LeNet: Convolutional Neural Network

Introduction to LeNet

LeNet is one of the earliest convolutional neural networks (CNNs) developed by Yann LeCun in 1989. It was designed for handwritten digit recognition and laid the foundation for modern deep learning.

Architecture of LeNet

LeNet consists of 7 layers, including:

1. Convolutional Layers

2. Subsampling (Pooling) Layers

3. Fully Connected Layers

4. Output Layer

It follows a hierarchical feature extraction approach.

Layers of LeNet

- Conv Layer 1: Extracts features from input images
- Pooling Layer 1: Reduces dimensions
- Conv Layer 2: Further feature extraction
- Pooling Layer 2: Reduces dimensions again
- Fully Connected Layers: Classifies features into categories

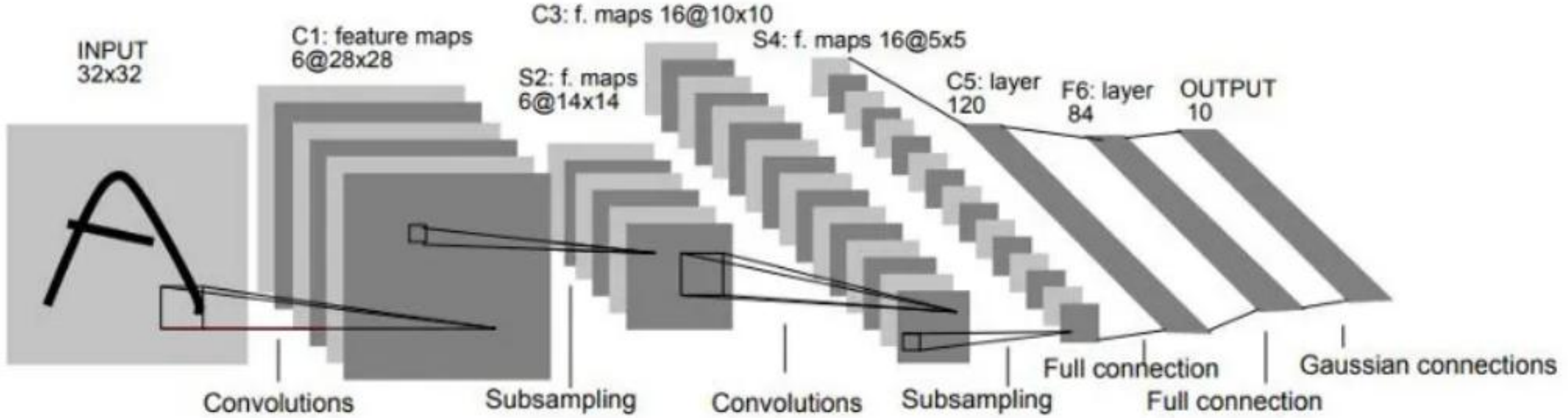
Layer #	Type	Details
1	Input Layer	32×32 grayscale image (MNIST images were 28×28 but padded to 32×32)
2	Convolutional Layer (C1)	6 filters of size 5×5, stride 1 → Output: 28×28×6
3	Subsampling Layer (S2)	Average Pooling (2×2), stride 2 → Output: 14×14×6
4	Convolutional Layer (C3)	16 filters of size 5×5, stride 1 → Output: 10×10×16
5	Subsampling Layer (S4)	Average Pooling (2×2), stride 2 → Output: 5×5×16
6	Fully Connected Layer (C5)	120 neurons, each connected to all 5×5×16 neurons from the previous layer
7	Output Layer (F6)	84 neurons (fully connected) → Final output: **10 softmax neurons** for classification

Features of LeNet-5

- Every convolutional layer includes three parts: convolution, pooling, and nonlinear activation functions
- Using convolution to extract spatial features (Convolution was called receptive fields originally)
- **The average pooling layer** is used for subsampling.
- ‘**tanh**’ is used as the activation function
- Using **Multi-Layered Perceptron** or **Fully Connected Layers** as the last classifier
- The sparse connection between layers reduces the complexity of computation

