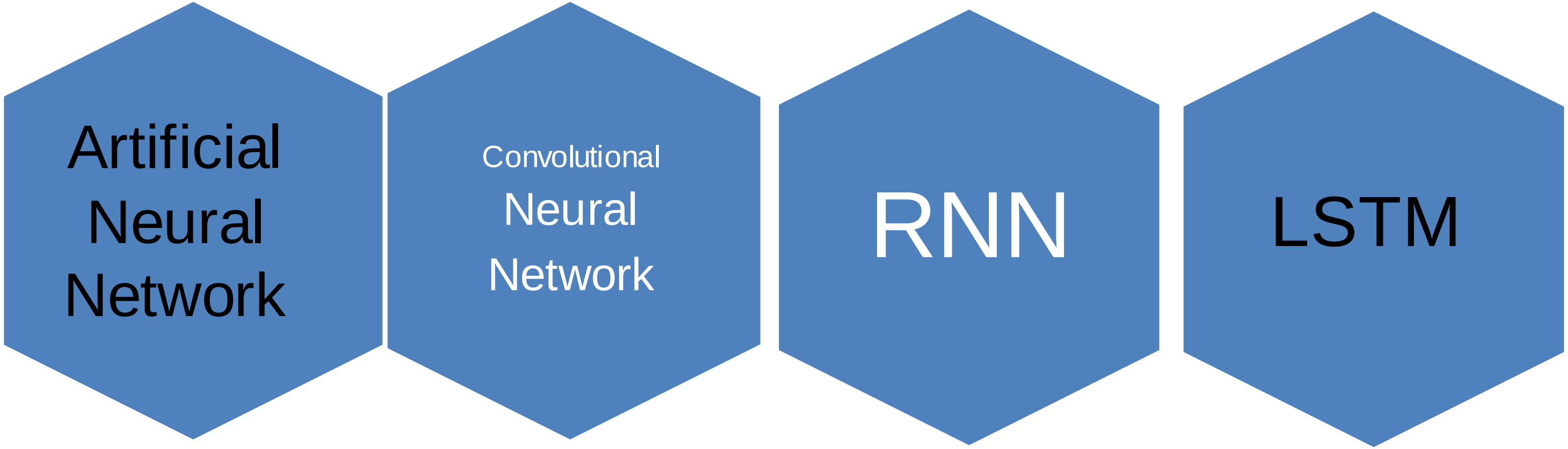
CONVOLUTIONAL NEURAL NETWORK

Effective at Dealing with Image Data

Disclaimer: The content is curated for educational purposes only.

