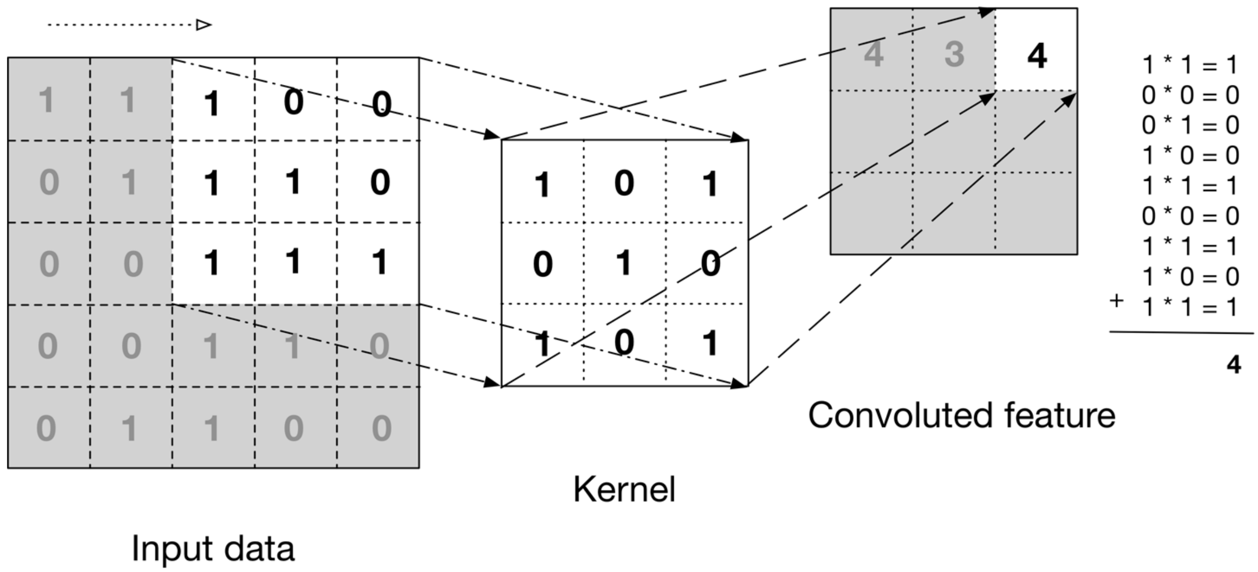
1. Convolutional Neural Networks (CNN):

reduce the images into a form that is easier to process, without losing features that are critical for getting a good prediction.

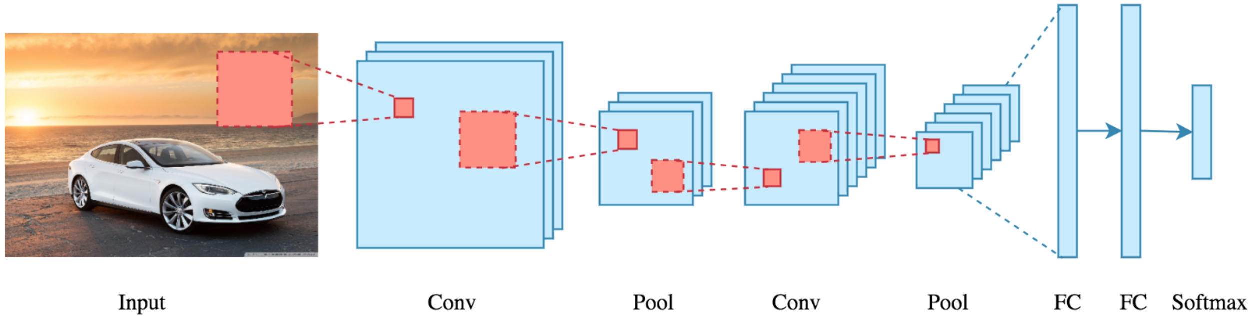


The first layer usually extracts basic features such as horizontal or diagonal edges. This output is passed on to the next layer which detects more complex features such as corners or combinational edges. As we move deeper into the network it can identify even more complex features such as objects, faces, etc.

