

COMPUTER VISION

VGG-Net Architecture Explained

The company Visual Geometry Group created VGGNet (by Oxford University). While GoogLeNet won the classification assignment at ILSVR2014, this architecture came first. Understanding VGGNet is important since many contemporary image classification models are constructed on top of it.



Siddhesh Bangar · Follow

6 min read · Jun 28, 2022



Listen



Share

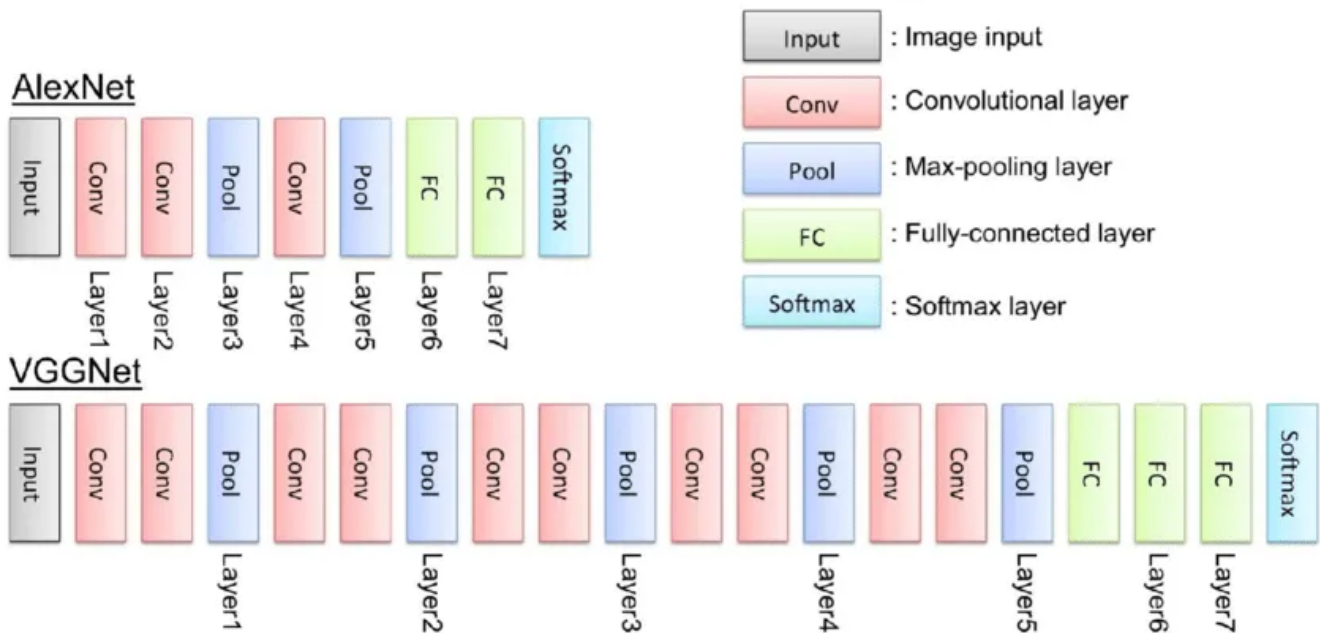


More

What is VGG-Net?

It is a typical deep Convolutional Neural Network (CNN) design with numerous layers, and the abbreviation VGG stands for Visual Geometry Group. The term “deep” describes the number of layers, with VGG-16 or VGG-19 having 16 or 19 convolutional layers, respectively.

Innovative object identification models are built using the VGG architecture. The VGGNet, created as a deep neural network, outperforms benchmarks on a variety of tasks and datasets outside of ImageNet. It also remains one of the most often used image recognition architectures today.