Types of Neural Network



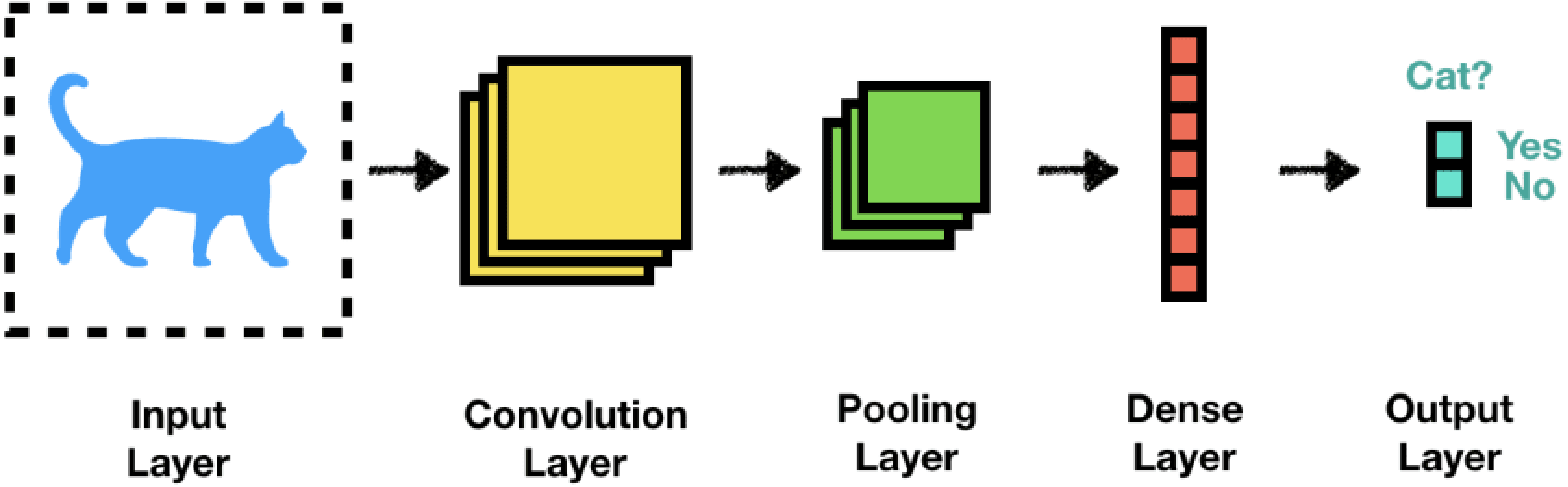
Artificial Neural Network

Convolutional Neural Network

RNN

LSTM

MLP – Multi Layer Perceptron



CNNs have stacked layered architecture of several Convolution and Pooling Layers

Understanding CNNs

1. Convolutional Layers.
2. Pooling Layers.
3. Fully-Connected Layers.
4. Padding

Blur Filters

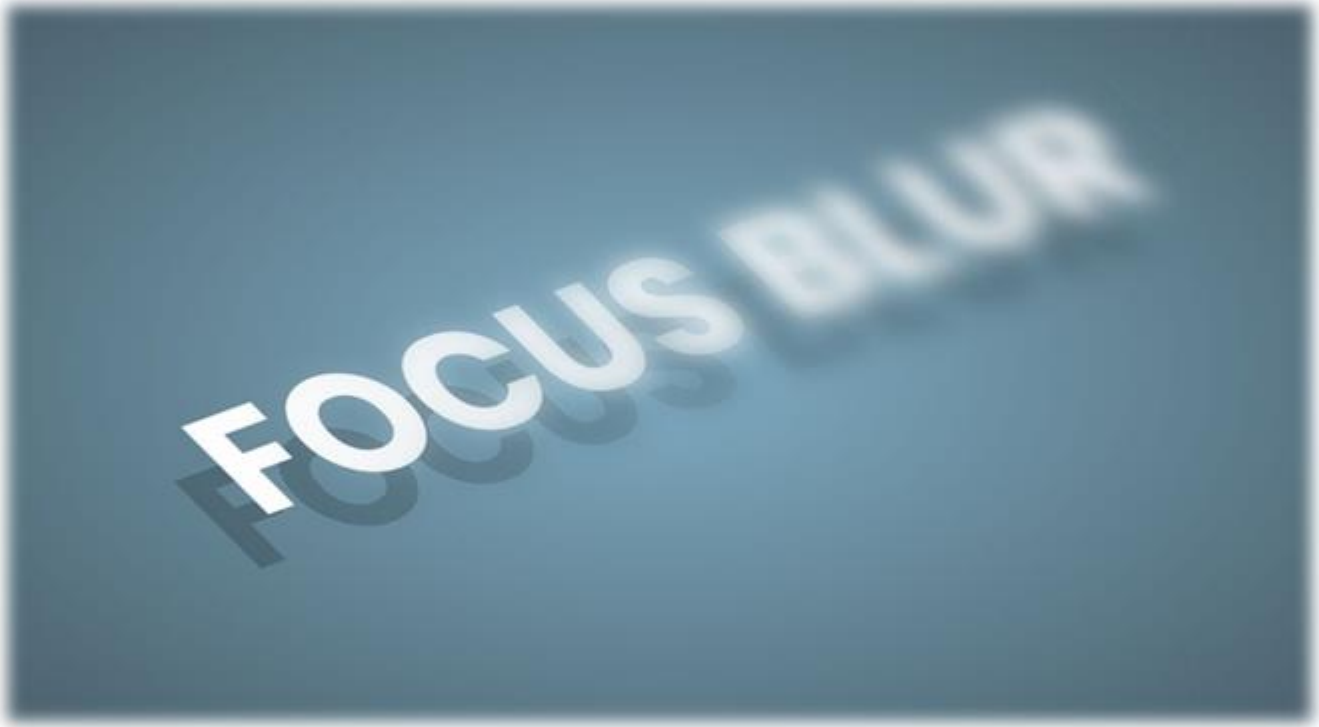
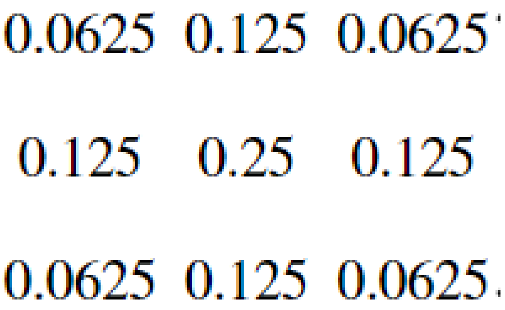


