Image Processing with Deep Learning

Tushar B. Kute,
http://tusharkute.com

ImageNet

- ImageNet is a large-scale image database designed for visual object recognition software research. It contains over 14 million images and 1,000 classes.
- The images are organized into a hierarchical taxonomy, with each class representing a different object or scene.
- ImageNet was created by the Stanford Vision Lab and first released in 2009.
- It has since become a benchmark for evaluating the performance of visual object recognition software.

ImageNet

- The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is an annual competition that tests the accuracy of visual object recognition software on the ImageNet dataset.
- ImageNet has been used to train a number of successful deep learning models, including AlexNet, VGGNet, ResNet, and Inception.
- These models have achieved state-of-the-art results on the ILSVRC challenge, and they have been used to develop a variety of real-world applications, such as self-driving cars and facial recognition software.

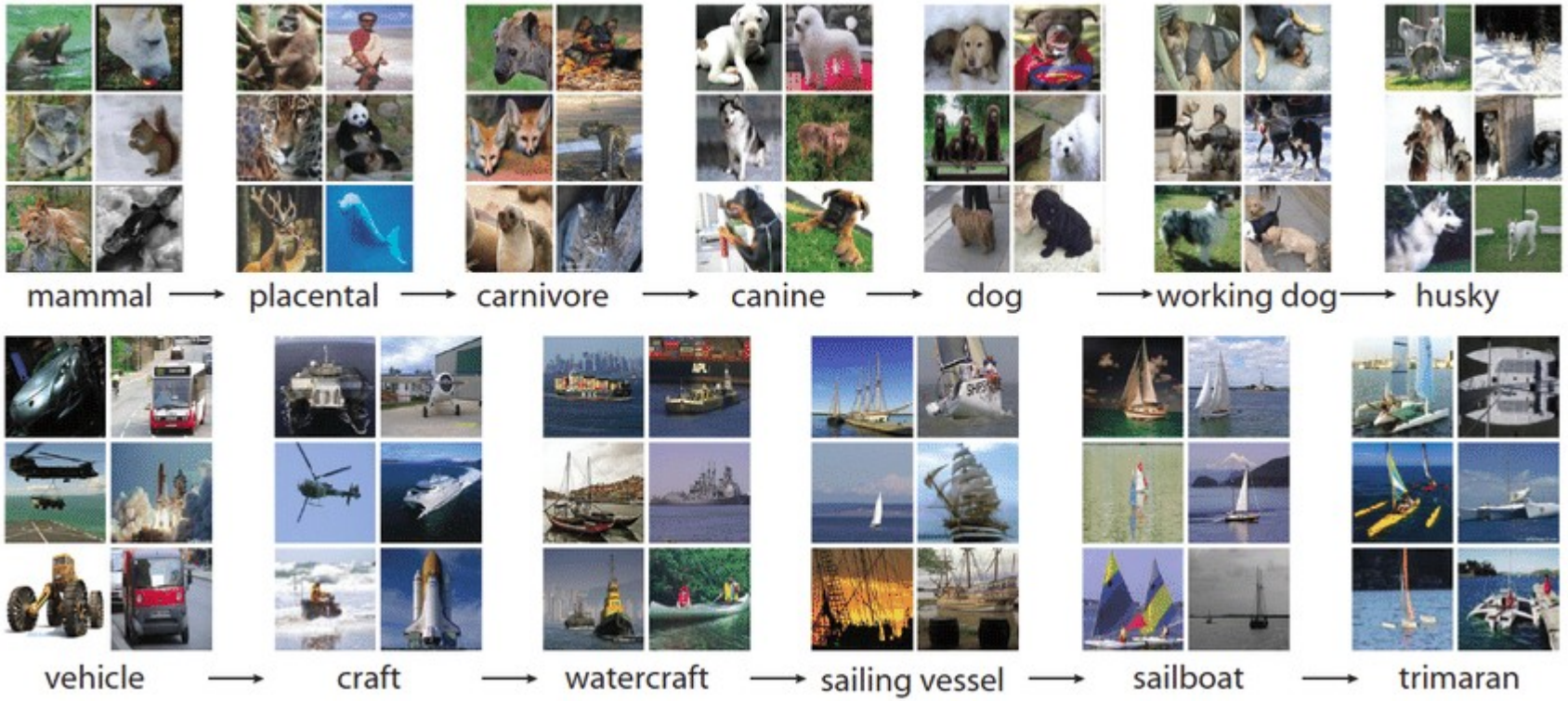
ImageNet: Benefits

- It is a large and diverse dataset, which allows for the training of more accurate models.
- The images are organized into a hierarchical taxonomy, which makes it easier to train models for specific tasks.
- The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) provides a benchmark for evaluating the performance of visual object recognition software.

