SMAI Project Final-Eval

Gradient-Based learning applied to document recognition

Team 54

- Kshitijaa 2019115005
- Ihita 2021701007
- Ashuthosh 2019112003

Contents

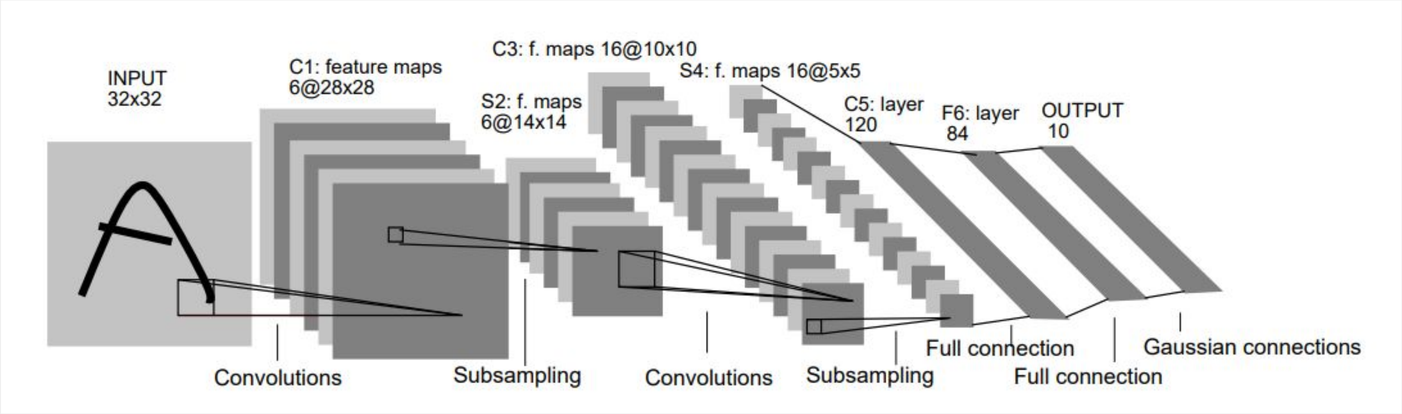
LeNet ?

What is LeNet?

Applications

What is LeNet?

- LeNet-5, 1996
- First generation CNN
- Preceded by 4 other architectures

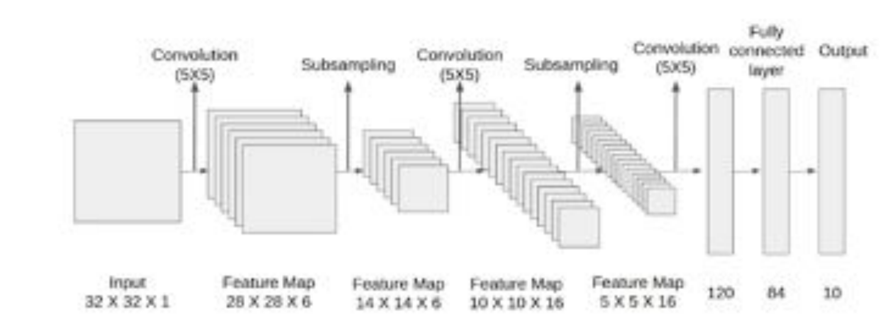


Applications

- Document recognition:
  - Many handwritten documents require digitization;
  - A well-trained LeNet can be used to digitize using maximal probable estimate
- Main use-case:
  - Post office automatic pincode classifier
  - Good accuracy, matches or exceeds human throughput

Architecture

Layer	# filters / neurons	Filter size	Stride	Size of feature map	Activation function
Input	-	-	-	32 X 32 X 1	
Conv 1	6	5 * 5	1	28 X 28 X 6	tanh
Avg. pooling 1		2 * 2	2	14 X 14 X 6	
Conv 2	16	5 * 5	1	10 X 10 X 16	tanh
Avg. pooling 2		2 * 2	2	5 X 5 X 16	
Conv 3	120	5 * 5	1	120	tanh
Fully Connected 1	-	-	-	84	tanh
Fully Connected 2	-	-	-	10	Softmax



Working

- The layers are a combination of Convolutional and average pooling ones, eventually leading to dense layers that do the classification from the extracted features.

