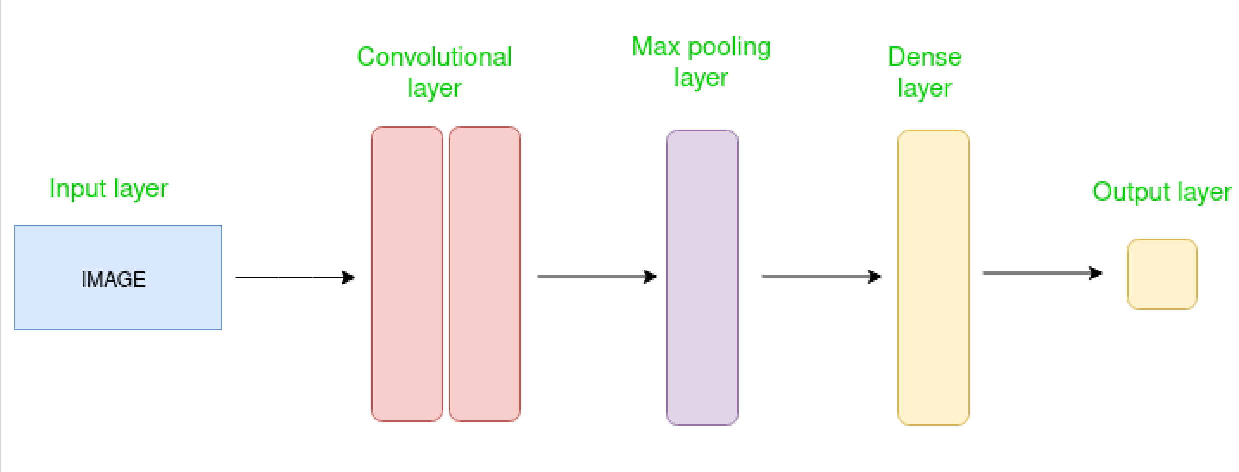
TASK 02

1] **Convolutional Neural Network (CNN)**

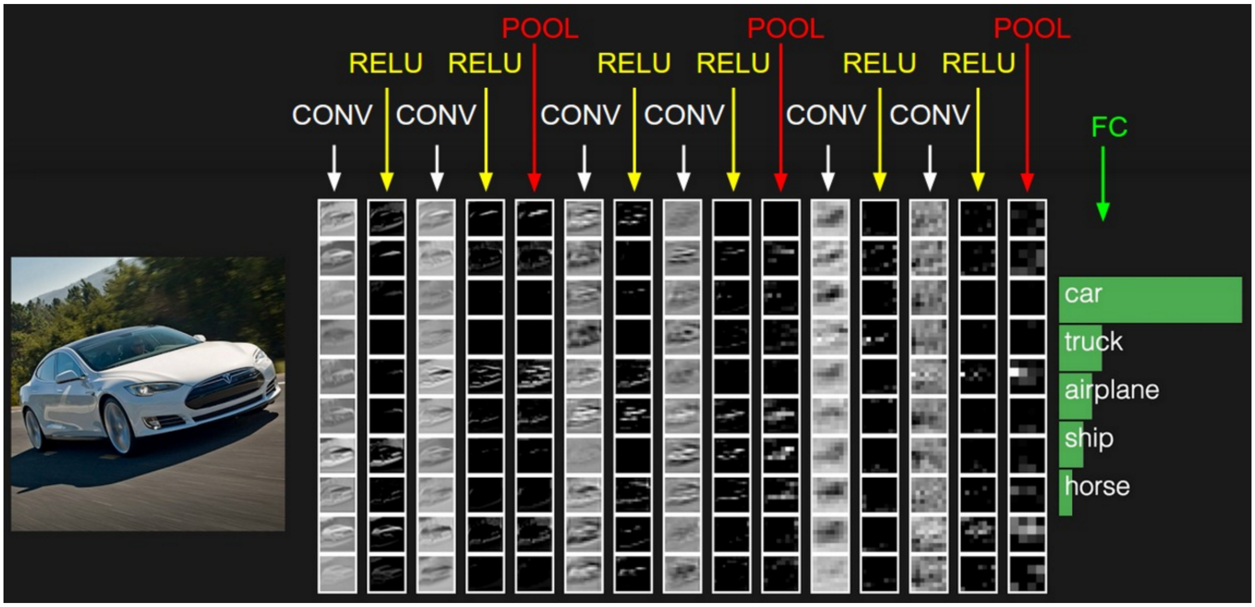
A CNN is a class of artificial neural networks that is particularly effective in processing and analyzing visual data, such as images and videos. It is inspired by the organization and functionality of the visual cortex in the human brain. The algorithm utilizes several key components, including convolutional layers, pooling layers, and fully connected layers.