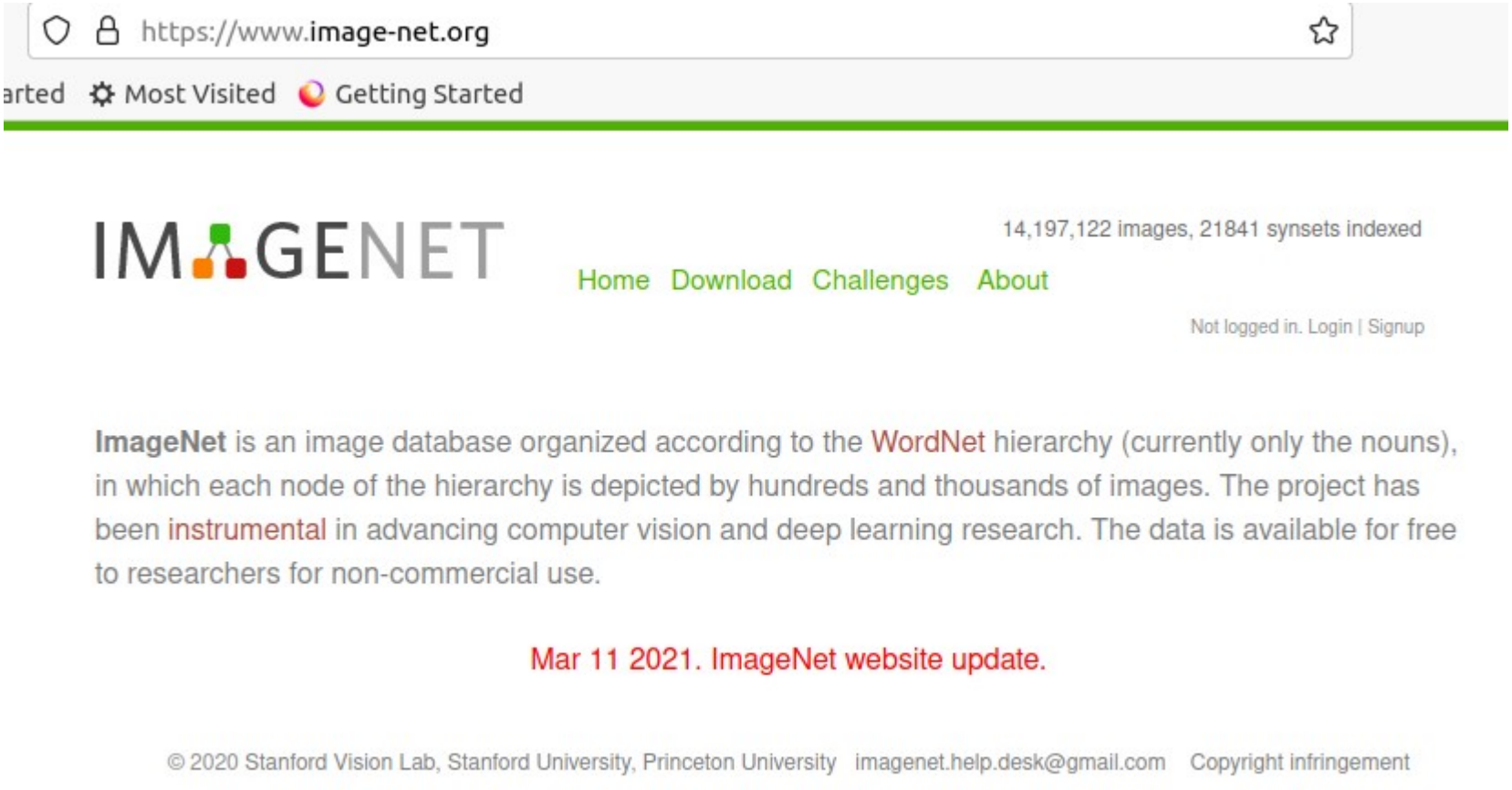
ImageNet: Limitations

- The images are all labeled by humans, which can be a time-consuming and expensive process.
- The dataset is biased towards Western culture, which can make it less effective for training models for other cultures.
- The dataset is not always up-to-date, which can make it less effective for training models for new objects or scenes.