Distinguish Between AlexNet (above) and VGGNet (below) architectures

VGG-16

[Open in app ↗](#)



Search

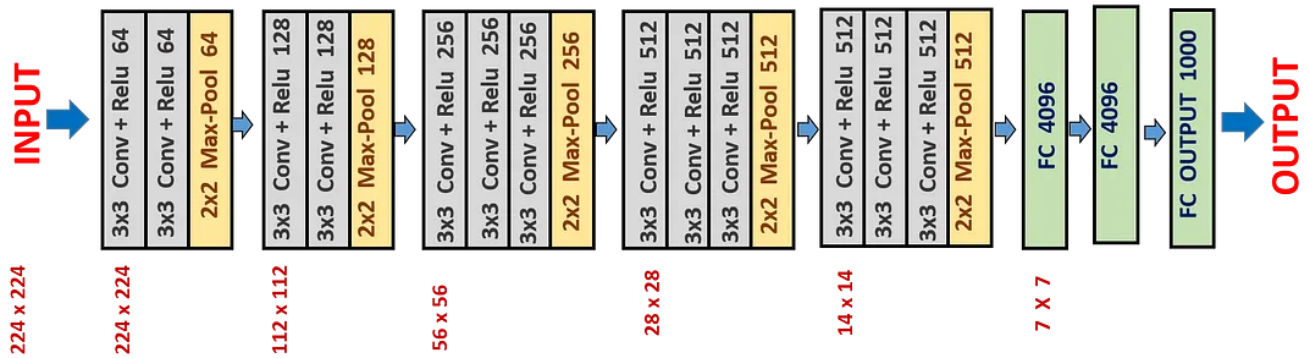


Convolutional Networks for Large-Scale Image Recognition contains the model that these researchers released.

In ImageNet, the VGG16 model achieves top-5 test accuracy of about 92.7 per cent. A dataset called ImageNet has over 14 million photos that fall into almost 1000 types. It was also among the most well-liked models submitted at ILSVRC-2014. It significantly outperforms AlexNet by substituting several 3x3 kernel-sized filters for the huge kernel-sized filters. Nvidia Titan Black GPUs were used to train the VGG16 model over many weeks.

The VGGNet-16 has 16 layers and can classify photos into 1000 different object categories, including keyboard, animals, pencil, mouse, etc., as discussed above. The model also accepts images with a resolution of 224 by 224.