Types of layers:
Let’s take an example by running a covnets on of image of dimension 32 x 32 x 3.

1. Input Layers: It’s the layer in which we give input to our model. In CNN, Generally, the input will be an image or a sequence of images. This layer holds the raw input of the image with width 32, height 32, and depth 3.
2. Convolutional Layers: This is the layer, which is used to extract the feature from the input dataset. It applies a set of learnable filters known as the kernels to the input images. The filters/kernels are smaller matrices usually 2×2, 3×3, or 5×5 shape. it slides over the input image data and computes the dot product between kernel weight and the corresponding input image patch. The output of this layer is referred ad feature maps. Suppose we use a total of 12 filters for this layer we’ll get an output volume of dimension 32 x 32 x 12.
3. Activation Layer: By adding an activation function to the output of the preceding layer, activation layers add nonlinearity to the network. it will apply an element-wise activation function to the output of the convolution layer. Some common activation functions are RELU: max(0, x), Tanh, Leaky RELU, etc. The volume remains unchanged hence output volume will have dimensions 32 x 32 x 12.

4. Pooling layer: This layer is periodically inserted in the covnets and its main function is to reduce the size of volume which makes the computation fast reduces memory and also prevents overfitting. Two common types of pooling layers are max pooling and average pooling. If we use a max pool with 2 x 2 filters and stride 2, the resultant volume will be of dimension 16x16x12.
5. Flattening: The resulting feature maps are flattened into a one-dimensional vector after the convolution and pooling layers so they can be passed into a completely linked layer for categorization or regression.
6. Fully Connected Layers: It takes the input from the previous layer and computes the final classification or regression task.



Advantages of Convolutional Neural Networks (CNNs):

- Good at detecting patterns and features in images, videos, and audio signals.
- Robust to translation, rotation, and scaling invariance.
- End-to-end training, no need for manual feature extraction.
- Can handle large amounts of data and achieve high accuracy.

Disadvantages of Convolutional Neural Networks (CNNs):

- Computationally expensive to train and require a lot of memory.
- Can be prone to overfitting if not enough data or proper regularization is used.
- Requires large amounts of labeled data.
- Interpretability is limited, it’s hard to understand what the network has learned.

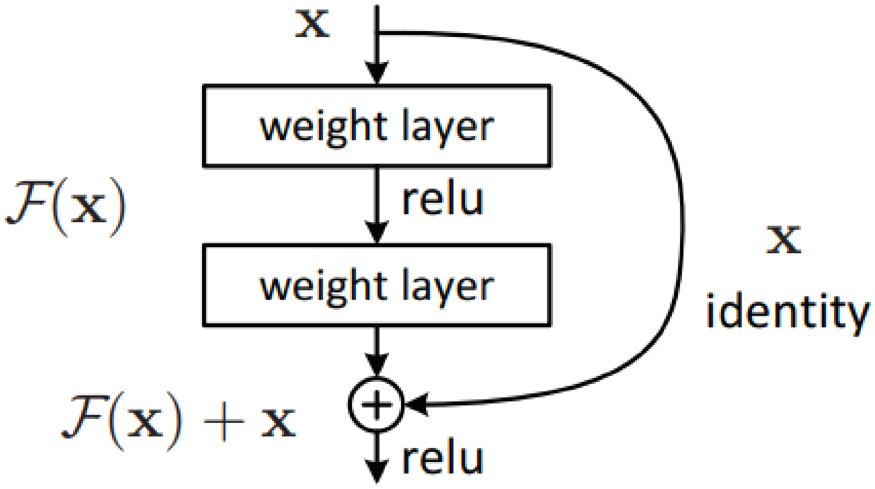
2] ResNet

After the first CNN-based architecture (AlexNet) that win the ImageNet 2012 competition, Every subsequent winning architecture uses more layers in a deep neural network to reduce the error rate. This works for less number of layers, but when we increase the number of layers, there is a common problem in deep learning associated with that called the Vanishing/Exploding gradient. This causes the gradient to become 0 or too large. Thus when we increases number of layers, the training and test error rate also increases.

