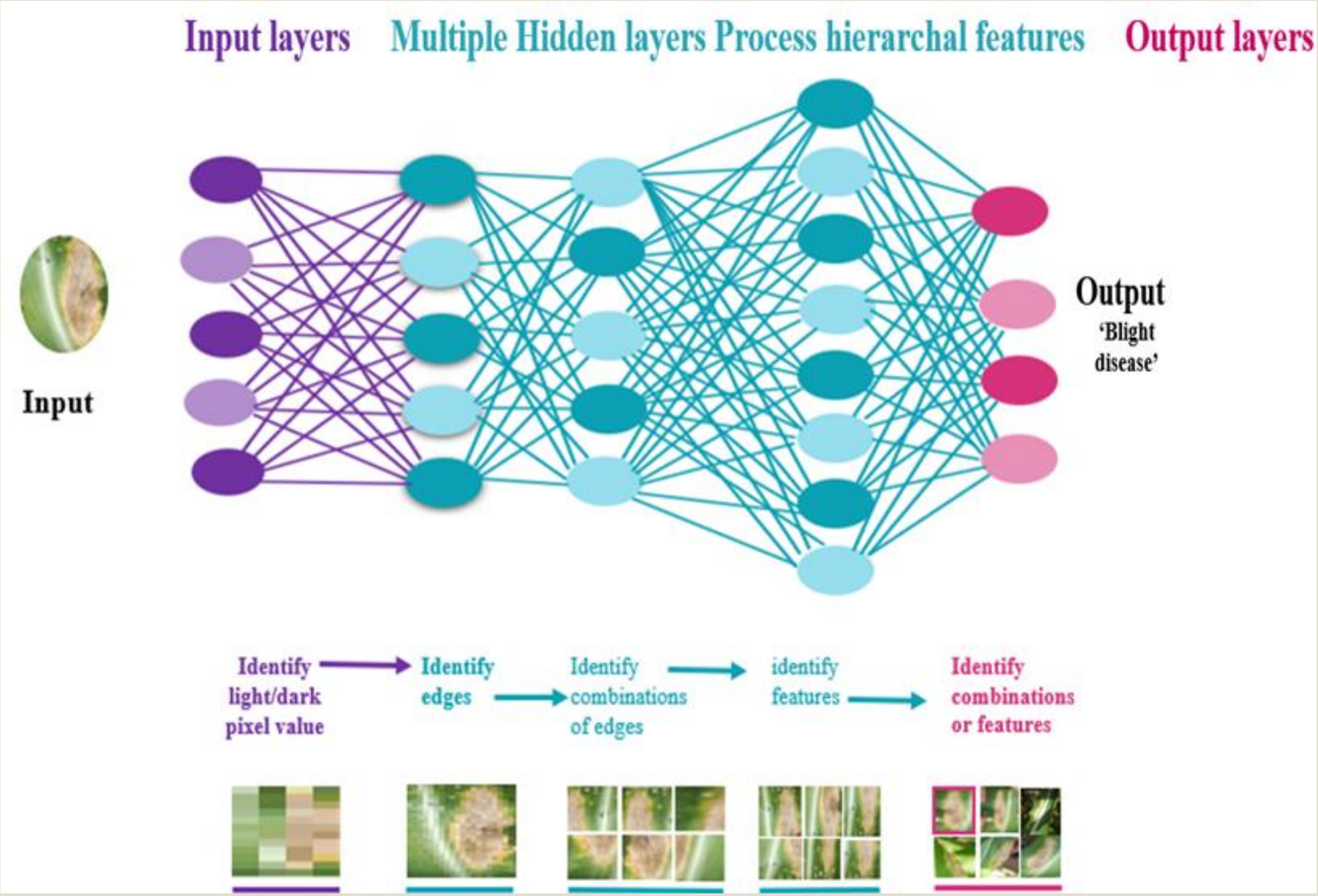
CNN ARCHITECTURES

Convolutional Neural Network



LeNet – First CNN Architecture

- LeNet was developed in 1998 by Yann LeCun, Corinna Cortes, and Christopher Burges for handwritten digit recognition problems.
- LeNet was one of the first successful CNNs and is often considered the “Hello World” of deep learning.
- It is one of the earliest and most widely-used CNN architectures and has been successfully applied to tasks such as handwritten digit recognition.
- The LeNet architecture consists of multiple convolutional and pooling layers, followed by a fully-connected layer.
- The model has five convolution layers followed by two fully connected layers.
- LeNet was the beginning of CNNs in deep learning for computer vision problems.