ImageNet Dataset



ImageNet Web link



AlexNet

- AlexNet is a convolutional neural network (CNN) architecture that was proposed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton in their 2012 paper, "ImageNet Classification with Deep Convolutional Neural Networks."
- AlexNet was the first CNN to achieve state-of-the-art results on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), and it is considered to be one of the most influential papers in the field of deep learning.

AlexNet

- AlexNet consists of eight layers, including five convolutional layers and three fully connected layers.
- The convolutional layers use rectified linear units (ReLUs) as activation functions, and the fully connected layers use softmax activation functions.
- AlexNet also uses dropout regularization to prevent overfitting.
- AlexNet was trained on the ImageNet dataset, which contains over 14 million images and 1,000 classes.
- AlexNet achieved a top-5 error rate of 15.3% on the ILSVRC 2012 challenge, which was a significant improvement over the previous state-of-the-art.

AlexNet

- AlexNet has been used as a baseline for many subsequent CNN architectures, and it has inspired a number of new research directions in deep learning.
- AlexNet is still a powerful CNN architecture, and it can be used for a variety of image classification tasks.

AlexNet

- Here are some of the key features of AlexNet:
  - It is a deep CNN architecture, with eight layers.
  - It uses ReLUs as activation functions.
  - It uses dropout regularization to prevent overfitting.
  - It was trained on the ImageNet dataset.
  - It achieved state-of-the-art results on the ILSVRC 2012 challenge.

AlexNet

- AlexNet is a significant milestone in the history of deep learning.
- It showed that deep CNNs could achieve state-of-the-art results on challenging image classification tasks.
- AlexNet has inspired a number of new research directions in deep learning, and it is still a powerful CNN architecture that can be used for a variety of image classification tasks.

AlexNet

- The architecture consists of eight layers: five convolutional layers and three fully-connected layers. But this isn't what makes AlexNet special; these are some of the features used that are new approaches to convolutional neural networks:
  - ReLU Nonlinearity.
    - AlexNet uses Rectified Linear Units (ReLU) instead of the tanh function, which was standard at the time.
    - ReLU's advantage is in training time; a CNN using ReLU was able to reach a 25% error on the CIFAR-10 dataset six times faster than a CNN using tanh.