VGG-16



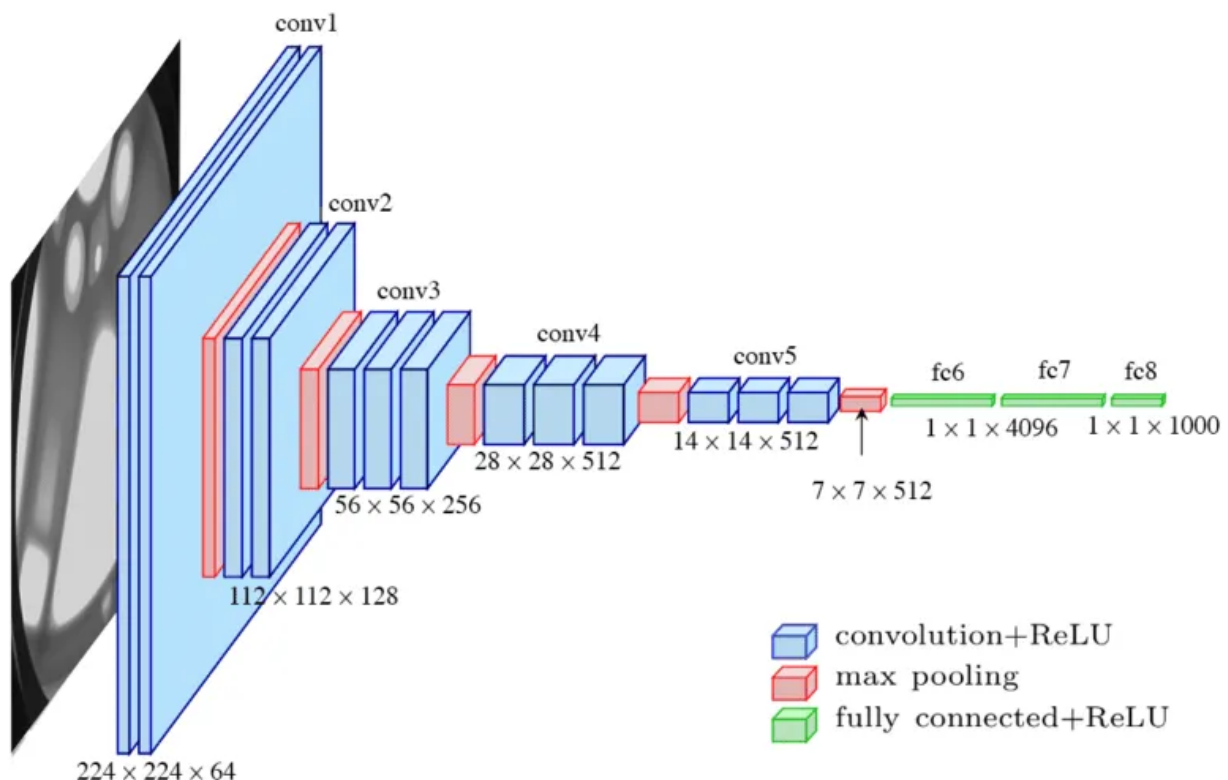
VGG 16 Architecture

VGG-19

The VGG19 model (also known as VGGNet-19) has the same basic idea as the VGG16 model, with the exception that it supports 19 layers. The numbers “16” and “19” refer to the model’s weight layers (convolutional layers). In comparison to VGG16, VGG19 contains three extra convolutional layers. In the final section of this essay, we’ll go into greater detail on the features of the VGG16 and VGG19 networks.

VGG-Net Architecture

Very tiny convolutional filters are used in the construction of the VGG network. Thirteen convolutional layers and three fully connected layers make up the VGG-16.



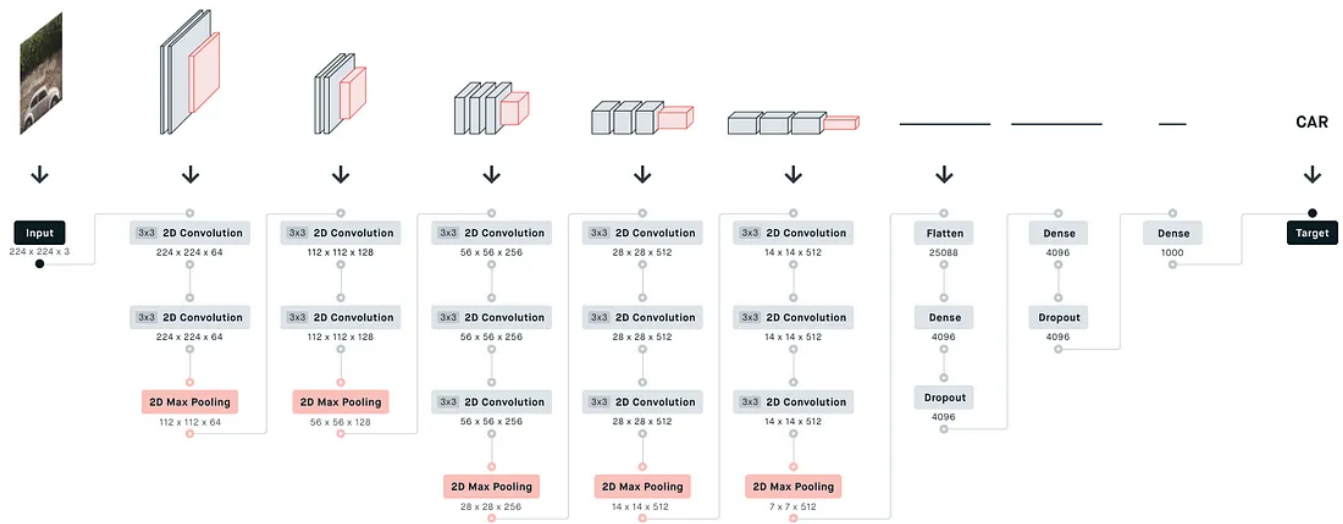
Let's quickly examine VGG's architecture:

- **Inputs:** The VGGNet accepts 224x224-pixel images as input. To maintain a consistent input size for the ImageNet competition, the model's developers chopped out the central 224x224 patches in each image.
- **Convolutional Layers:** VGG's convolutional layers use the smallest feasible receptive field, or 3x3, to record left-to-right and up-to-down movement. Additionally, 11 convolution filters are used to transform the input linearly. The next component is a ReLU unit, a significant advancement from AlexNet that shortens training time. Rectified linear unit activation function, or ReLU, is a piecewise linear function that, if the input is positive, outputs the input; otherwise, the output is zero. The convolution stride is fixed at 1 pixel to keep the spatial resolution preserved after convolution (stride is the number of pixel shifts over the input matrix).
- **Hidden Layers:** The VGG network's hidden layers all make use of ReLU. Local Response Normalization (LRN) is typically not used with VGG as it increases memory usage and training time. Furthermore, it doesn't increase overall accuracy.
- **Fully Connected Layers:** The VGGNet contains three layers with full connectivity. The first two levels each have 4096 channels, while the third layer has 1000 channels with one channel for each class.

Understanding VGG-16

The deep neural network's 16 layers are indicated by the number 16 in their name, which is VGG (VGGNet). This indicates that the VGG16 network is quite large, with a total of over 138 million parameters. Even by today's high standards, it is a sizable network. The network is more appealing due to the simplicity of the VGGNet16 architecture, nevertheless. Its architecture alone can be used to describe how uniform it is.

The height and width are decreased by a pooling layer that comes after a few convolution layers. There are around 64 filters available, which we can then multiply by two to get about 128 filters, and so on up to 256 filters. In the last layer, we can use 512 filters.



VGG 16 Architecture

VGG Configuration, Training, and Results

The VGG network has five configurations named A to E. The depth of the configuration increases from left (A) to right (B), with more layers added. Below is a table describing all the potential network architectures:

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

All configurations adhere to the same design and simply differ in depth; for example, network A has 11 weight layers (8 convolutional and 3 fully-connected layers), whereas network E has 19 weight layers (16 convolutional and 3 fully-connected layers). Convolutional layers have a relatively low number of channels, starting at 64 in the first layer and rising by a factor of 2 after each max-pooling layer to a maximum of 512. The following picture displays the overall amount of parameters (in millions):

Network	A,A-LRN	B	C	D	E
Number of parameters	133	133	134	138	144

VGG-16 Summary

	Layer	Feature Map	Size	Kernel Size	Stride	Activation
Input	Image	1	224 x 224 x 3	-	-	-
1	2 X Convolution	64	224 x 224 x 64	3x3	1	relu
	Max Pooling	64	112 x 112 x 64	3x3	2	relu
3	2 X Convolution	128	112 x 112 x 128	3x3	1	relu
	Max Pooling	128	56 x 56 x 128	3x3	2	relu
5	2 X Convolution	256	56 x 56 x 256	3x3	1	relu
	Max Pooling	256	28 x 28 x 256	3x3	2	relu
7	3 X Convolution	512	28 x 28 x 512	3x3	1	relu
	Max Pooling	512	14 x 14 x 512	3x3	2	relu
10	3 X Convolution	512	14 x 14 x 512	3x3	1	relu
	Max Pooling	512	7 x 7 x 512	3x3	2	relu
13	FC	-	25088	-	-	relu
14	FC	-	4096	-	-	relu
15	FC	-	4096	-	-	relu
Output	FC	-	1000	-	-	Softmax

Summary Table of VGG 16

Limitations Of VGG 16:

- It is very slow to train (the original VGG model was trained on Nvidia Titan GPU for 2–3 weeks).
- The size of VGG-16 trained imageNet weights is 528 MB. So, it takes quite a lot of disk space and bandwidth which makes it inefficient.
- 138 million parameters lead to exploding gradients problem.