Applications

The LeNet CNN is a simple yet powerful model that has been used for various tasks such as handwritten digit recognition, traffic sign recognition, and face detection.

LeNet CNN Architecture

LeNet-5, a pioneering convolutional network that classifies digits, was applied by several banks to recognize hand-written numbers on cheques digitized in 32x32 pixel greyscale input images. The ability to process higher resolution images requires larger and more convolutional layers, so this technique is constrained by the availability of computing resources.

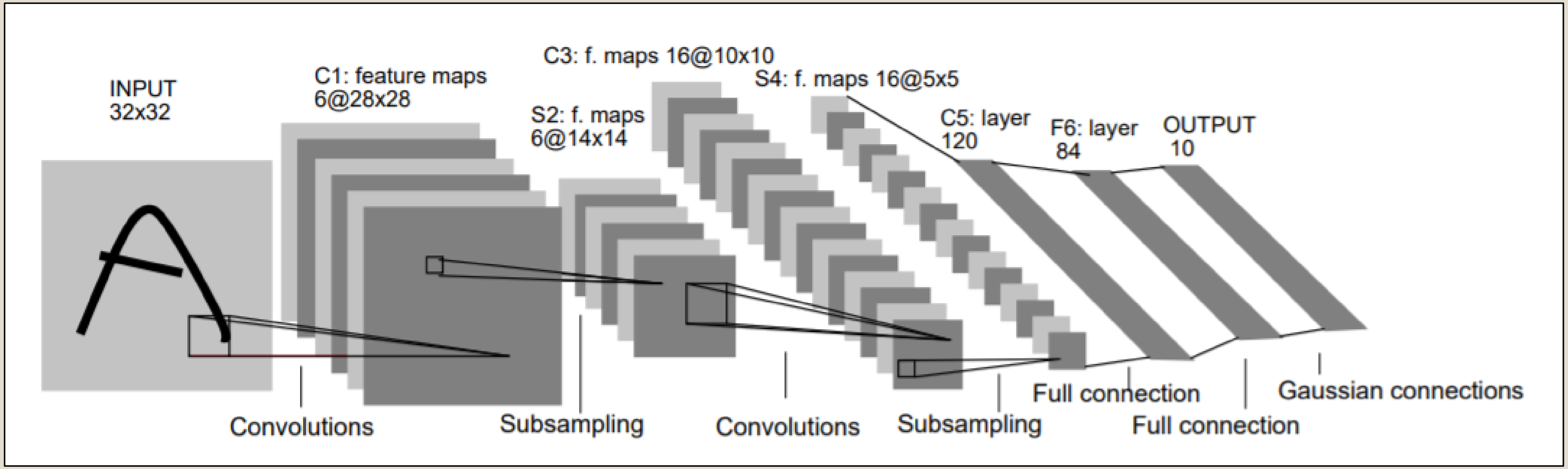
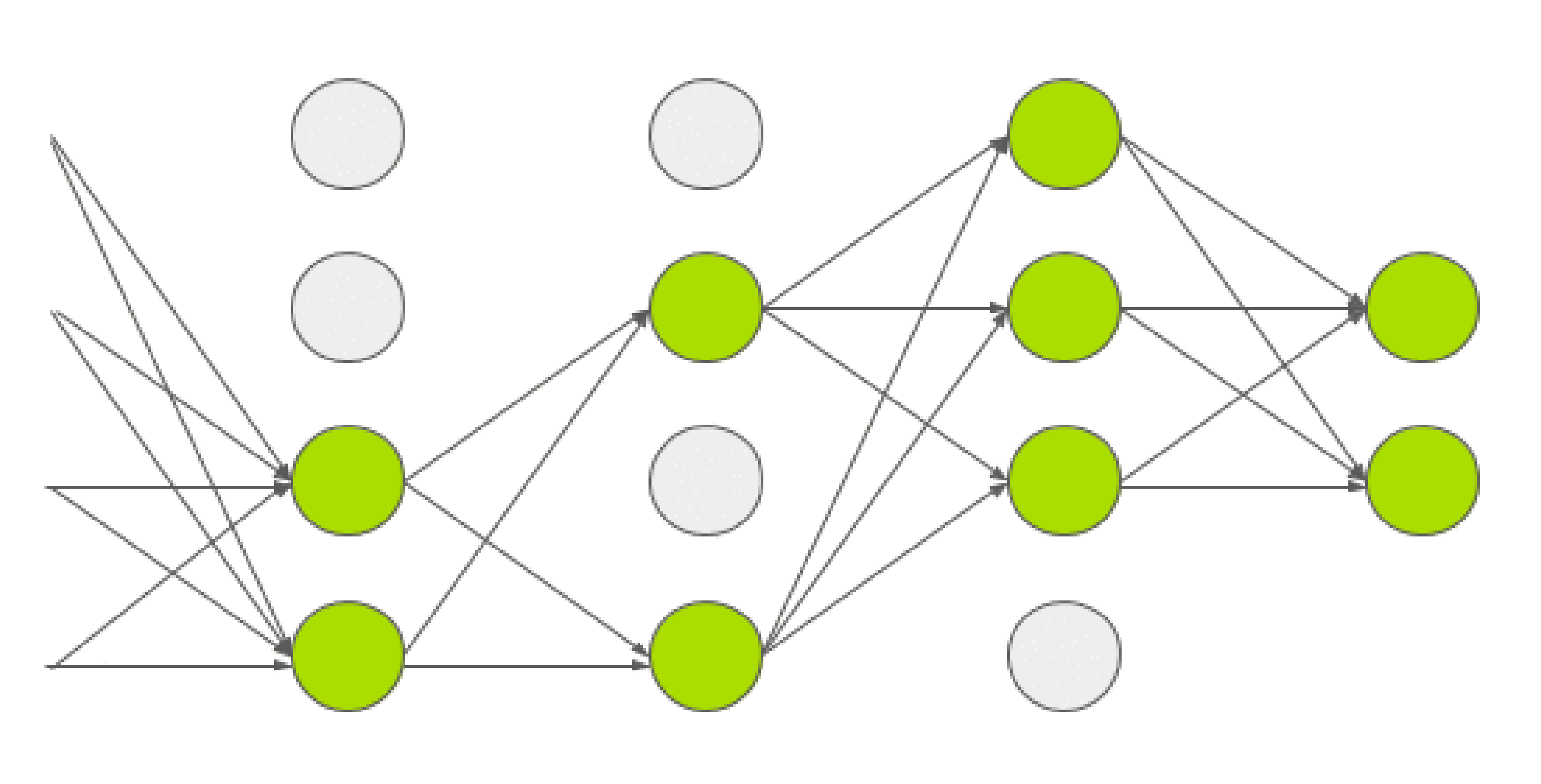