Image Kernel

Small matrix applied to an entire image

1/16	1/8	1/16
1/8	1/4	1/8
1/16	1/8	1/16



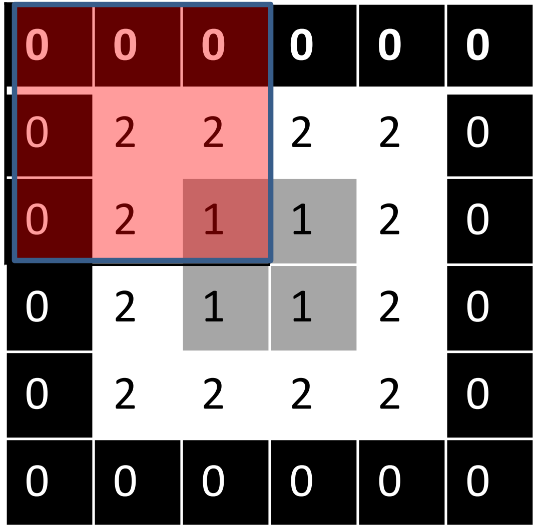
0.0625	0.125	0.0625
0.125	0.25	0.125
0.0625	0.125	0.0625

Blur Filter

- Filters are referred to as Convolution Kernels.
- The process of passing the Kernels over an image is called Convolution