- The kernels for convolution are learned using backpropagation;

- Non-linearities used:
  - Tanh

$$A\tanh(s \cdot x) = A\frac{e^{sx} - e^{-sx}}{e^{sx} + e^{-sx}}, \ A = 1.7159, s = 2/3$$

  - Softmax

$$\mathrm{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}}$$

A note on Sparse connectivity

	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

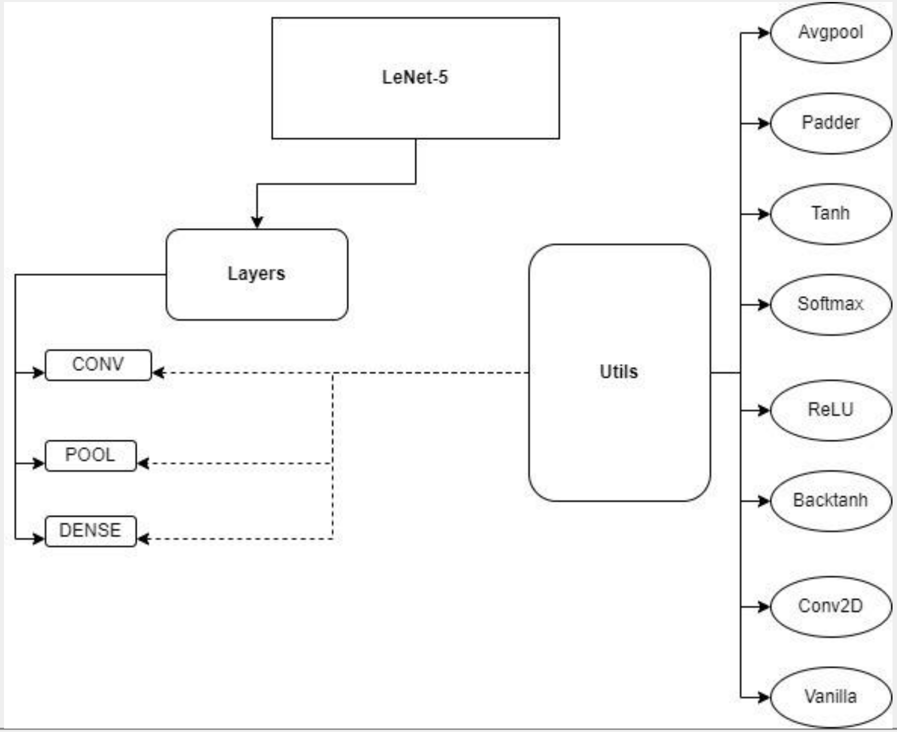
EACH COLUMN INDICATES WHICH FEATURE MAP IN S2 ARE COMBINED BY THE UNITS IN A PARTICULAR FEATURE MAP OF C3.

- In the network, between S2 and C3 the authors use a constructed map to take the 6 different channels of the output to 16 channels using 2D Convolution.

- Why we use Sparse connectivity:

Implementation

Framework


LeNet-5
Layers
CONV
POOL
DENSE
Utils
Avgpool
Padder
Tanh
Softmax
ReLU
Backtanh
Conv2D
Vanilla

Training on MNIST

- *Modified National Institute of Standards and Technology*
- ~70000 images of handwritten digits, original paper uses SVM to classify

Lenet using inbuilt (keras)

- We also made a model of LeNet on Keras using the inbuilt sequential model to compare with our implementation

- Benchmark

Idea of an alternative approach

- A CNN uses shared weights to identify image based features using the convolution operation.

- W can reshape the input image to a normal column vector and interpret it as the input layer to an MLP.

- Every conv layer ⟶ a normal weight,bias MLP layer
  But the weight matrix will be sparse circulant.
  Thus, backprop becomes very intuitive because we can implement a normal MLP and only change the values of the weights and not the positions.

$$\begin{pmatrix} k1 & k2 & 0 & k3 & k4 & 0 & 0 & 0 & 0 \\ 0 & k1 & k2 & 0 & k3 & k4 & 0 & 0 & 0 \\ 0 & 0 & 0 & k1 & k2 & 0 & k3 & k4 & 0 \\ 0 & 0 & 0 & 0 & k1 & k2 & 0 & k3 & k4 \end{pmatrix} \cdot \begin{pmatrix} x1 \\ x2 \\ x3 \\ x4 \\ x5 \\ x6 \\ x7 \\ x8 \\ x9 \end{pmatrix}$$

THANK YOU!