In order to solve the problem of the vanishing/exploding gradient, this architecture introduced the concept called Residual Blocks. In this network, we use a technique called skip connections. The skip connection connects activations of a layer to further layers by skipping some layers in between. This forms a residual block. Resnets are made by stacking these residual blocks together.

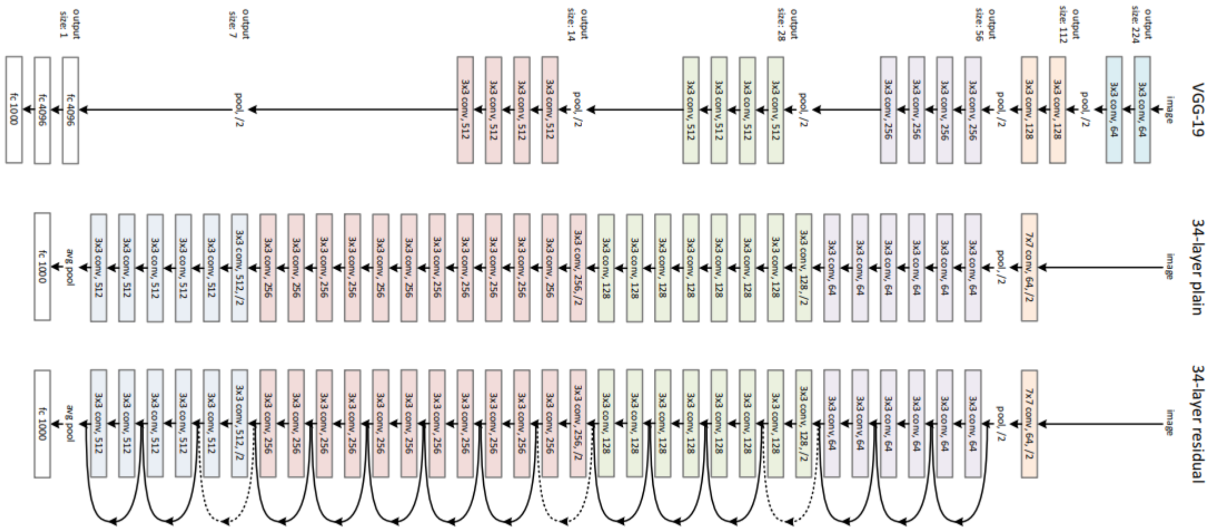
The approach behind this network is instead of layers learning the underlying mapping, we allow the network to fit the residual mapping. So, instead of say H(x), initial mapping, let the network fit,

$F(x) := H(x) - x$ which gives $H(x) := F(x) + x$.


x
weight layer
$\mathcal{F}(\mathbf{x})$
relu
weight layer
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

The advantage of adding this type of skip connection is that if any layer hurt the performance of architecture then it will be skipped by regularization.

Network Architecture: This network uses a 34-layer plain network architecture inspired by VGG-19 in which then the shortcut connection is added. These shortcut connections then convert the architecture into a residual network.



Advantages of ResNet

- Deep network training: ResNet allows for the training of extremely deep neural networks, surpassing the limitations of traditional architectures. This is achieved through the use of residual connections, which allow the gradients to flow directly from the earlier layers to the later layers, mitigating the vanishing gradient problem. As a result, ResNet can effectively learn from deeper layers and capture more complex patterns and features.
- Improved accuracy: ResNet has demonstrated improved accuracy in various computer vision tasks, such as image classification, object detection, and image segmentation. The deep architecture and residual connections enable the network to learn more discriminative features, leading to better classification and detection performance.
- Faster convergence: ResNet's residual connections provide shortcuts for gradient propagation, enabling faster convergence during training. This allows the network to converge more quickly and effectively learn the desired representations, reducing the time required for training deep models.
- Transfer learning: Due to their deep architecture and superior performance, pre-trained ResNet models are often used as a starting point for transfer learning in computer vision tasks. By utilizing the knowledge and feature representations learned from large-scale datasets, pre-trained ResNet models can be fine-tuned on smaller, specialized datasets, resulting in improved performance with less data.