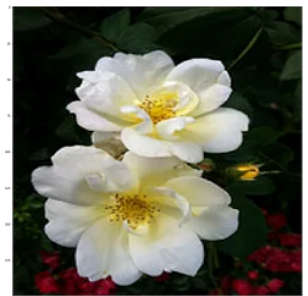
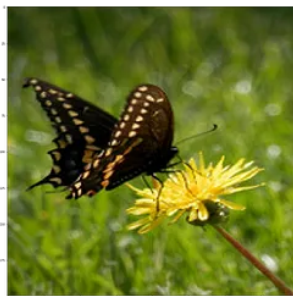
Implementation of VGGNet-16

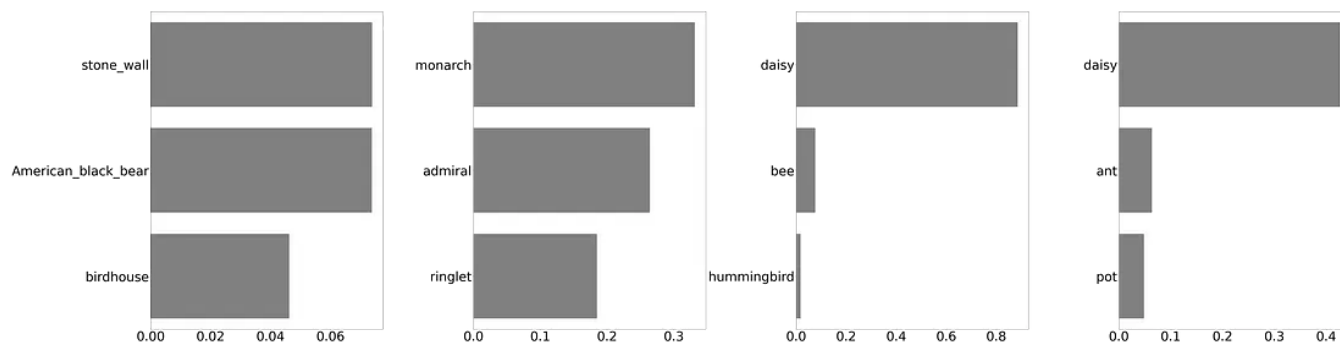
Importing Libraries, Defining the input image shape and Building VGG-16 Model

Working with pretrained model

- Keras library also provides the pre-trained model in which one can load the saved model weights, and use them for different purposes: transfer learning, image feature extraction, and object detection. We can load the model architecture given in the library, and then add all the weights to the respective layers.
- Before using the pre-trained models, let's write a few functions which will be used to make some predictions. First, load some images and pre-process them.

Using pretrained weights to save some time.





Code Credits go to — [VGGNet-16 Architecture: A Complete Guide by Paras Varshney](#)

And with that, we have completed the VGG-Net Architecture, if anyone would like to understand this architecture deeply can check out the paper which was published (The link has been added in the introduction part of this blog). Keep Learning.