Architecture

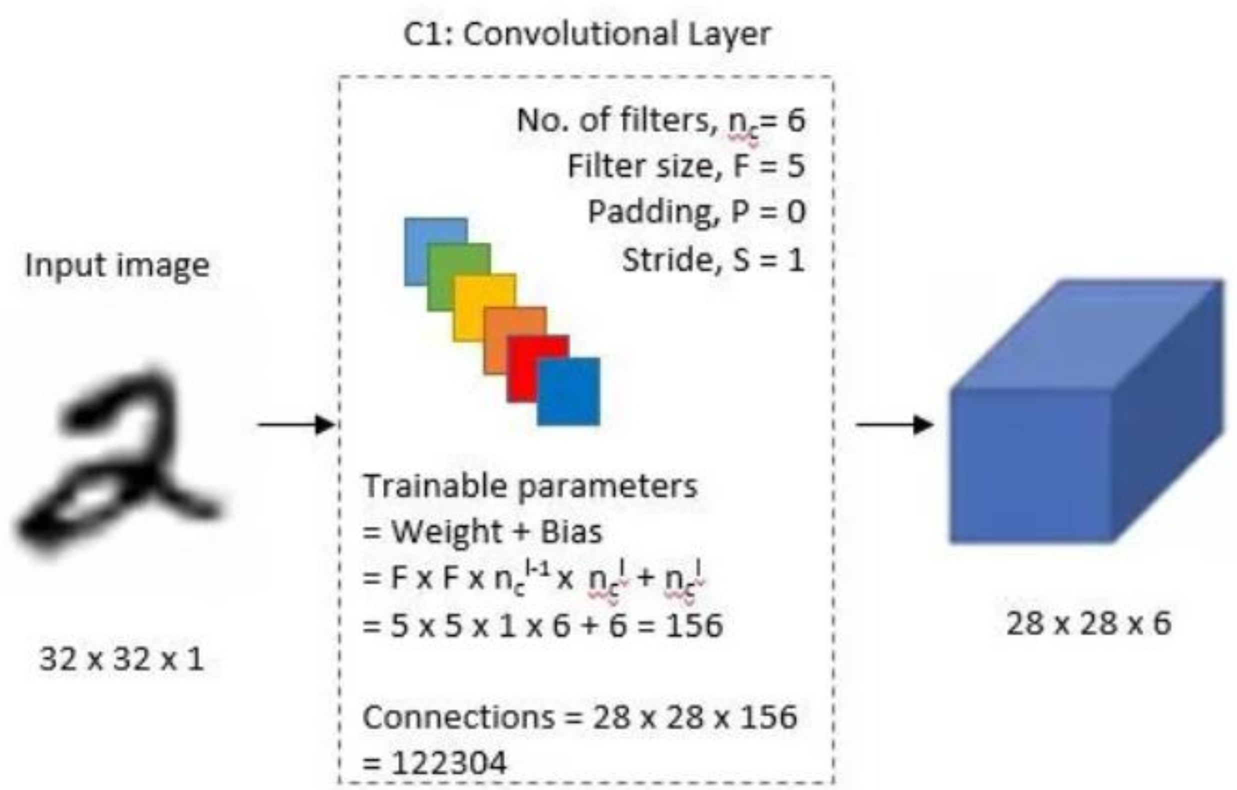
The LeNet-5 CNN architecture has seven layers. Three convolutional layers, two subsampling layers, and two fully linked layers make up the layer composition.