Note : Padding can be used if you don't want to lose border information

Filter



0	0	0	0	0	0
0	2	2	2	2	0
0	2	1	1	2	0
0	2	1	1	2	0
0	2	2	2	2	0
0	0	0	0	0	0

0	0	0
0	-1	1
0	1	-1

3*3 FILTER

0X0	0X0	0X0
0X0	2X-1	2X1
0X0	2X1	1X-1

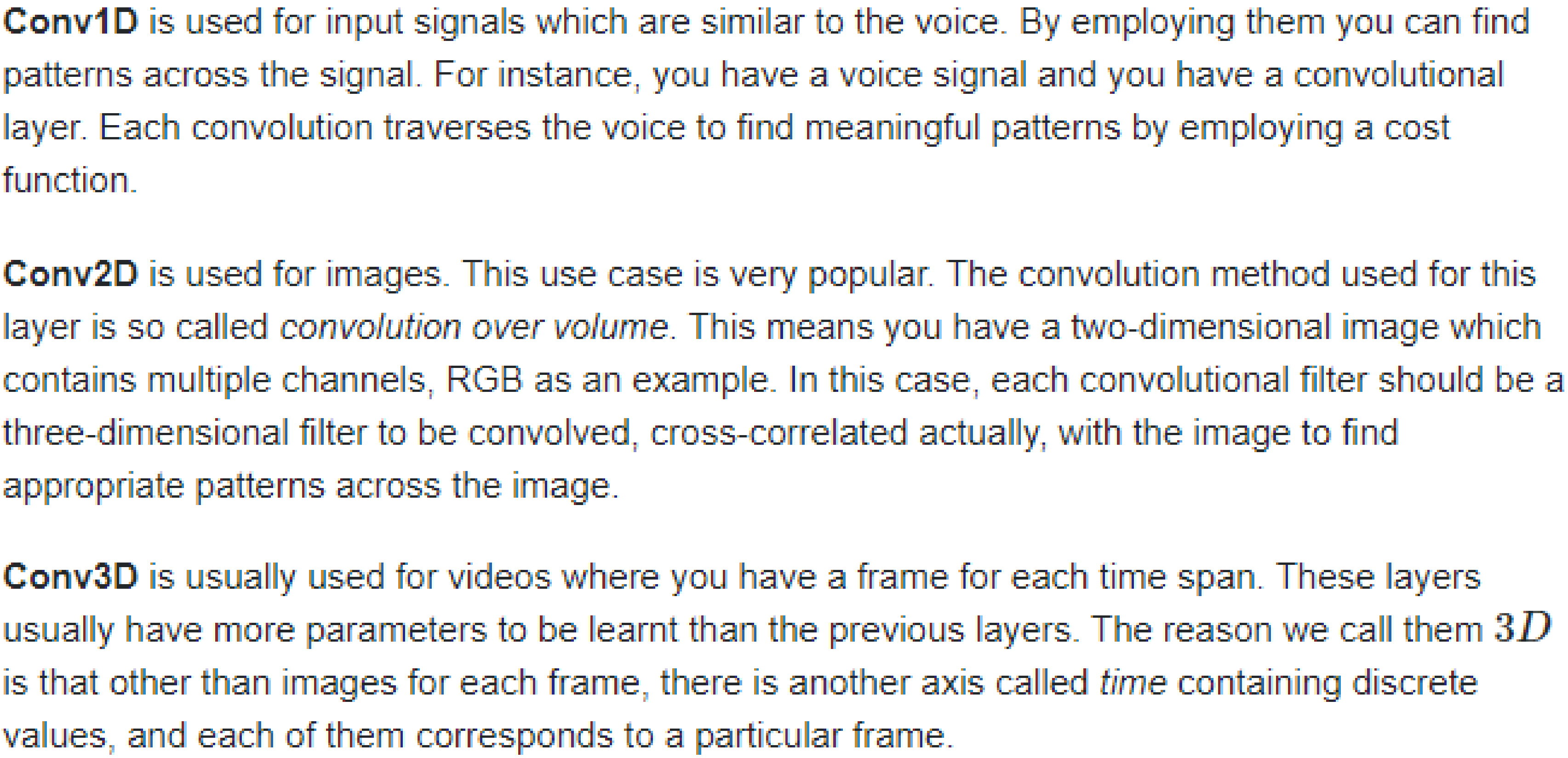
Multiply by Filter Weights

0	0	0
0	-2	2
0	2	-1

SUM THE RESULT = 1

Stride Distance = (1,1)

Note: The resolution will decrease because we are taking 9 input values and one output value



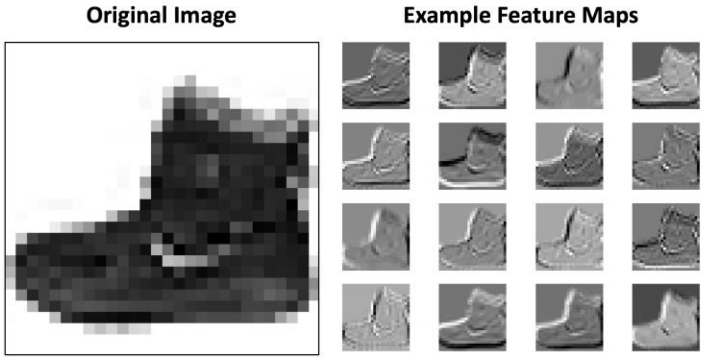
Conv1D is used for input signals which are similar to the voice. By employing them you can find patterns across the signal. For instance, you have a voice signal and you have a convolutional layer. Each convolution traverses the voice to find meaningful patterns by employing a cost function.

Conv2D is used for images. This use case is very popular. The convolution method used for this layer is so called *convolution over volume*. This means you have a two-dimensional image which contains multiple channels, RGB as an example. In this case, each convolutional filter should be a three-dimensional filter to be convolved, cross-correlated actually, with the image to find appropriate patterns across the image.

