Inception V1

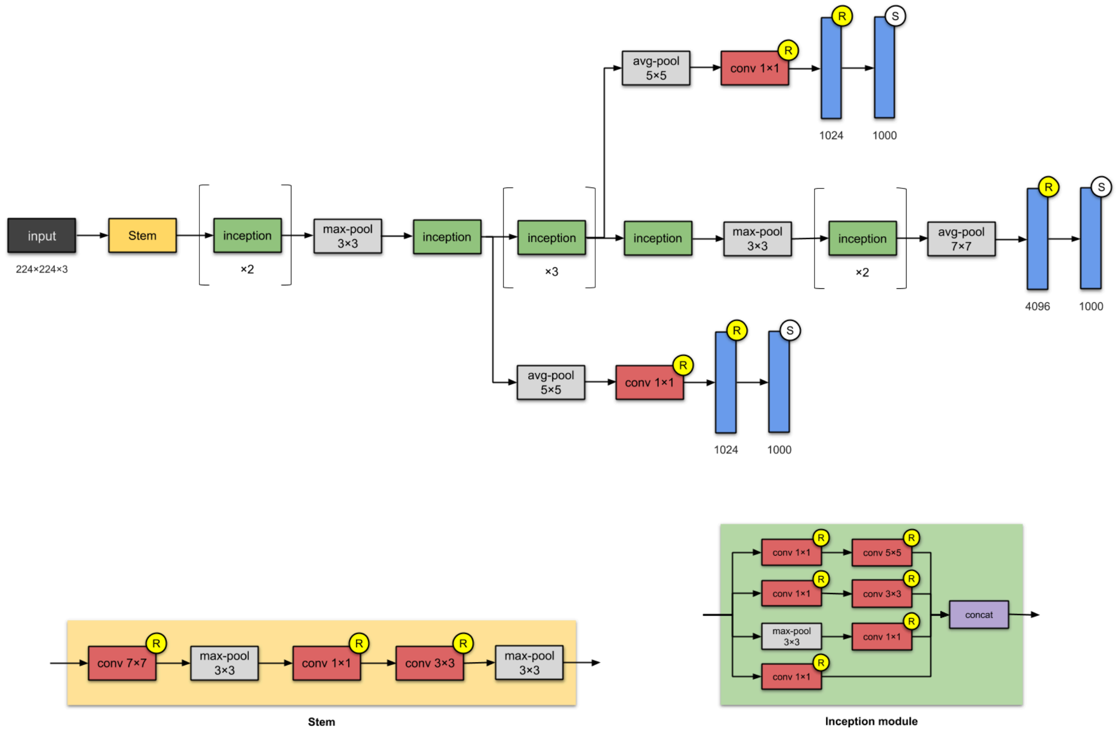

Figure 1: Inception V1

Inception v1, also known as GoogLeNet, introduced the **Inception module**, which combines multi-scale feature extraction with dimensionality reduction to achieve an efficient and powerful architecture.

Architecture

- Inception module
  - Consists of parallel branches with convolutional filters of sizes 1x1, 3x3, 5x5, and a 3x3 max-pooling operation
  - Outputs of all branches are concatenated along the depth dimension.
  - Convolutional layer with filters of sizes 1x1 are applied to reduce the depth of feature maps, minimizing computational cost and enabling deeper architectures.