LeNet-5 Architecture

First Layer

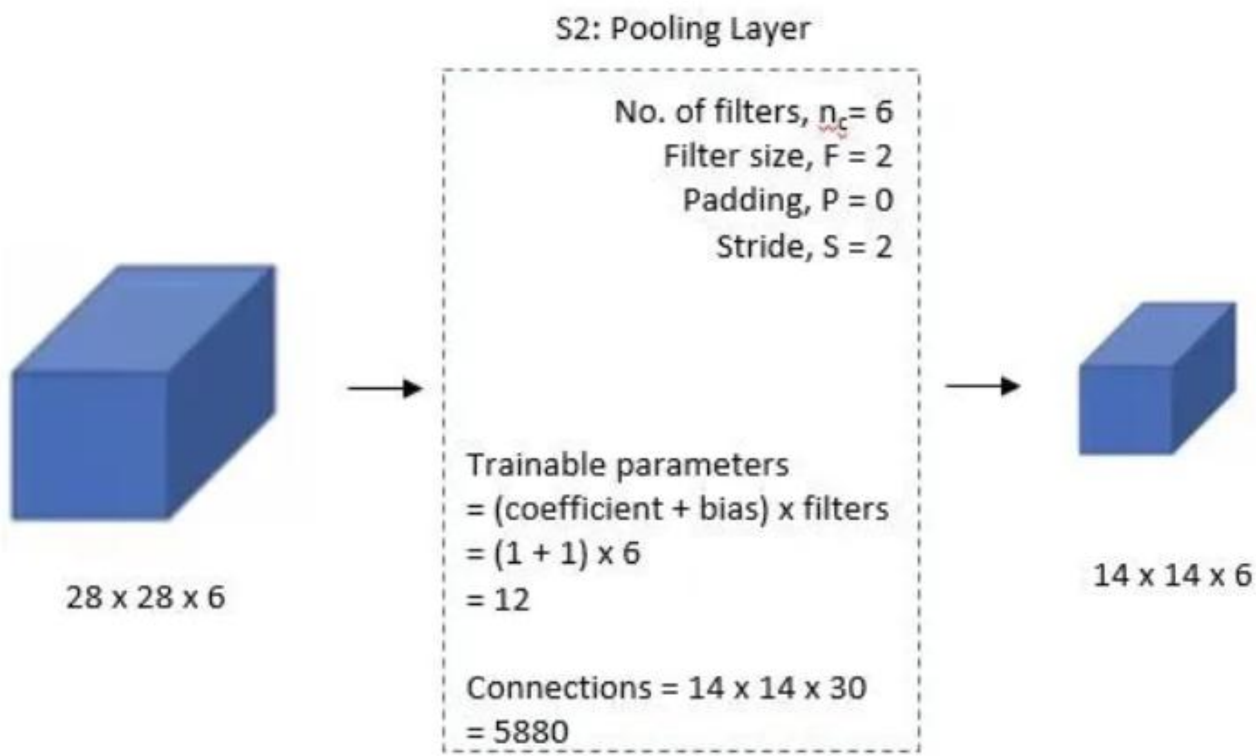
A 32x32 grayscale image serves as the input for LeNet-5 and is processed by the first convolutional layer comprising six feature maps or filters with a stride of one. From 32x32x1 to 28x28x6, the image's dimensions shift.



First Layer

Second Layer

Then, using a filter size of 22 and a stride of 2, the LeNet-5 adds an average pooling layer or sub-sampling layer. 14x14x6 will be the final image's reduced size.