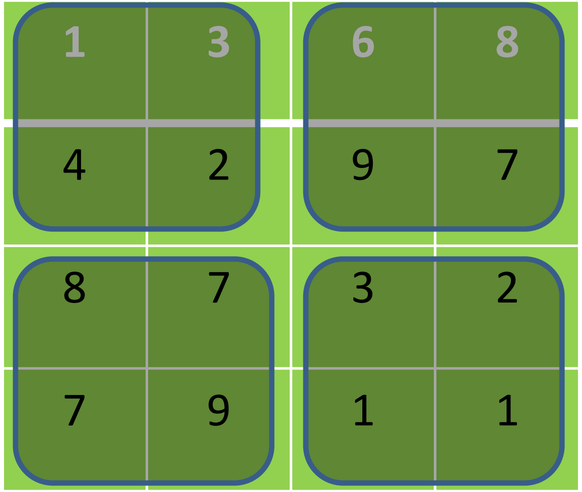
Conv3D is usually used for videos where you have a frame for each time span. These layers usually have more parameters to be learnt than the previous layers. The reason we call them $3D$ is that other than images for each frame, there is another axis called *time* containing discrete values, and each of them corresponds to a particular frame.

Convolutional Layers

- Perform FEATURE EXTRACTION on **Source Images** by applying **filters**(eg. 16)
- Understand various pattern with regard to Image Structures (Textures, Edges, Corners and Patterns)
- They extract **Feature Maps**



Max Pooling



1	3	6	8
4	2	9	7
8	7	3	2
7	9	1	1

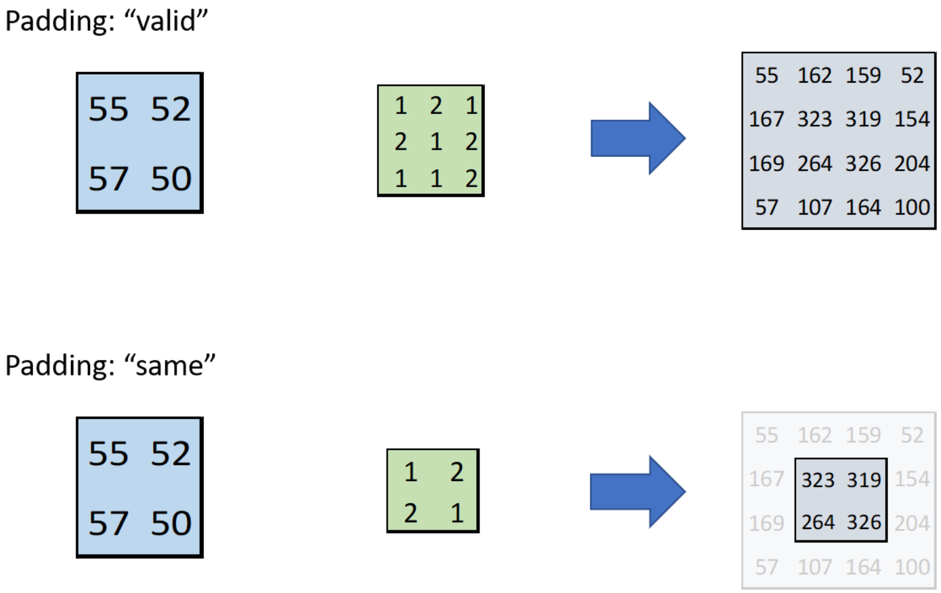
Downsamples Feature maps from Conv Layers

4	9
9	3

Window 2x2 Stride : 2

Pooling Layers help in reducing Dimensionality after convolutions(compression)

Padding



Padding: “valid”

55	52
57	50

1	2	1
2	1	2
1	1	2

55	162	159	52
167	323	319	154
169	264	326	204
57	107	164	100

Padding: “same”

