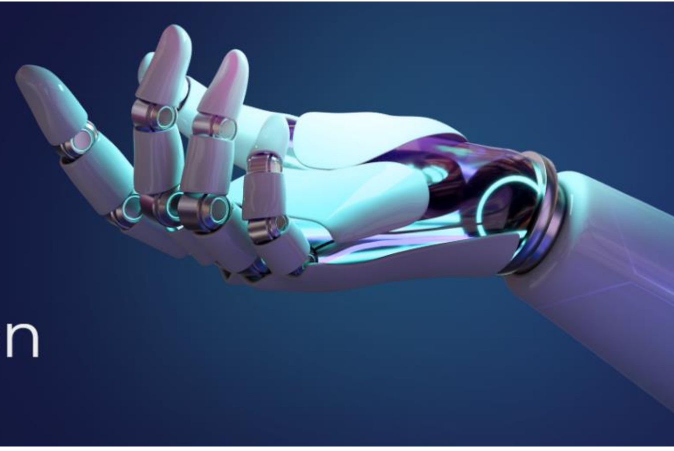
ARTIFICIAL INTELLIGENCE



ResNet-50 Documentation

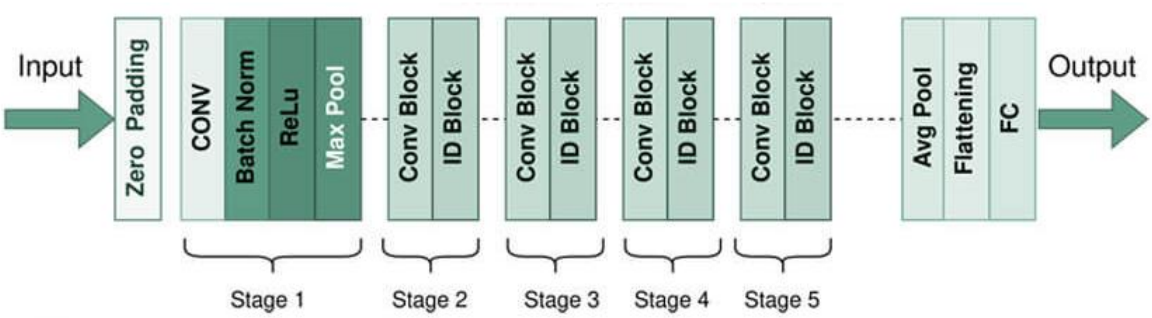
1- Introduction

ResNet-50 ("Residual Network") is a deep convolutional neural network (CNN) designed to address the degradation problem in deep learning. Introduced by researchers from Microsoft Research in their 2015 paper "Deep Residual Learning for Image Recognition," ResNet-50 is an architecture that uses residual connections to ease the training of deep networks while maintaining high accuracy.

2- Key Features

- Depth: ResNet-50 consists of 50 learning layers.
- Residual Connections (skip connections): These connections bypass one or more layers, alleviating the vanishing gradient problem.
- Residual Block: Each residual block contains convolutional layers, batch normalization, and a ReLU (Rectified Linear Unit) activation function.
- Bottleneck Design: To reduce computational costs, ResNet-50 employs a bottleneck approach where convolutions are performed with reduced dimensions.

3- Architecture



4- The architecture of ResNet-50 comprises the following components:

Layer	Description
Input	Input image sized 224 x 224 x 3 (RGB).
Initial Convolution	7x7 convolution with stride 2, followed by batch normalization and ReLU.
MaxPooling	3x3 pooling with stride 2, reduces spatial dimensions.
Residual Block 1	1 block: 3 bottleneck layers (64, 256).
Residual Block 2	4 blocks: each with 3 bottleneck layers (128, 512).
Residual Block 3	6 blocks: each with 3 bottleneck layers (256, 1024).
Residual Block 4	3 blocks: each with 3 bottleneck layers (512, 2048).
Global Average Pooling (GAP)	Reduces spatial dimensions to a single-dimensional output.
Fully Connected Layer	Dense layer producing the final output with class probabilities.

5- Functionality of Each Layer

Layer	Functionality
Input	Accepts the input image and prepares it for processing in the network.
Initial Convolution	Extracts low-level features such as edges and textures.
MaxPooling	Reduces spatial resolution to focus on prominent features.
Residual Block 1	Captures simple patterns and features (e.g., edges, corners).
Residual Block 2	Identifies intermediate features, such as shapes or object parts.
Residual Block 3	Detects higher-level features, such as complex shapes and textures.
Residual Block 4	Encodes abstract and semantic information for objects in the image.
Global Average Pooling (GAP)	Aggregates spatial information into a single vector for classification.
Fully Connected Layer	Outputs probabilities for each class in the dataset.

6- A residual block in ResNet-50 includes:

- **<u>Three convolutional layers:</u>**
  - A 1x1 convolution to reduce dimensions.
  - A 3x3 convolution to capture spatial features.

  - Another 1x1 convolution to restore original dimensions.

- A residual connection that directly adds the block's input to its output, enabling better information flow.

7- Advantages of ResNet-50

1. Better Convergence: Residual connections allow deeper networks to be trained without loss of accuracy.
2. Improved Accuracy: ResNet-50 has set performance records on datasets like ImageNet.
3. Generalizable Model: ResNet-50 is widely used in applications such as computer vision, medical diagnostics, and facial recognition.

8- Applications

ResNet-50 is commonly used as a baseline model for:

- Image Classification: Tasks where each image is assigned a class.
- Object Detection: Identifying and locating objects within an image.
- Image Segmentation: Dividing an image into segments based on the objects present.

9- Training and Fine-Tuning

ResNet-50 can be trained from scratch on a new dataset, but it is often used with transfer learning. Pretrained weights from ImageNet can be fine-tuned for specific tasks, achieving good performance with fewer data and reduced computational costs.

10- Limitations

1. Computational Cost: ResNet-50 is relatively heavy in terms of computation and memory.
2. Overfitting: If not properly regularized, the model may overfit smaller datasets.