Disadvantages of ResNet

- Increased computational complexity: ResNet's deeper architecture and residual connections lead to increased computational complexity compared to shallower networks.

The additional layers and connections require more memory and computational resources during both training and inference.

- Overfitting potential: The increased depth of ResNet models can make them more prone to overfitting, especially when training data is limited. Regularization techniques, such as dropout and weight decay, are commonly used to mitigate this issue and improve generalization.
- Gradient explosion: Although ResNet addresses the problem of vanishing gradients, there is still a risk of gradients exploding during training. Careful initialization, regularization, and gradient clipping techniques are often necessary to prevent unstable training caused by gradient explosion.
- Model interpretability: The deep and complex architecture of ResNet models can make it challenging to interpret and understand the learned representations. The numerous layers and non-linear transformations make it harder to extract human-understandable insights from the network's internal representations.

3] YOLO

The YOLO (You Only Look Once) algorithm is a popular object detection algorithm used in computer vision tasks. It stands out for its ability to achieve real-time object detection with high accuracy.

YOLO algorithm is important because of the following reasons:

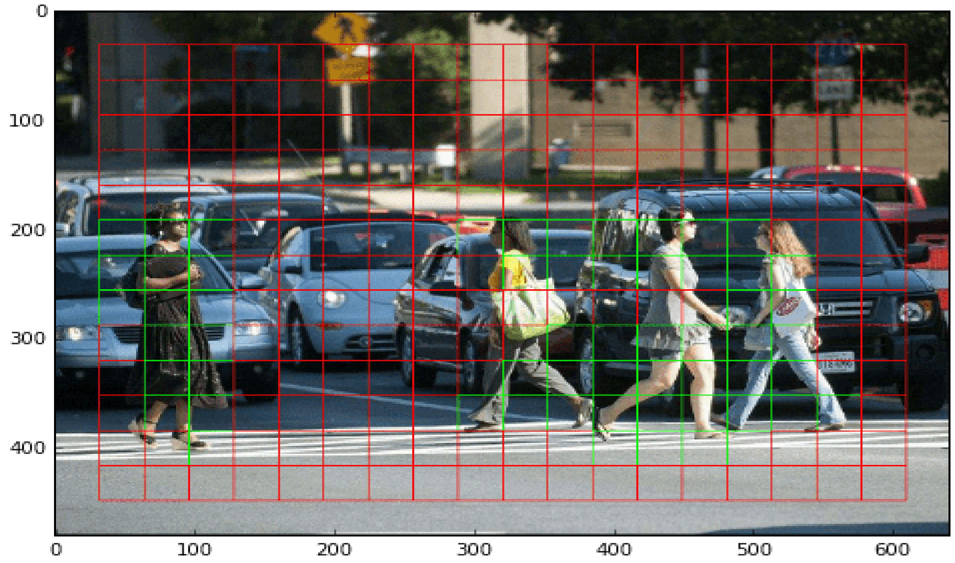
- Speed: This algorithm improves the speed of detection because it can predict objects in real-time.
- High accuracy: YOLO is a predictive technique that provides accurate results with minimal background errors.
- Learning capabilities: The algorithm has excellent learning capabilities that enable it to learn the representations of objects and apply them in object detection.

YOLO algorithm works using the following three techniques:

- Residual blocks
- Bounding box regression
- Intersection Over Union (IOU)

Residual blocks

First, the image is divided into various grids. Each grid has a dimension of S x S. The following image shows how an input image is divided into grids.