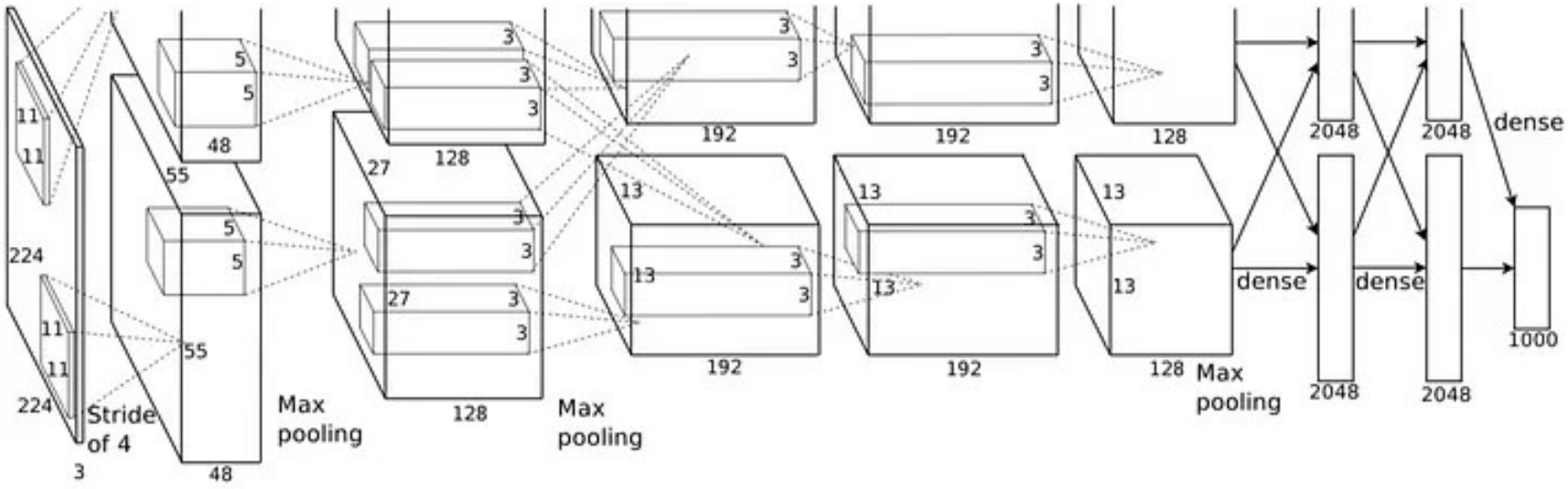
AlexNet

- Multiple GPUs. Back in the day, GPUs were still rolling around with 3 gigabytes of memory (nowadays those kinds of memory would be rookie numbers).
- This was especially bad because the training set had 1.2 million images.
- AlexNet allows for multi-GPU training by putting half of the model's neurons on one GPU and the other half on another GPU.
- Not only does this mean that a bigger model can be trained, but it also cuts down on the training time.

AlexNet

- Overlapping Pooling. CNNs traditionally “pool” outputs of neighboring groups of neurons with no overlapping.
- However, when the authors introduced overlap, they saw a reduction in error by about 0.5% and found that models with overlapping pooling generally find it harder to overfit.

AlexNet


224
224
3
11
11
11
11
Stride
of 4
55
55
48
48
5
5
5
5
Max
pooling
27
27
128
128
3
3
3
3
3
3
Max
pooling
192
192
13
13
13
13
13
13
192
192
128
128
Max
pooling
dense
dense
2048
2048
2048
2048
dense
1000

AlexNet

Layer	# filters / neurons	Filter size	Stride	Padding	Size of feature map	Activation function
Input	-	-	-	-	227 x 227 x 3	-
Conv 1	96	11 x 11	4	-	55 x 55 x 96	ReLU
Max Pool 1	-	3 x 3	2	-	27 x 27 x 96	-
Conv 2	256	5 x 5	1	2	27 x 27 x 256	ReLU
Max Pool 2	-	3 x 3	2	-	13 x 13 x 256	-
Conv 3	384	3 x 3	1	1	13 x 13 x 384	ReLU
Conv 4	384	3 x 3	1	1	13 x 13 x 384	ReLU
Conv 5	256	3 x 3	1	1	13 x 13 x 256	ReLU
Max Pool 3	-	3 x 3	2	-	6 x 6 x 256	-
Dropout 1	rate = 0.5	-	-	-	6 x 6 x 256	-

AlexNet: Overfitting

- AlexNet had 60 million parameters, a major issue in terms of overfitting. Two methods were employed to reduce overfitting:
  - Data Augmentation. The authors used label-preserving transformation to make their data more varied. Specifically, they generated image translations and horizontal reflections, which increased the training set by a factor of 2048.
  - They also performed Principle Component Analysis (PCA) on the RGB pixel values to change the intensities of RGB channels, which reduced the top-1 error rate by more than 1%.

AlexNet: Dropout

- This technique consists of “turning off” neurons with a predetermined probability (e.g. 50%).
- This means that every iteration uses a different sample of the model’s parameters, which forces each neuron to have more robust features that can be used with other random neurons.
- However, dropout also increases the training time needed for the model’s convergence.

AlexNet: The Results

- On the 2010 version of the ImageNet competition, the best model achieved 47.1% top-1 error and 28.2% top-5 error.
- AlexNet vastly outpaced this with a 37.5% top-1 error and a 17.0% top-5 error.
- AlexNet is able to recognize off-center objects and most of its top five classes for each image are reasonable.
- AlexNet won the 2012 ImageNet competition with a top-5 error rate of 15.3%, compared to the second place top-5 error rate of 26.2%.

AlexNet: What now?

- AlexNet is an incredibly powerful model capable of achieving high accuracies on very challenging datasets.
- However, removing any of the convolutional layers will drastically degrade AlexNet's performance.
- AlexNet is a leading architecture for any object-detection task and may have huge applications in the computer vision sector of artificial intelligence problems.
- In the future, AlexNet may be adopted more than CNNs for image tasks.

