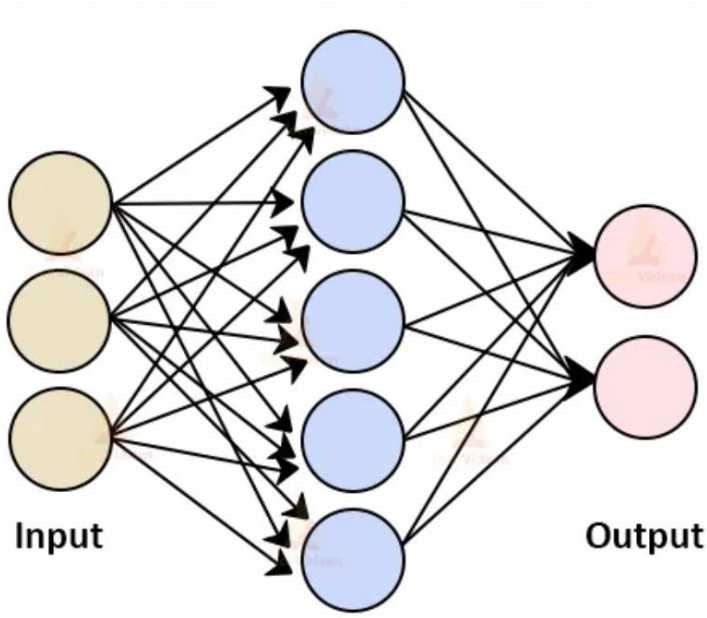
Question (Task 3_1)

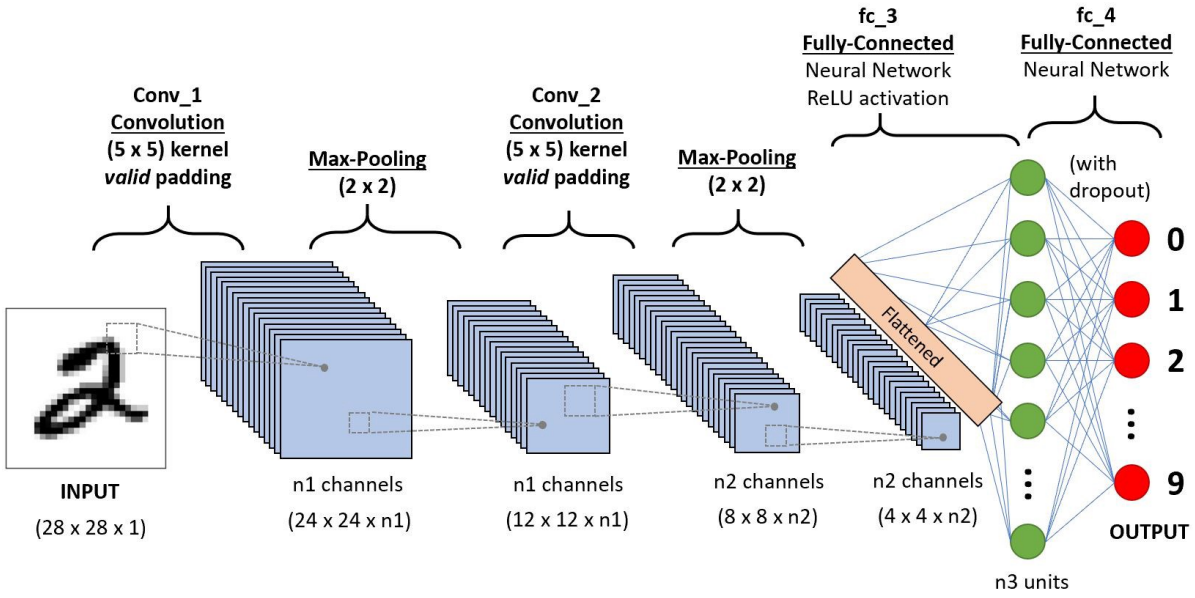
As a part of the wildlife and environmental monitoring project, you need to identify the species of animals and birds. Consider yourself to be a part of a wildlife sanctuary/refuge team with access to video devices. There are designated camera spots within the territory.
How would you employ Machine Learning and its aiding technology for this purpose? Explain how you can preprocess the sources, and classify the animals or birds in the video clips. You have to mention things like the algorithm or method you will be using and the reasoning for the same.
Mathematical explanations would be beneficial.

Answer (Using CNN)

ANN -



CNN-



ANN disadvantages-

1] Computation Cost
2] Overfitting
3] Loss of important features (While converting from 2D to 1D directly)

CNN advantage over ANN-

CNN helps us to reduce the dimensionality without reducing the actual quality of the image used (it helps us to preserve the features of the image so that we can get a precise output) it helps us to reduce the amount of time and also the computational complexity required by ANN. In CNN we filter out important info and then apply ANN

CONVOLUTION-

A convolution in neural network training is a reduction of the number of weights that the model needs to learn by summarizing the image into fewer pixels.
A convolution combines information about local pixels such that pixels close to each other in an image are ”summarized” by a smaller set of pixels. This ”summarization” is done by ”sliding” a kernel across an image to output a final product for each position of the kernel.

CNN TERMINOLOGIES-

A kernel in a convolution is an n x n matrix of numbers. Since the image is a 4x4 image and the kernel is a 3x3 matrix, the kernel can have 4 unique positions on the image. As a result, the output is a 4-pixel, or 2x2 image. It's a filter.