Second Layer

Third Layer

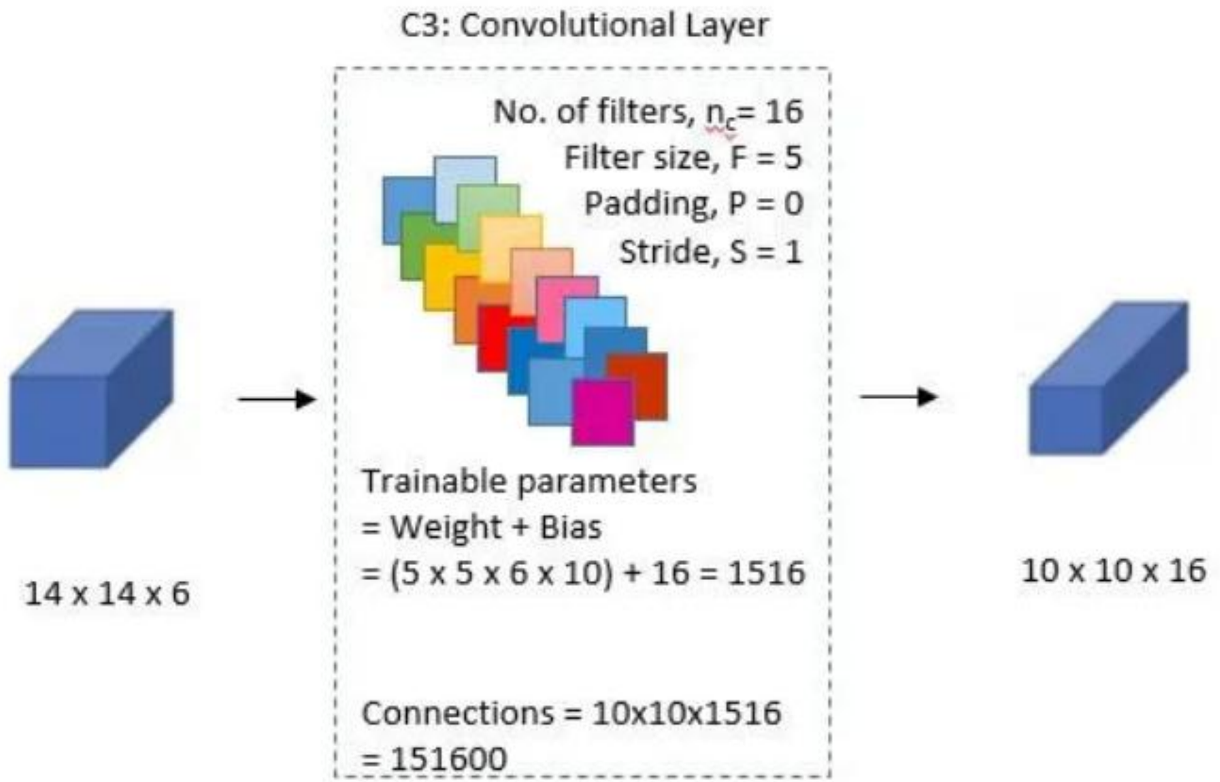
A second convolutional layer with 16 feature maps of size 55 and a stride of 1 is then present. Only 10 of the 16 feature maps in this layer are linked to the six feature maps in the layer below, as can be seen in the illustration below.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	X				X	X	X			X	X	X	X		X	X
1	X	X				X	X	X			X	X	X	X		X
2	X	X	X				X	X	X			X		X	X	X
3		X	X	X			X	X	X	X			X		X	X
4			X	X	X			X	X	X	X		X	X		X
5				X	X	X			X	X	X	X		X	X	X

TABLE I

EACH COLUMN INDICATES WHICH FEATURE MAP IN S2 ARE COMBINED BY THE UNITS IN A PARTICULAR FEATURE MAP OF C3.

The primary goal is to disrupt the network's symmetry while maintaining a manageable number of connections. Because of this, there are 1516 training parameters instead of 2400 in these layers, and similarly, there are 151600 connections instead of 240000.