Computer Vision

Artificial Intelligence

Machine Learning

Deep Learning

Technology



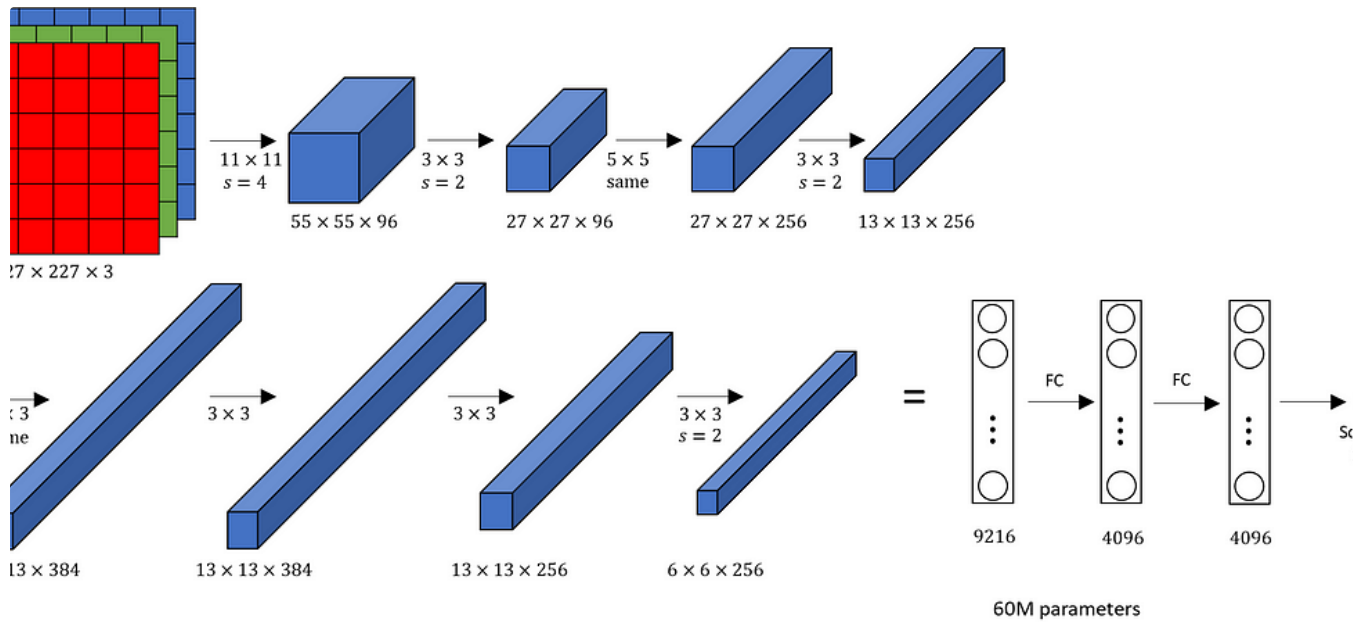
Follow

Written by Siddhesh Bangar

120 Followers

Computer Science Student @Trinity College Dublin | Exploring Jobs in Tech | Queries siddheshb008@gmail.com