Figure 1: LeNet CNN architecture.

AlexNet – Deep Learning Architecture that popularized CNN

- AlexNet was developed by Alex Krizhevsky, Ilya Sutskever, and Geoff Hinton.
- AlexNet network had a very similar architecture to LeNet, but was deeper, bigger, and featured Convolutional Layers stacked on top of each other.
- AlexNet was the first large-scale CNN and was used to win the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012.
- The AlexNet architecture was designed to be used with large-scale image datasets and it achieved state-of-the-art results at the time of its publication.
- AlexNet is composed of 5 convolutional layers with a combination of max-pooling layers, 3 fully connected layers, and 2 dropout layers.
- The activation function used in all layers is Relu.
- The activation function used in the output layer is Softmax.
- The total number of parameters in this architecture is around 60 million.

Dropout Layer



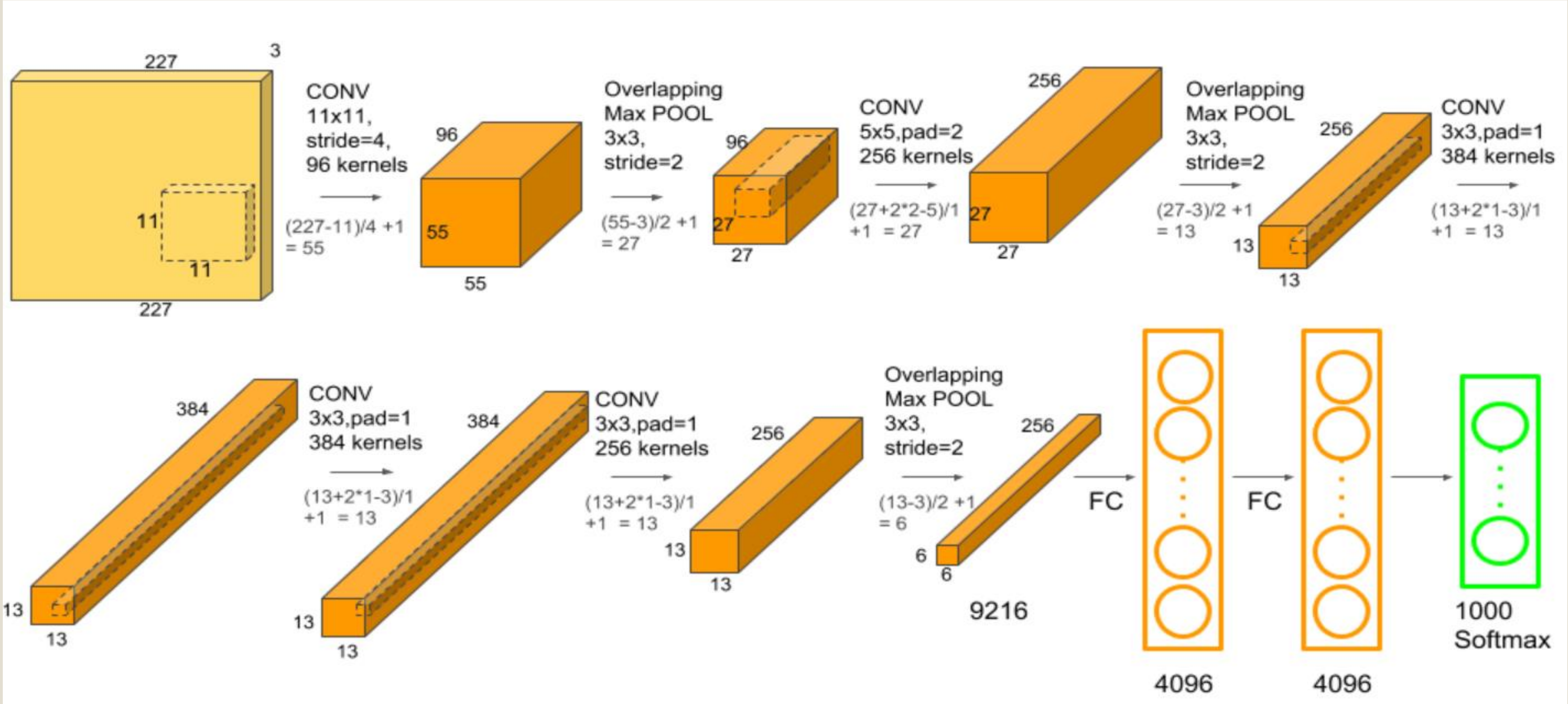


Figure 2: AlexNet CNN architecture.

GoogLeNet – CNN Architecture used by Google

- GoogLeNet is the CNN architecture used by Google to win ILSVRC 2014 classification task.
- It was developed by Jeff Dean, Christian Szegedy, Alexandro Szegedy et al..
- It has been shown to have a notably reduced error rate in comparison with previous winners AlexNet (Ilsvrc 2012 winner).
- It achieves deeper architecture by employing a number of distinct techniques, including 1×1 convolution and global average pooling.
- Real-world applications/examples of GoogLeNet CNN architecture include Street View House Number (SVHN) digit recognition task, which is often used as a proxy for roadside object detection.
- It used batch normalization, image distortions and RMSprop. This module is based on several very small convolutions in order to drastically reduce the number of parameters. Their architecture consisted of a 22 layer deep CNN