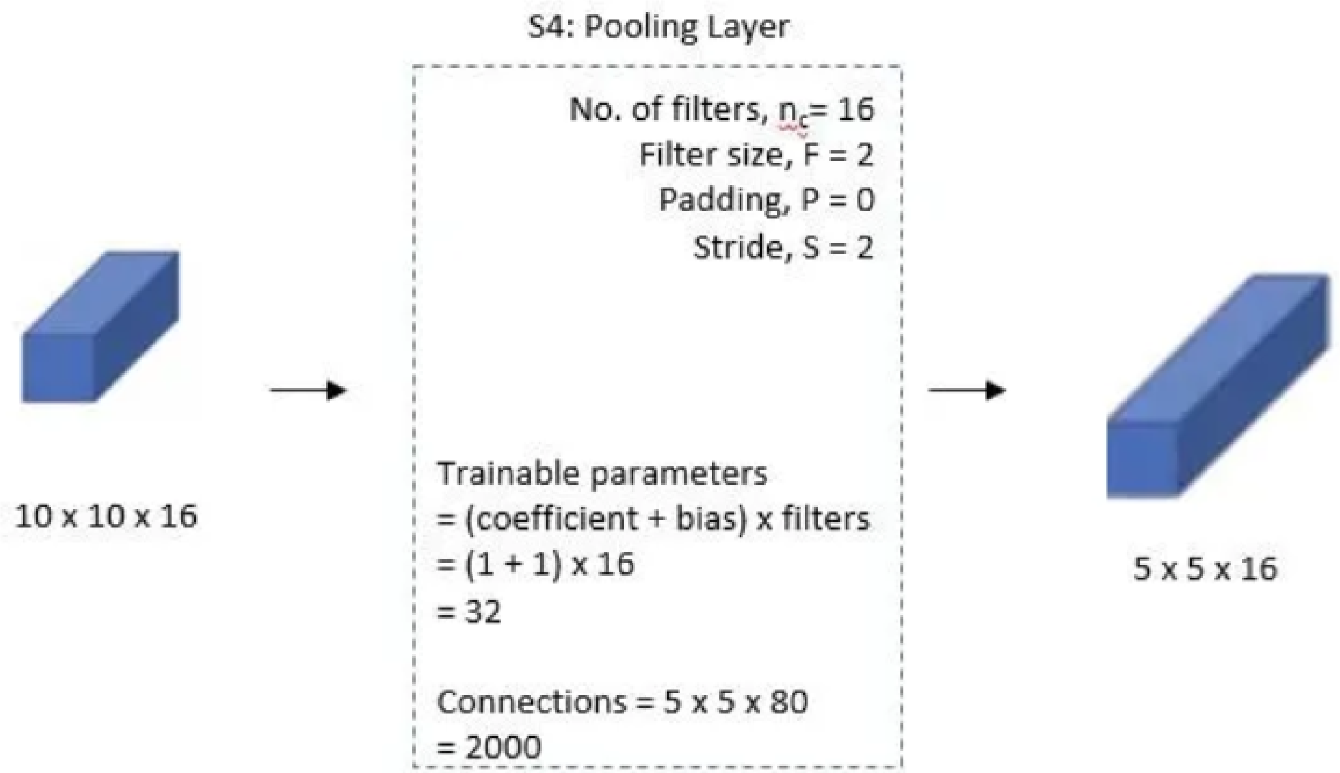


C3: Convolutional Layer
No. of filters, n_c= 16
Filter size, F = 5
Padding, P = 0
Stride, S = 1
Trainable parameters
= Weight + Bias
= (5 x 5 x 6 x 10) + 16 = 1516
Connections = 10x10x1516
= 151600
14 x 14 x 6
10 x 10 x 16

Third Layer

Fourth Layer

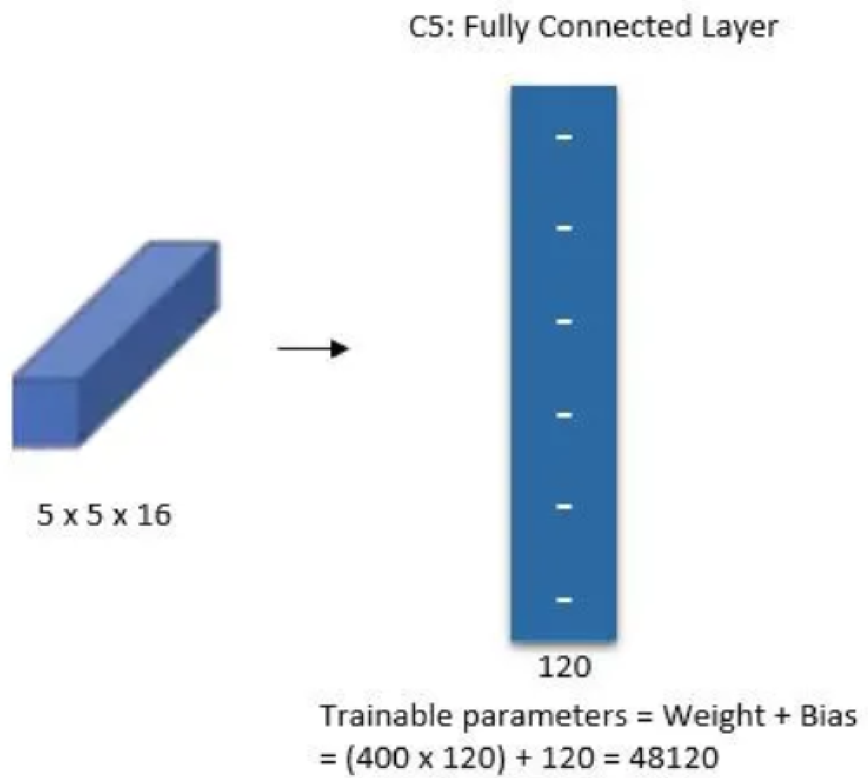
With a filter size of 22 and a stride of 2, the fourth layer (S4) is once more an average pooling layer. The output will be decreased to 5x5x16 because this layer is identical to the second layer (S2) but has 16 feature maps.