VGG

- VGG, or Visual Geometry Group, is a type of convolutional neural network (CNN) architecture that was proposed by Karen Simonyan and Andrew Zisserman of the Visual Geometry Group (VGG) at the University of Oxford in 2014.
- The VGG model is characterized by its simplicity and its use of small 3x3 convolution filters.
- The model was one of the most successful entries in the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC), and it has been used as a basis for many other CNN architectures.

VGG

- The VGG model is composed of a series of convolutional layers, followed by max pooling layers, and then fully connected layers.
- The convolutional layers use small 3x3 filters, which helps to preserve spatial information in the images.
- The max pooling layers downsample the images, which helps to reduce the computational complexity of the model.
- The fully connected layers classify the images into different classes.

VGG

- The VGG model has been shown to be very effective for image classification tasks.
- It achieved a top-5 error rate of 16.4% on the ILSVRC 2014 dataset, which was a significant improvement over the previous state-of-the-art.
- The VGG model has also been used for other tasks, such as object detection and segmentation.

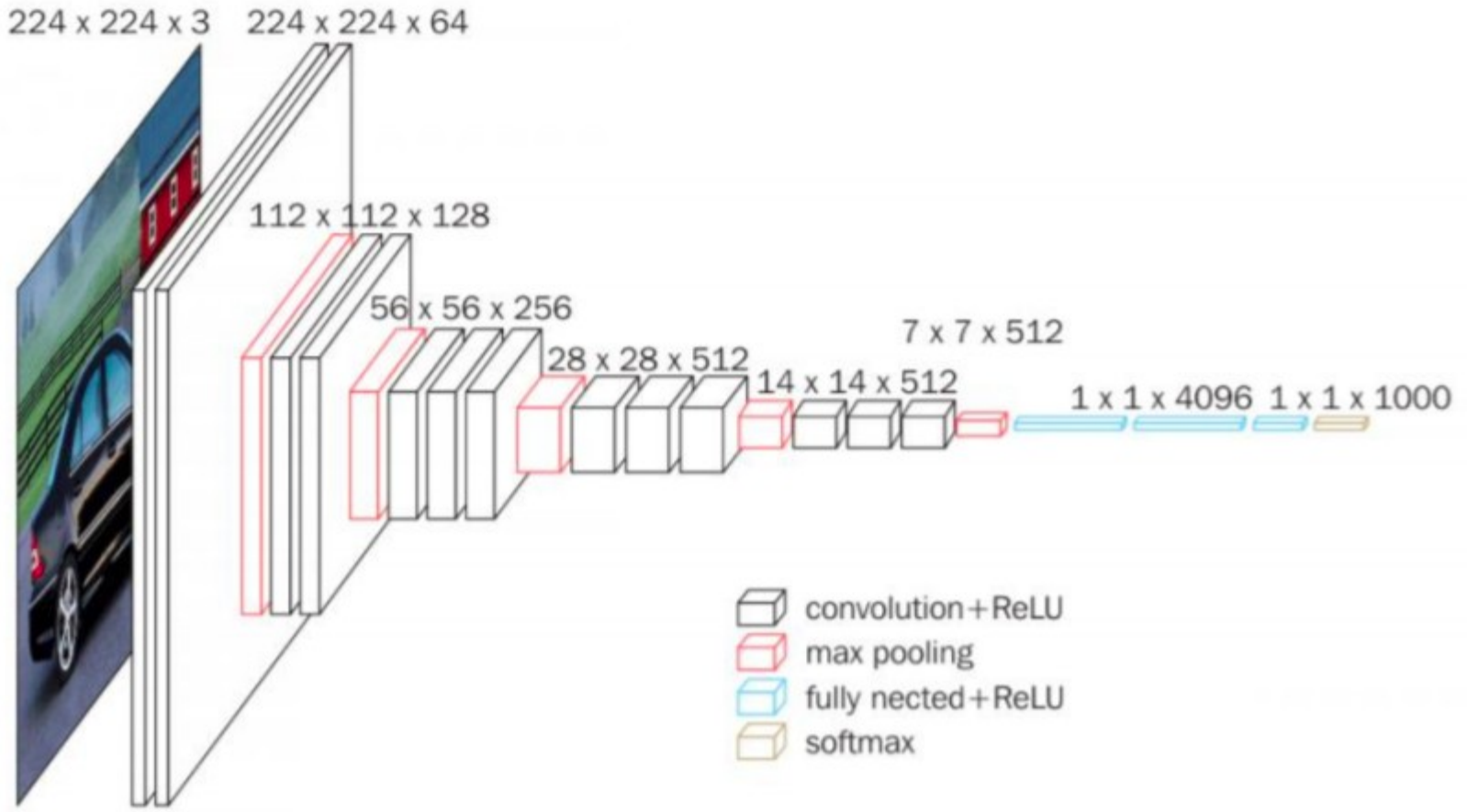
VGG: Architecture

- The input to the convolution neural network is a fixed-size 224 × 224 RGB image.
- The only preprocessing it does is subtracting the mean RGB values, which are computed on the training dataset, from each pixel.
- Then the image is running through a stack of convolutional (Conv.) layers, where there are filters with a very small receptive field that is 3 × 3, which is the smallest size to capture the notion of left/right, up/down, and center part.

VGG: Architecture

- In one of the configurations, it also utilizes 1 × 1 convolution filters, which can be observed as a linear transformation of the input channels followed by non-linearity.
