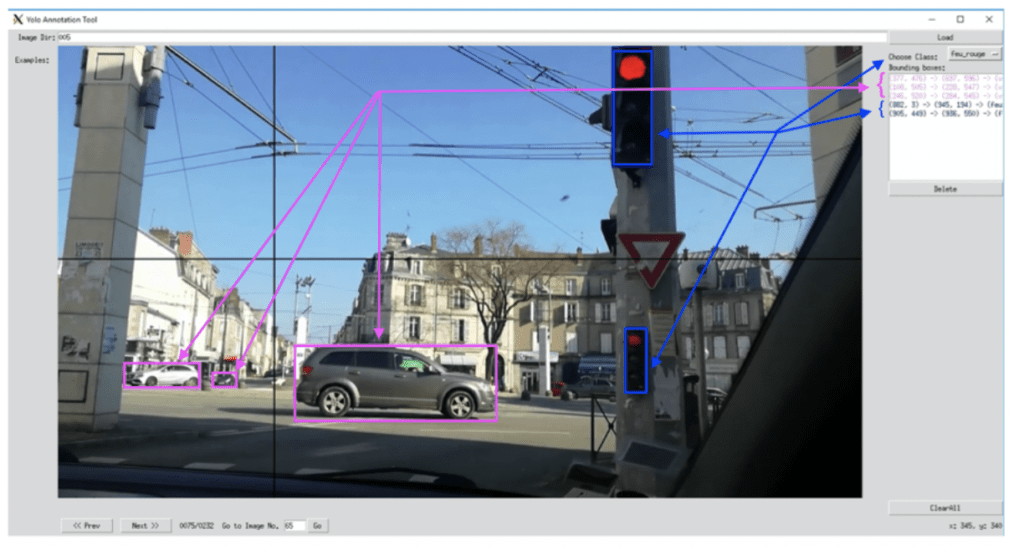
To achieve this goal, we start by installing a camera in a moving car in a built-up area. We film the road from the passenger side of the car on a summer day (in this practical case, the different weather conditions have not been taken into account) for 2 minutes 30 minutes. Then, by means of a script, we recover one frame for every second of video, totaling 150 frames.

**Labeling**

Once this database has been created, we label these images thanks to an open source project called “YOLO Annotation Tool”, particularly suitable for using data on YOLO

(https://github.com/ManzarIMalik/YOLO-Annotation-Tool ).

This labeling consists, for each image of the dataset, in drawing rectangles around the objects that one wishes to recognize, specifying for each rectangle the class of the object that it designates (red light, car, pedestrians, etc.) . As can be seen in the image below, we have drawn the 3 pink rectangles assigning them the car class and then the 2 blue rectangles representing the red lights.



<https://blog.octo.com/you-only-look-once-un-reseau-de-neurones-pour-la-detection-dobjets/>

3-BEHAVIORAL CLONING

This last component determines The ability of the model to drive the car is learned from cloning the behavior of a human driver.

Convolution Neural networks also have a crucial role in these tasks. This kind of neural network will present the ability to predict the steering angle using a frame from the car's front camera.

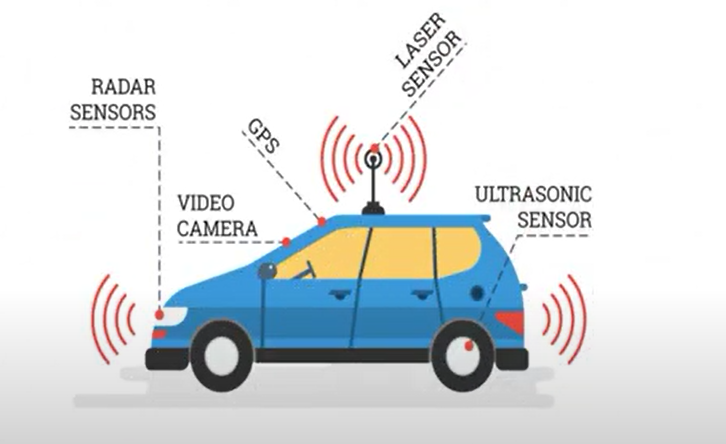
This an end-to-end approach that lets the car drive without lane markings on highways and inroads with unclear visual guidance such as in parking lots and on unpaved roads.

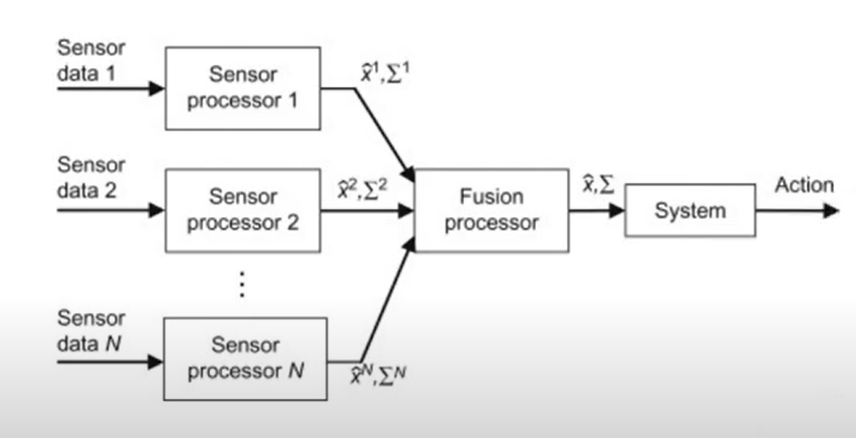
The training data will consist of images and corresponding direction angles. to obtain three labels right, left and center.

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8855753> <https://ichi.pro/fr/clonage-comportemental-176301480485298>

**SENSOR FUSION :**

Autonomous vehicles also need the computing power and advanced machine intelligence to analyze multiple, sometimes conflicting data streams to create a single, accurate view of their environment. Known as ‘sensor fusion’, it is clear that this is an important prerequisite for self-driving cars.





sensors

inertial sensor :