Pooling layer is responsible for reducing the spatial size of the Convolved Feature. This is to decrease the computational power required to process the data by reducing the dimensions. There are two types of pooling, average pooling and max pooling.

Despite the power and resource complexity of CNNs, they provide in-depth results. At the root of it all, it is just recognizing patterns and details that are so minute and inconspicuous that it goes unnoticed to the human eye. But when it comes to understanding the contents of an image it fails.

CNN’s are used in many computer vision applications such as **facial recognition**, image search, and editing, augmented reality.

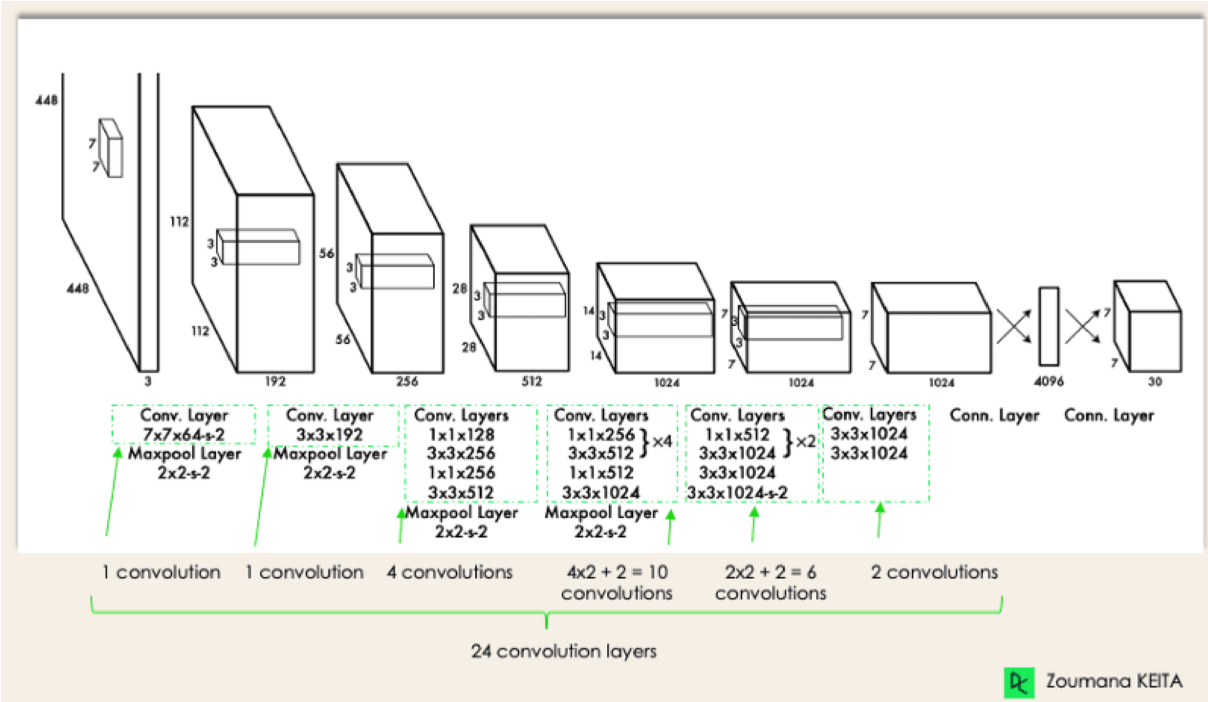


- Pros:
  - Effective in image and video recognition tasks due to their ability to capture local spatial dependencies through convolutional layers.
  - Parameter sharing and spatial invariance properties reduce the number of parameters required, making CNNs computationally efficient.
  - Can learn hierarchical representations of data through multiple convolutional layers.
  - Widely used and well-established in computer vision tasks.
- Cons:
  - Lack of interpretability in deeper layers due to the complex transformations learned.
  - Limited ability to handle varying input sizes without additional preprocessing.
  - Sensitive to changes in input scale and translation.
  - Difficulty in capturing long-range dependencies in sequential data.