Fourth Layer

Fifth Layer

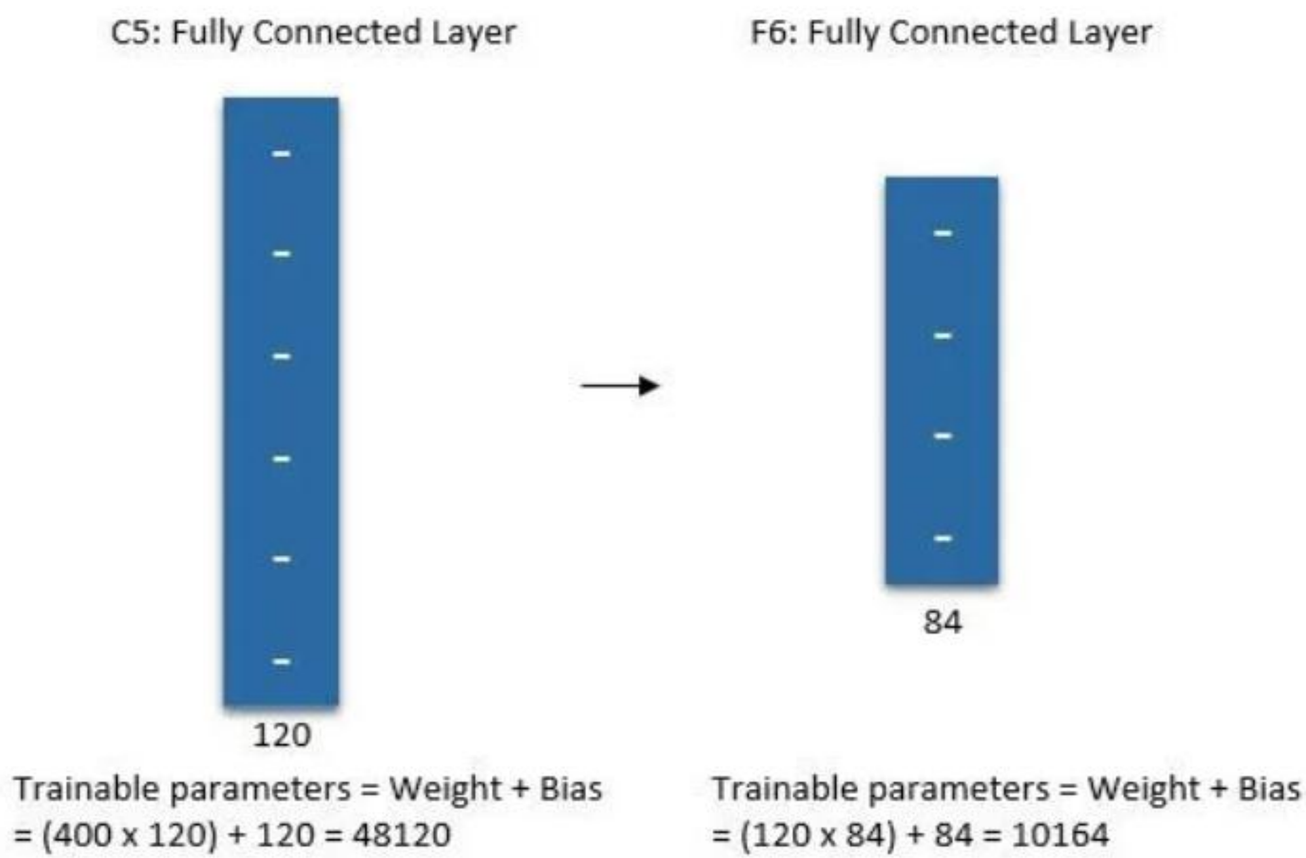
With 120 feature maps, each measuring 1 x 1, the fifth layer (C5) is a fully connected convolutional layer. All 400 nodes (5x5x16) in layer four, S4, are connected to each of the 120 units in C5's 120 units.