- The convolutional strides are fixed to 1 pixel; the spatial padding of convolutional layer input is such that the spatial resolution is maintained after convolution, that is the padding is 1 pixel for 3 × 3 Conv. layers.
- Then the Spatial pooling is carried out by five max-pooling layers, 16 which follow some of the Conv. layers but not all the Conv. layers are followed by max-pooling.
- This Max-pooling is performed over a 2 × 2-pixel window, with stride 2.

VGG: Architecture



VGG: Advantages

- It is simple and easy to understand.
- It is very effective for image classification tasks.
- It has been used as a basis for many other CNN architectures.

VGG: Disadvantages

- It is computationally expensive to train.
- It is not as good as some newer CNN architectures for some tasks.

ResNet

- ResNet, short for Residual Network, is a type of convolutional neural network (CNN) architecture that was introduced in 2015 by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun in their paper “Deep Residual Learning for Image Recognition”.
- ResNets are very deep CNNs, which means that they have a large number of layers.
- However, training very deep CNNs can be difficult, because the gradients can become very small as they propagate through the network. This can make it difficult for the network to learn.

ResNet

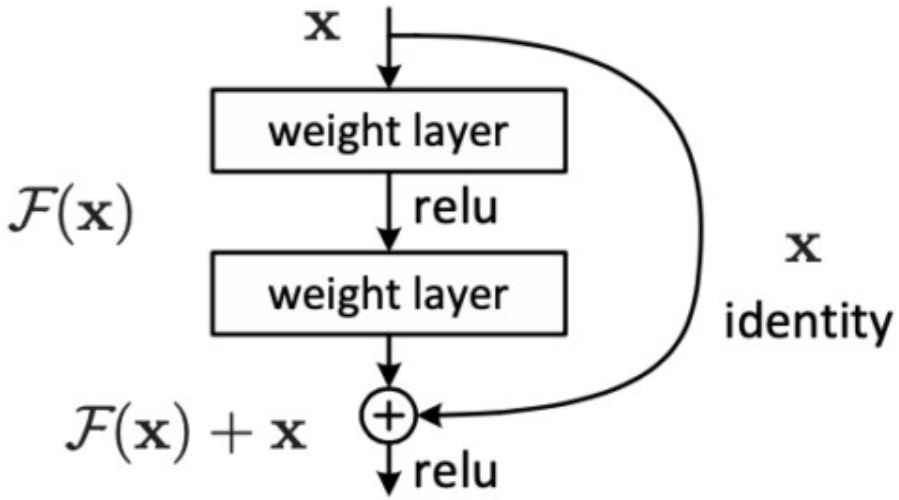
- ResNets solve this problem by introducing the concept of residual connections.
- A residual connection is a direct connection between the input of a layer and the output of the layer.
- This means that the output of the layer is not simply the result of the convolution operation, but it is also the sum of the convolution operation and the input of the layer.
- This allows the gradient to flow more easily through the network, which makes it easier for the network to learn.