Inception V3

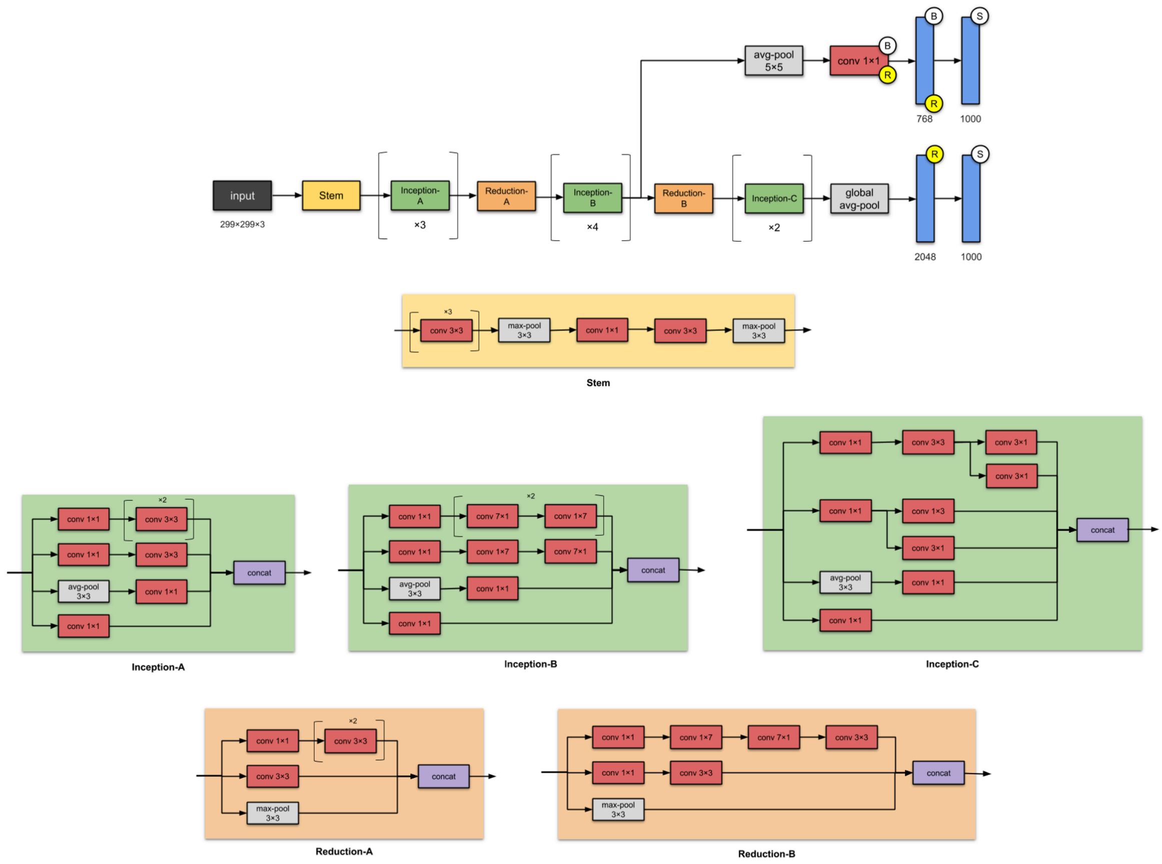

Figure 2: Inception V3

Inception v3 improved upon v1 with optimization techniques that enhanced efficiency and accuracy.

Architecture

- Factorized Convolutions:
  - Large filters (e.g., 7x7) are replaced with smaller ones (e.g., 1x7 followed by 7x1) to reduce computational cost.
  - Improves parameter efficiency while maintaining spatial resolution.
  - Solve representational bottlenecks problem
- Batch Normalization
  - Applied after every convolutional layer to stabilize training and improve convergence.
- Inception modules
  - Inception-A: replace 5x5 convolutional layer by 2 3x3 convolutional layer
  - Inception-B: replace 3x3 convolutional layer by 7x1 and 1x7 convolutional layer
  - Inception-C: replace 7x1 and 1x7 convolutional layer by 3x1 and 1x3 convolutional layer

ResNet-50

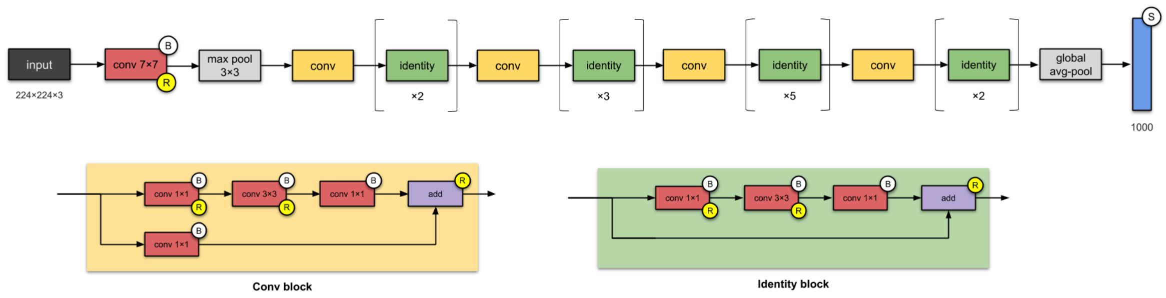

Figure 3: ResNet-50

ResNet, short for Residual Network, introduced **skip connections** to train ultra-deep networks effectively.

Architecture

Convolutional block

- Uses 1x1 convolutions before and after 3x3 convolutions to reduce dimensionality and computation

Identity block

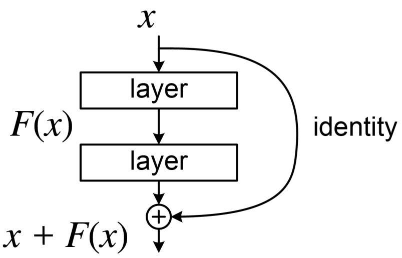

Figure 4: Residual connection

Consider network with N layers, where the output of layer i is

$$z_i = f(W_i \cdot z_{i-1})$$

The gradient of the loss L is

$$\frac{\partial L}{\partial z_0} = \frac{\partial L}{\partial z_N} \prod_{i=1}^{N} \frac{\partial f(W_i \cdot z_{i-1})}{\partial z_{i-1}}$$

Consider the same network with residual connections

$$z_i = z_{i-1} + f(W_i \cdot z_{i-1})$$

The gradient of the loss L is

$$\begin{aligned} \frac{\partial L}{\partial z_0} &= \frac{\partial L}{\partial z_N} \prod_{i=1}^{N} \left(1 + \frac{\partial f(W_i \cdot z_{i-1})}{\partial z_{i-1}}\right) \\ &= \frac{\partial L}{\partial z_N} + \frac{\partial L}{\partial z_N} \prod_{i=1}^{N} \frac{\partial f(W_i \cdot z_{i-1})}{\partial z_{i-1}} \end{aligned}$$

Variants

- ResNet-50: 50 layers, includes bottleneck blocks for optimized depth.
- ResNet-101 and ResNet-152: Extended depths for more complex feature hierarchies.