Fifth Layer

Sixth Layer

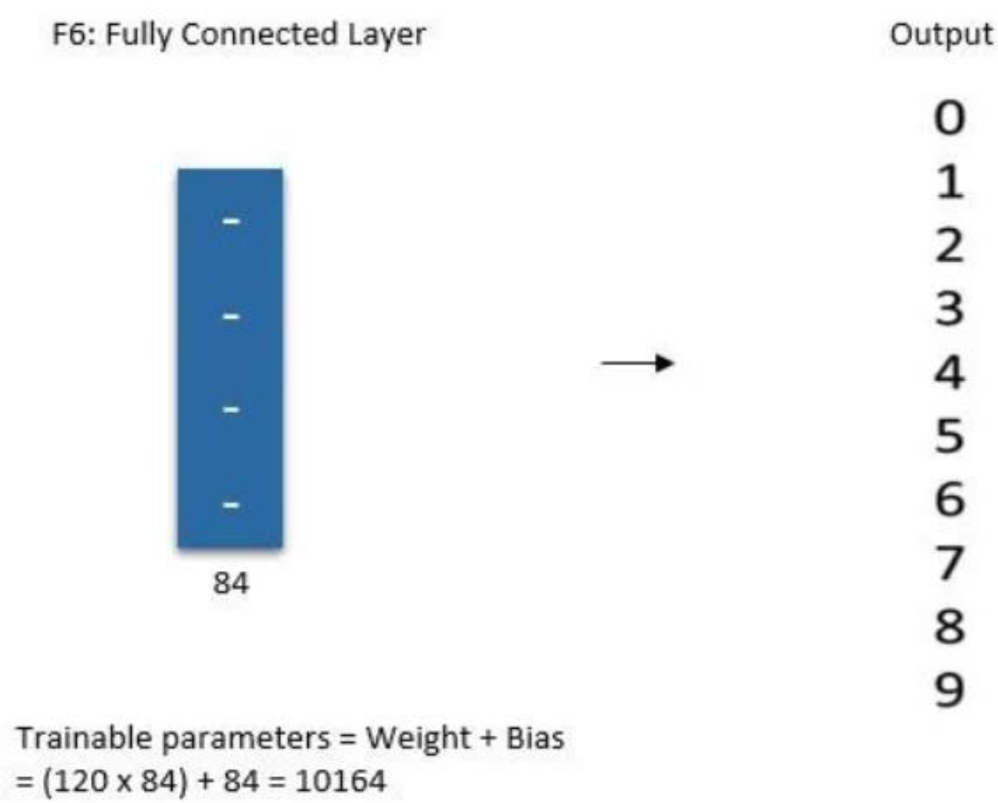
A fully connected layer (F6) with 84 units makes up the sixth layer.



Sixth Layer

Output Layer

The SoftMax output layer, which has 10 potential values and corresponds to the digits 0 to 9, is the last layer.



Output Layer

Summary of LeNet-5 Architecture

Layer		Feature Map	Size	Kernel Size	Stride	Activation
Input	Image	1	32x32	-	-	-
1	Convolution	6	28x28	5x5	1	tanh
2	Average Pooling	6	14x14	2x2	2	tanh
3	Convolution	16	10x10	5x5	1	tanh
4	Average Pooling	16	5x5	2x2	2	tanh
5	Convolution	120	1x1	5x5	1	tanh
6	FC	-	84	-	-	tanh
Output	FC	-	10	-	-	softmax

Summarized table for LeNet 5 Architecture

Applications of LeNet

LeNet is primarily used for:

- Handwritten digit recognition (MNIST dataset)

- OCR (Optical Character Recognition)

- Early deep learning experiments

- A foundation for modern CNN architectures

Summary

- LeNet was one of the first CNN architectures
- It uses convolution, pooling, and fully connected layers
- It played a crucial role in deep learning advancements
- Still used for educational purposes and basic image classification