Residual Block

- Residual blocks are an important part of the ResNet architecture. In older architectures such as VGG16, convolutional layers are stacked with batch normalization and nonlinear activation layers such as ReLu between them.
- This method works with a small number of convolutional layers—the maximum for VGG models is around 19 layers.
- However, subsequent research discovered that increasing the number of layers could significantly improve CNN performance.

Residual Block

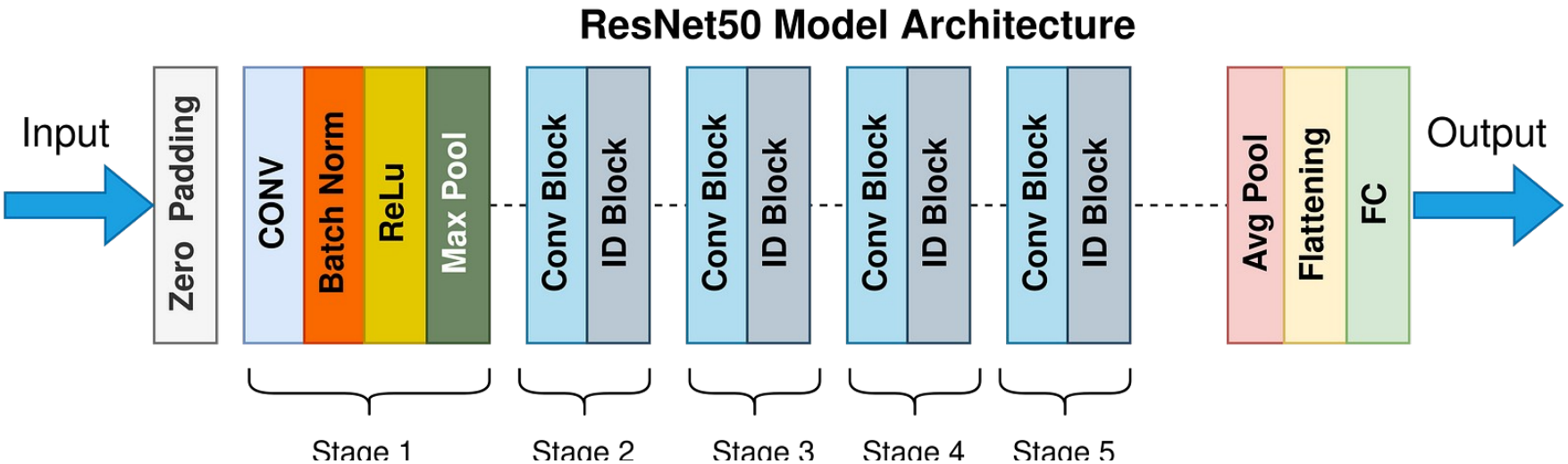
- The ResNet architecture introduces the simple concept of adding an intermediate input to the output of a series of convolution blocks. This is illustrated below.



ResNet

- ResNets have been very successful for image classification tasks.
- They have achieved state-of-the-art results on the ImageNet dataset, which is a large dataset of images with their corresponding labels.
- ResNets have also been used for other tasks, such as object detection and segmentation.

ResNet


ResNet50 Model Architecture
Input
Zero Padding
CONV
Batch Norm
ReLu
Max Pool
Conv Block
ID Block
Conv Block
ID Block
Conv Block
ID Block
Conv Block
ID Block
Avg Pool
Flattening
FC
Output
Stage 1
Stage 2
Stage 3
Stage 4
Stage 5

ResNet: Advantages

- They are very deep, which allows them to learn complex features.
- They are able to learn even when the gradients are very small.
- They have been shown to be very effective for image classification tasks.

ResNet: Disadvantages

- They can be computationally expensive to train.
- They can be difficult to understand and debug.

ResNet: Architectures

- Here are some of the most popular ResNet architectures:
  - ResNet50
  - ResNet101
  - ResNet152
  - ResNet200
  - ResNet34
  - ResNet18

Thank you

@mitu_skillologies

@mITuSkillologies

@mitu_group

@mitu-skillologies

@MITUSkillologies

kaggle

@mituskillologies

Web Resources
https://mitu.co.in
http://tusharkute.com

@mituskillologies

contact@mitu.co.in

tushar@tusharkute.com