More from Siddhesh Bangar



Siddhesh Bangar

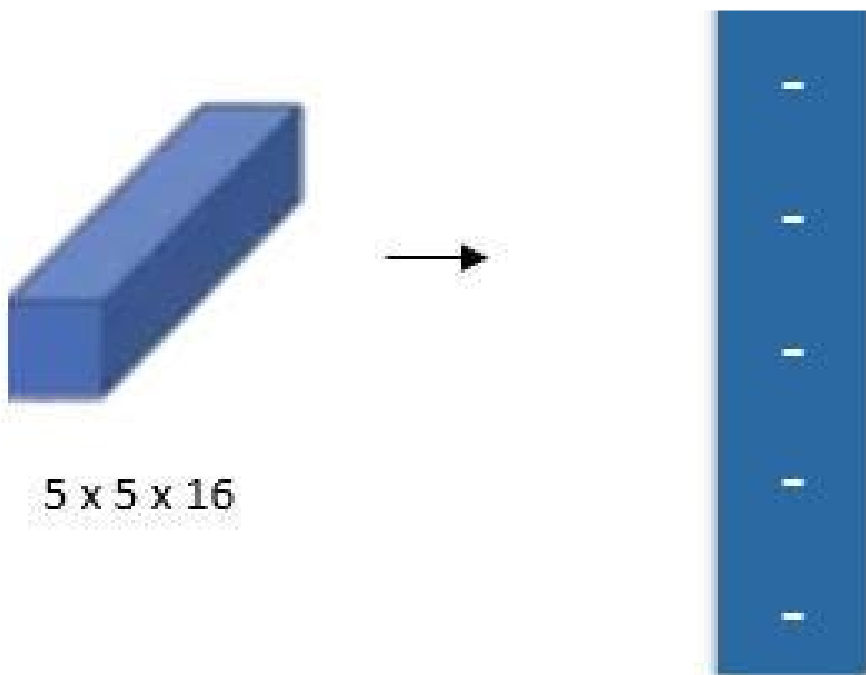
AlexNet Architecture Explained

The convolutional neural network (CNN) architecture known as AlexNet was created by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton...

6 min read · Jun 24, 2022



40



Siddhesh Bangar

LeNet 5 Architecture Explained

In the 1990s, Yann LeCun, Leon Bottou, Yosuha Bengio, and Patrick Haffner proposed the LeNet-5 neural network design for character...