55	52
57	50

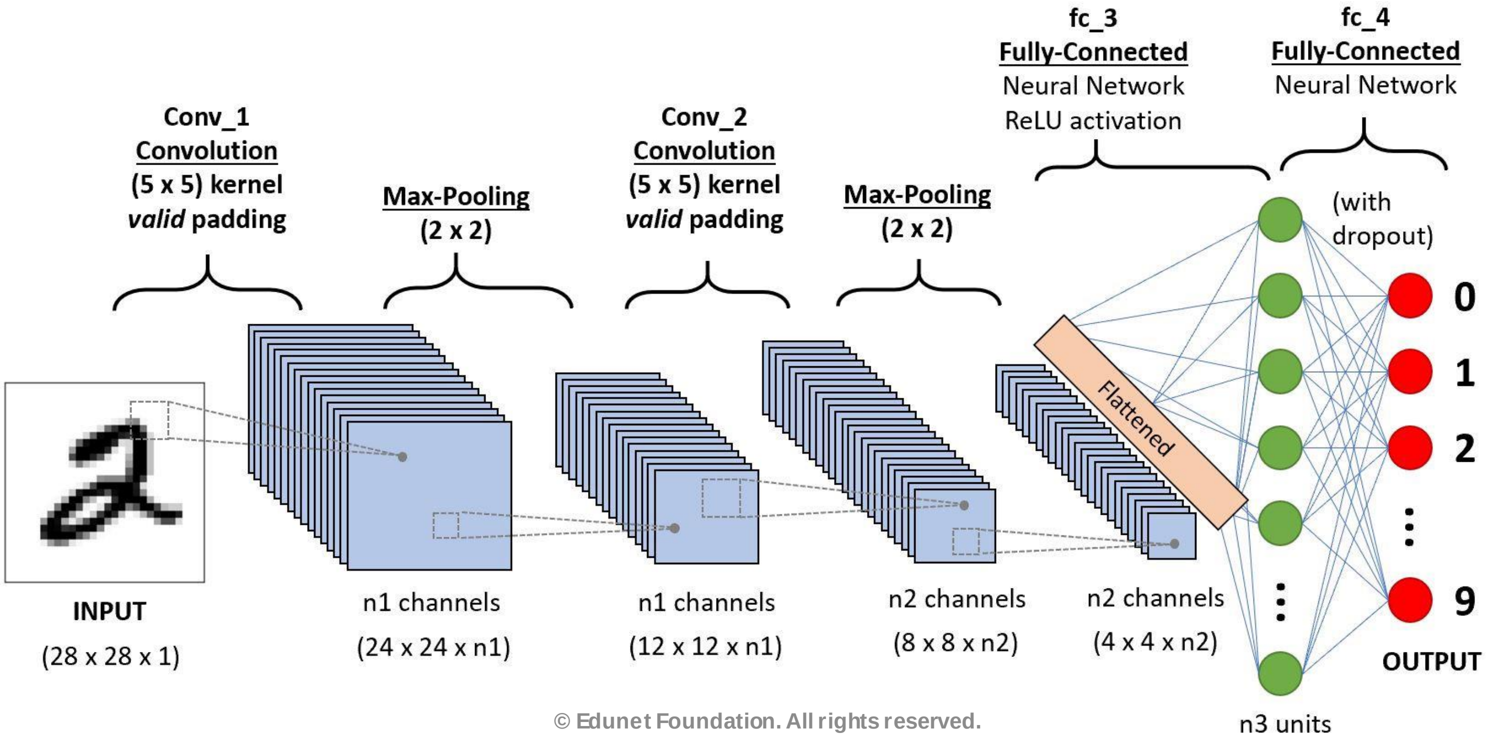
1	2
2	1

55	162	159	52
167	323	319	154
169	264	326	204
57	107	164	100

1. valid
2. same

https://miro.medium.com/max/2264/1*XUJwcj95wk-LsILwzJHaAg.png

CNN


fc_3
Fully-Connected
Neural Network
ReLU activation
fc_4
Fully-Connected
Neural Network
Conv_1
Convolution
(5 x 5) kernel
valid padding
Max-Pooling
(2 x 2)
Conv_2
Convolution
(5 x 5) kernel
valid padding
Max-Pooling
(2 x 2)
(with
dropout)
0
1
2
9
Flattened
INPUT
(28 x 28 x 1)
n1 channels
(24 x 24 x n1)
n1 channels
(12 x 12 x n1)
n2 channels
(8 x 8 x n2)
n2 channels
(4 x 4 x n2)
OUTPUT
n3 units

References

1. https://www.cs.ryerson.ca/~aharley/vis/
2. https://setosa.io/ev/image-kernels/
3. https://towardsdatascience.com/understand-transposed-convolutions-and-build-your-own-transposed-convolution-layer-from-scratch-4f5d97b2967

THANK YOU