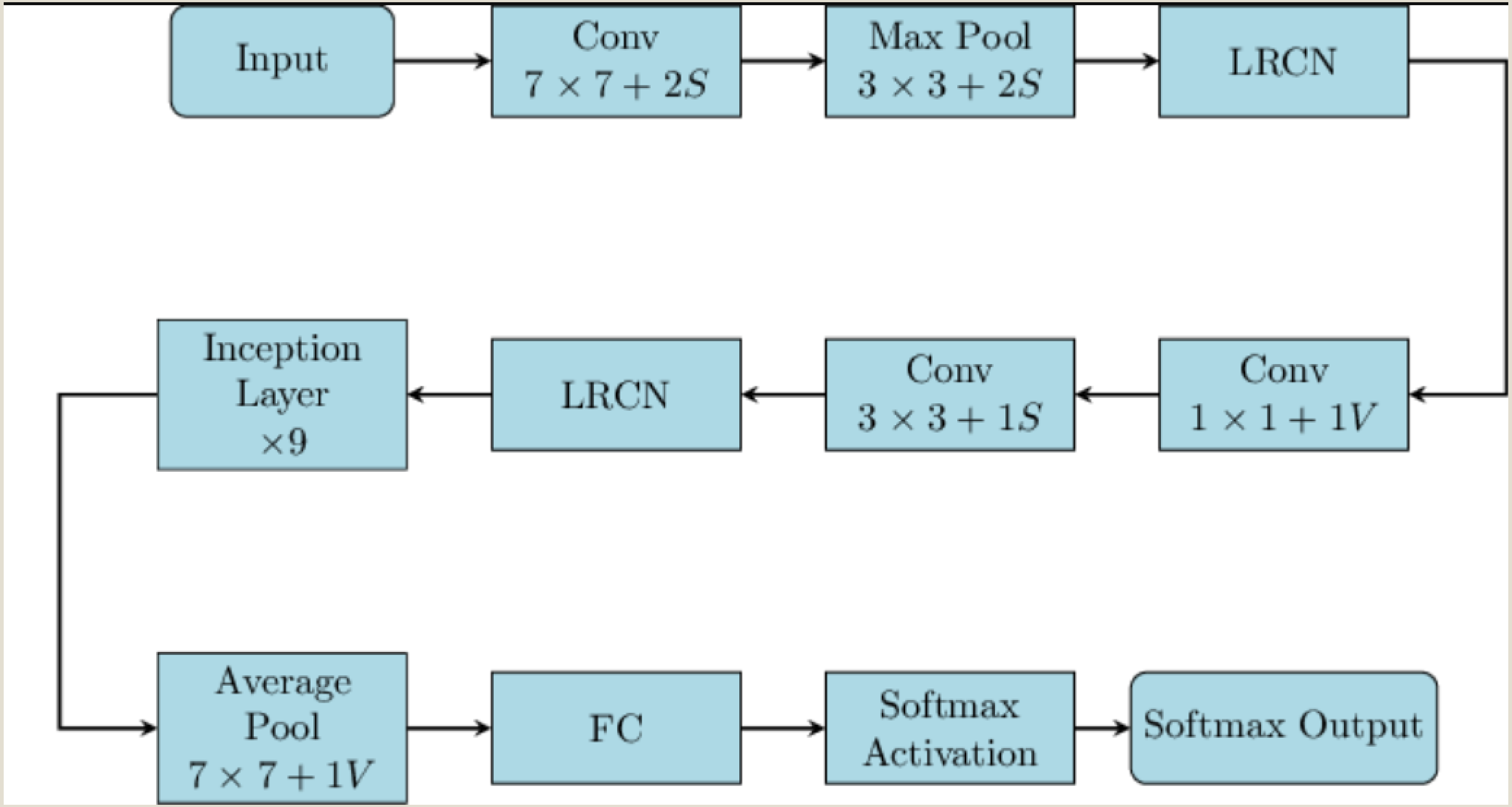


Figure 3: GoogleNet CNN architecture.