2. Residual Networks (ResNet):

- Pros:
  - Overcome the vanishing gradient problem by using skip connections, allowing the network to learn residual mappings.
  - Enable training of much deeper networks (e.g., 50, 101, or 152 layers) without degradation in performance.
  - Improved optimization and convergence speed due to the skip connections.
  - State-of-the-art performance in image classification and other computer vision tasks.
- Cons:
  - Increased computational complexity and memory requirements compared to shallower networks.
  - More challenging to interpret and visualize due to the skip connections and deep layer interactions.
  - Higher risk of overfitting with very deep architectures.

3. You Only Look Once (YOLO):



○ Pros:

■ Real-time object detection and localization with impressive speed and accuracy.

■ Single pass through the network makes it efficient for applications requiring fast inference.

■ Unified framework for object detection and classification.

■ Can handle multi-object detection in a single prediction.

■ YOLOv4 and YOLOv5 introduced various improvements, including enhanced backbone networks, feature fusion techniques, and attention mechanisms.

○ Cons:

■ YOLO struggles with detecting small objects due to the inherent downsampling in the network.

■ Localization accuracy may be lower compared to region-based methods.

■ Vulnerable to occlusion and overlapping objects.

■ Limited ability to handle object scales and aspect ratios.

Analysis Report on Deep Learning Models:

1. CNN:

○ Working: CNNs consist of convolutional layers for extracting features, pooling layers for downsampling, and fully connected layers for classification. Convolutional layers perform convolution operations on input data using filters/kernels to detect features.

○ Architecture: CNNs typically have an input layer, multiple convolutional and pooling layers, and one or more fully connected layers for classification.

2. ResNet:

○ Working: ResNet introduces skip connections, allowing the network to learn residual mappings by adding the input to the output of each layer. This enables the network to avoid vanishing gradients and facilitates training of deeper networks.

○ Architecture: ResNet architectures have multiple residual blocks, each containing several convolutional layers and skip connections. The skip connections add the input to the output of the block.

3. YOLO (You Only Look Once):

○ Working: YOLO divides the input image into a grid and predicts bounding boxes and class probabilities directly from the grid cells. It uses anchor boxes to handle object scales and aspect ratios and predicts bounding box coordinates relative to the grid cell.

○ Architecture: YOLO architectures consist of a backbone network for feature extraction, followed by additional convolutional layers for detection. YOLOv4 and YOLOv5 introduced enhancements like feature fusion, attention mechanisms, and other architectural improvements.

Useful features of each model:

- CNN: Local spatial dependency capture, parameter sharing, hierarchical representations.
- ResNet: Skip connections, improved training of deep networks, state-of-the-art performance.
- YOLO: Real-time object detection, unified framework, fast inference, multi-object detection.

YOLO

YOLO is a real-time object detection algorithm that operates on entire images, aiming to detect objects and their bounding boxes in a single pass. Unlike traditional object detection methods that involve sliding windows or region proposals, YOLO takes a different approach to achieve efficiency and accuracy. introduced in many practical industries such as healthcare and agriculture.

Key Features of YOLO:

1. Grid-based Approach: YOLO divides the input image into a grid and assigns each cell responsibility for predicting objects. Each cell predicts a fixed number of bounding boxes along with their class probabilities.
2. Single Pass Detection: YOLO performs object detection in a single pass through the neural network, making it extremely fast compared to other methods that require multiple stages or region proposals.
3. Feature Extraction: YOLO uses a convolutional neural network to extract features from the input image. These features are then used for predicting bounding boxes and class probabilities.
4. Multi-Scale Training: YOLO trains on images of different scales to improve its ability to detect objects at various sizes. It uses anchor boxes to handle objects of different aspect ratios.

References :

1. https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/
2. https://www.geeksforgeeks.org/introduction-convolution-neural-network/
3. https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939
4. https://www.v7labs.com/blog/yolo-object-detection
5. https://www.datacamp.com/blog/yolo-object-detection-explained