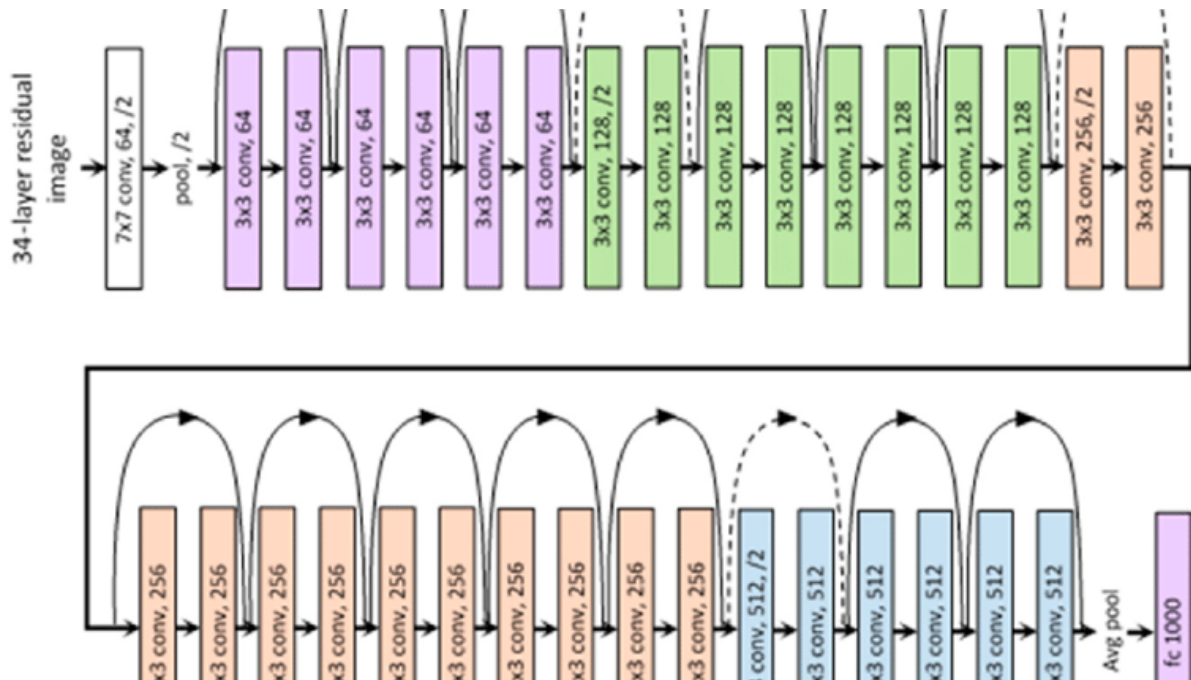
6 min read · Jun 22, 2022



94



1



Siddhesh Bangar

Resnet Architecture Explained

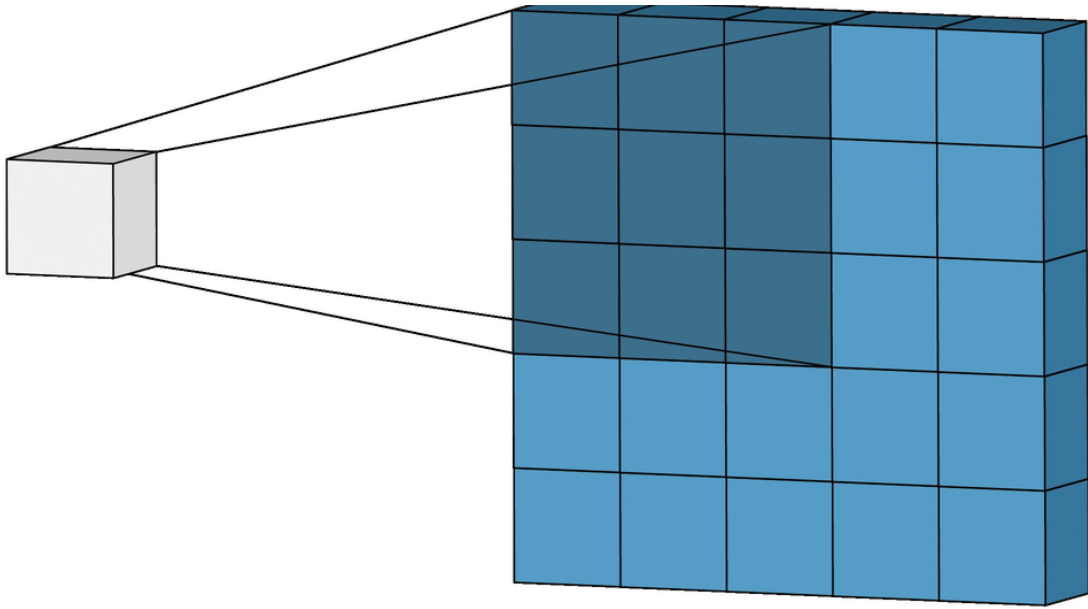
In their 2015 publication “Deep Residual Learning for Image Recognition,” Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun created a...

5 min read · Jul 5, 2022



27





Siddhesh Bangar

Understanding Convolution Neural Networks (Part 1)

Convolutional Neural Networks (ConvNets or CNNs) are a category of Neural Networks that have proven very effective in areas such as image...

8 min read · Jun 16, 2022




30



See all from Siddhesh Bangar

Recommended from Medium



 Muhammadshabrozshahab

Comparison and Architecture of Pre-Trained Models (AlexNet, VGG-16, VGG-19, ResNet, GoogleNet...

AlexNet

3 min read · Oct 23, 2023



 Azeem - I

Understanding ResNet Architecture: A Deep Dive into Residual Neural Network

Residual Network is a deep Learning model used for computer vision applications. The ResNet (Residual Neural Network) architecture was...

6 min read · Nov 14, 2023



12

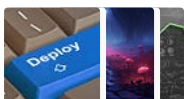


Lists



AI Regulation

6 stories · 311 saves



Predictive Modeling w/ Python

20 stories · 892 saves



Natural Language Processing

1184 stories · 657 saves



ChatGPT prompts

37 stories · 1104 saves




Vaibhav Rastogi

Residual Networks (ResNet)

ResNet, which stands for Residual Networks, is a type of convolutional neural network (CNN) architecture that was introduced to address the...

2 min read · Sep 8, 2023



 Tahera Firdose

Transfer Learning with MobileNet

★ · 11 min read · Aug 27, 2023




 noplaxochia

Pytorch ResNet from Scratch

Pytorch implementation of

1 min read · Jan 16



 Sandaruwan Herath in Data Science and Machine Learning

AlexNet

In 2012, image recognition saw its blurry world come into razor-sharp focus with the arrival of AlexNet, a game-changing deep neural...

5 min read · Dec 31, 2023



See more recommendations