VGGNet – CNN Architecture with Large Filters

- VGGNet is the CNN architecture that was developed by Karen Simonyan, Andrew Zisserman et al. at Oxford University.
- The full form of "VGG16" stands for "Visual Geometry Group 16".
- This name comes from the Visual Geometry Group at the University of Oxford, where this neural network architecture was developed.
- The "16" in the name indicates that the model contains 16 layers that have weights; this includes convolutional layers as well as fully connected layers.
- VGGNet is a 16-layer CNN with up to 95 million parameters and trained on over one billion images (1000 classes).
- It can take large input images of 224 x 224-pixel size for which it has 4096 convolutional features.
- The VGG CNN model is computationally efficient and serves as a strong baseline for many applications in computer vision due to its applicability for numerous tasks including object detection. Its deep feature representations are used across multiple neural network architectures like YOLO, SSD, etc.

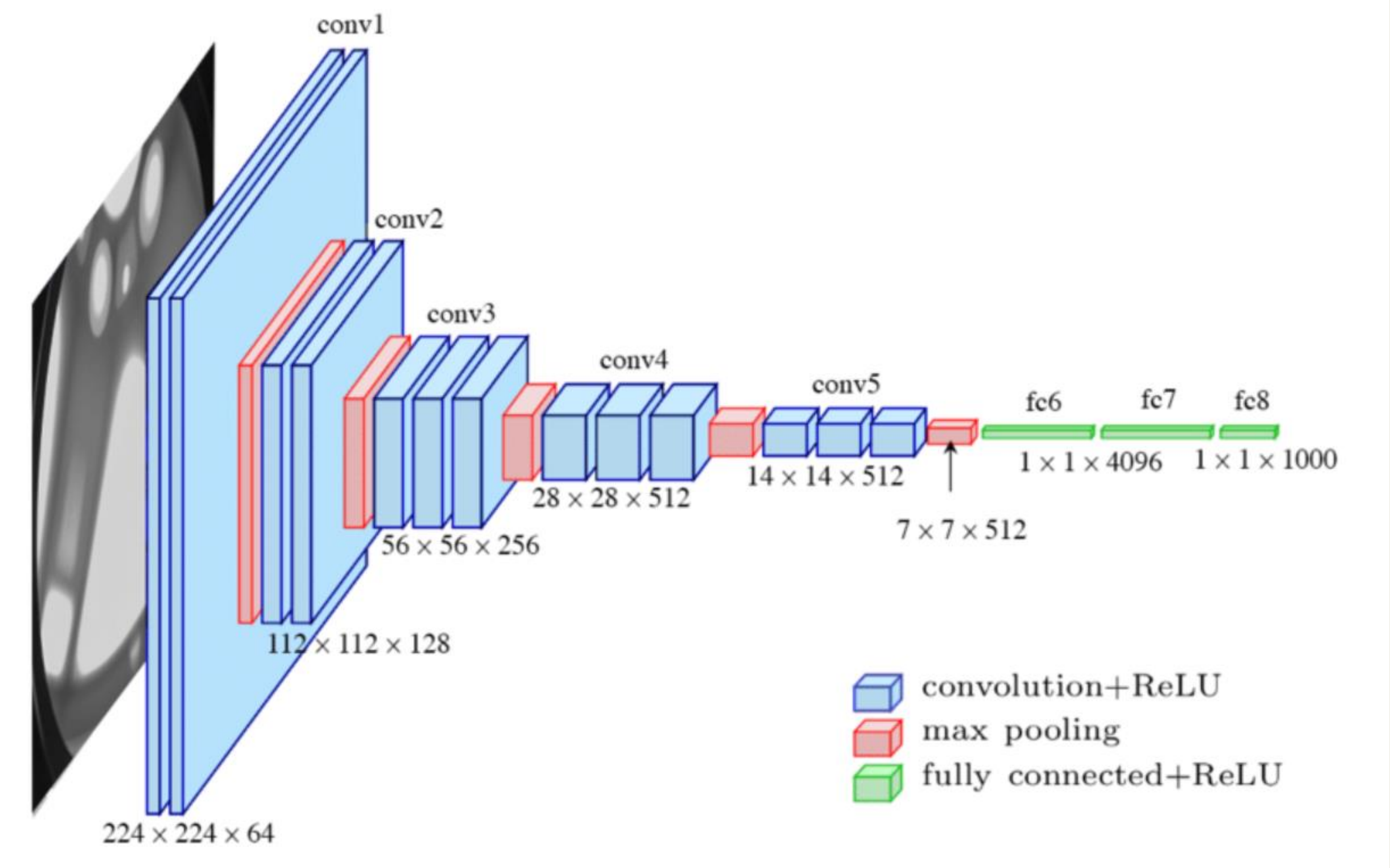


Figure 4: VGGNet CNN architecture.