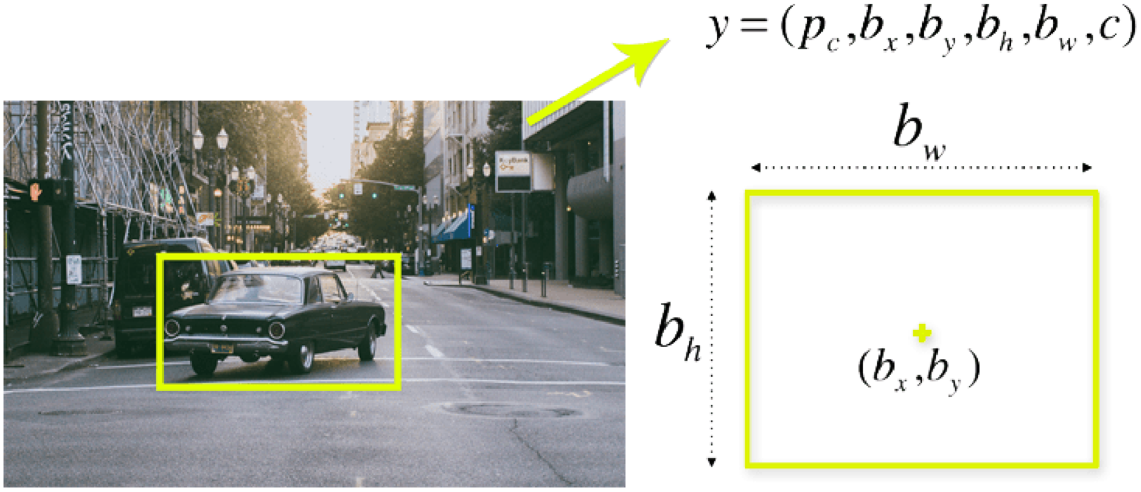
Bounding box regression

A bounding box is an outline that highlights an object in an image.

Every bounding box in the image consists of the following attributes:

- Width (bw)
- Height (bh)
- Class (for example, person, car, traffic light, etc.)- This is represented by the letter c.
- Bounding box center (bx,by)

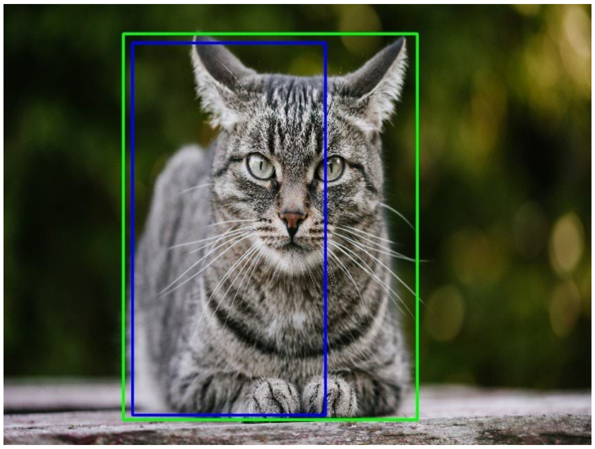
The following image shows an example of a bounding box. The bounding box has been represented by a yellow outline.



Intersection over union (IOU)

Intersection over union (IOU) is a phenomenon in object detection that describes how boxes overlap. YOLO uses IOU to provide an output box that surrounds the objects perfectly.

Each grid cell is responsible for predicting the bounding boxes and their confidence scores. The IOU is equal to 1 if the predicted bounding box is the same as the real box. This mechanism eliminates bounding boxes that are not equal to the real box.
The following image provides a simple example of how IOU works.



YOLO algorithm can be applied in the following fields:

- Autonomous driving: YOLO algorithm can be used in autonomous cars to detect objects around cars such as vehicles, people, and parking signals. Object detection in autonomous cars is done to avoid collision since no human driver is controlling the car.
- Wildlife: This algorithm is used to detect various types of animals in forests. This type of detection is used by wildlife rangers and journalists to identify animals in videos (both recorded and real-time) and images. Some of the animals that can be detected include giraffes, elephants, and bears.
- Security: YOLO can also be used in security systems to enforce security in an area. Let’s assume that people have been restricted from passing through a certain area for security reasons. If someone passes through the restricted area, the YOLO algorithm will detect him/her, which will require the security personnel to take further action.