MobileNets – CNN Architecture for Mobile Devices

- MobileNets are CNNs that can be fit on a mobile device to classify images or detect objects with low latency.
- MobileNets have been developed by Andrew G Trillion et al..
- They are usually very small CNN architectures, which makes them easy to run in real-time using embedded devices like smartphones and drones.
- The architecture is also flexible so it has been tested on CNNs with 100-300 layers and it still works better than other architectures like VGGNet.
- Real-life examples of MobileNets CNN architecture include CNNs that is built into Android phones to run Google's Mobile Vision API, which can automatically identify labels of popular objects in images.

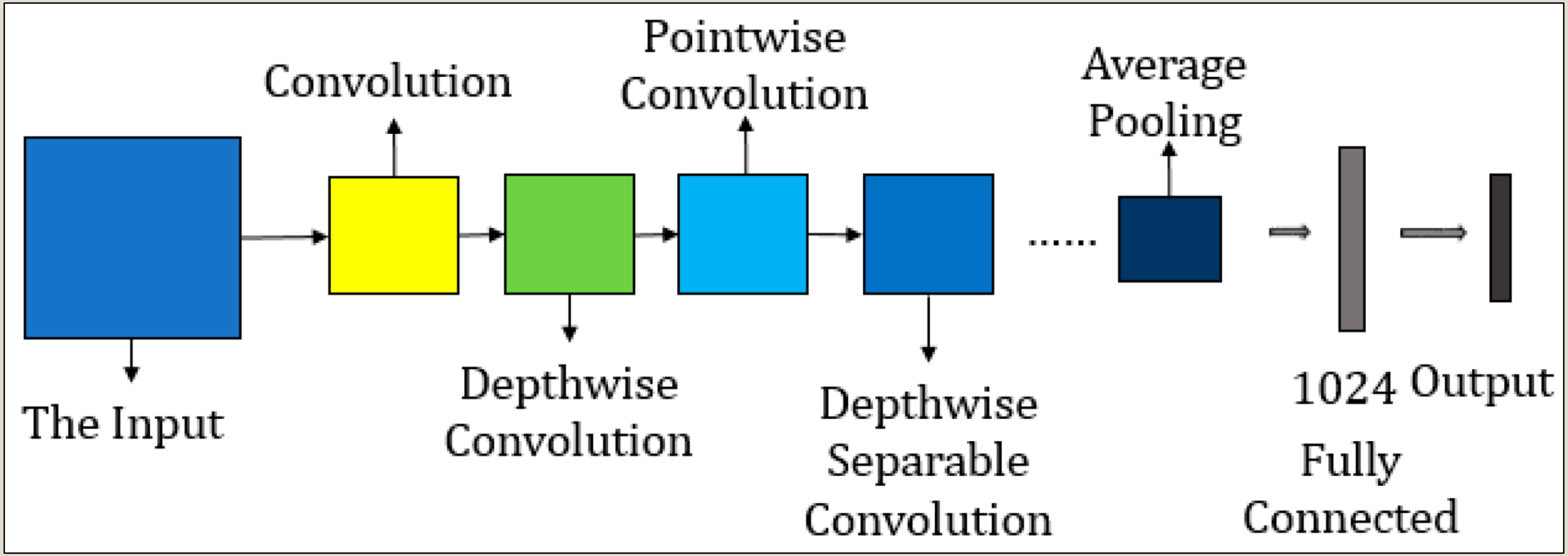


Figure 5: MobileNets CNN architecture.

Depth-wise convolution

- The depth-wise convolutions are used to apply a single filter into each input channel. This is different from a standard convolution in which the filters are applied to all of the input channels.

Point-wise convolution

- Since the depth wise convolution is only used to filter the input channel, it does not combine them to produce new features. So an additional layer called pointwise convolution layer is made, which computes a linear combination of the output of depth wise convolution using a 1×1 convolution.

Helping Websites

- https://medium.com/
- https://towardsdatascience.com/
- https://machinelearningmastery.com/