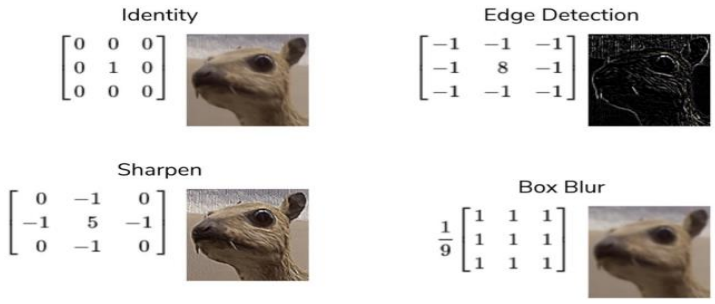
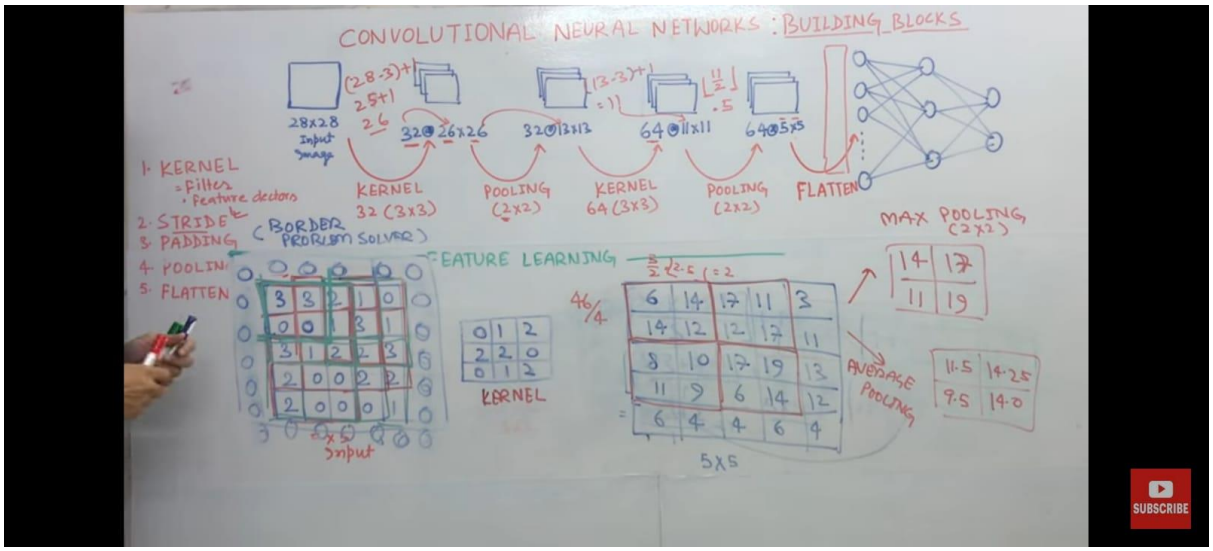


Figure 18.96: Special kernels which have specific properties such as maintaining the same image, detecting edges, sharpening the image, and applying a box blur

padding: a layer of pixels of some static (unchanging) value wrapped around the image so that the kernel can align with the image more conveniently. (padding is actually added so that more concentration is given to border liner element too).(After padding we use padding. This is to reduce the dimensionality after adding padding)



Pooling of (2x2) halves the matrix dimensions if even (32x32 after pooling with 2x2 will become 16x16) (as shown in the above pic.

dimension: how large the kernel should be
stride: how far the kernel should slide across the image each iteration.

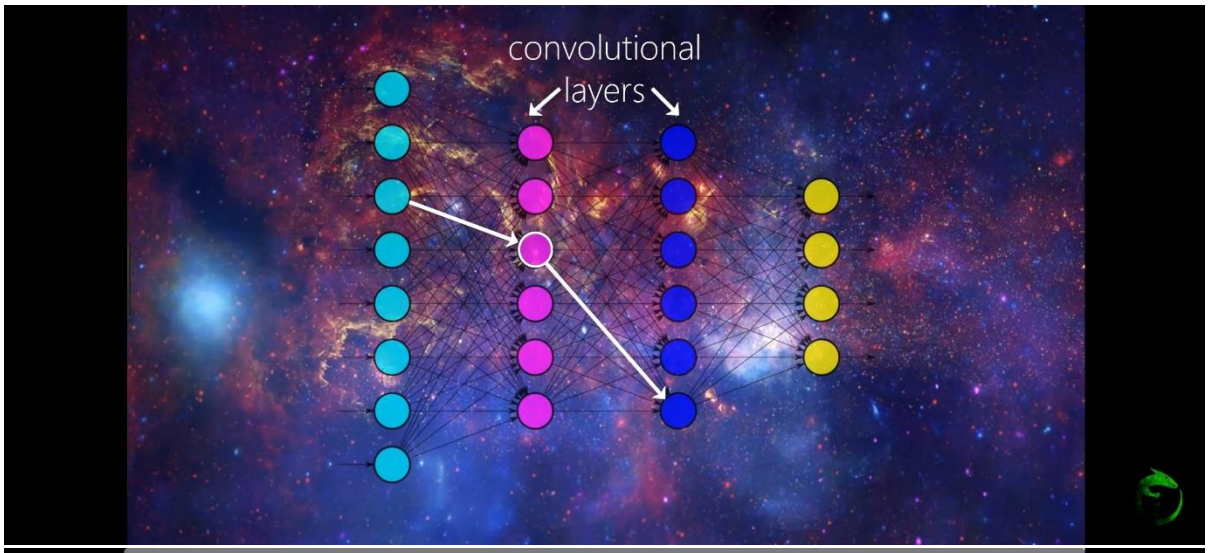
Flattening-Once the pooled featured map is obtained, the next step is to flatten it. It involves Transforming the entire pooled feature map matrix into a single column which is then fed to the neural network for processing.

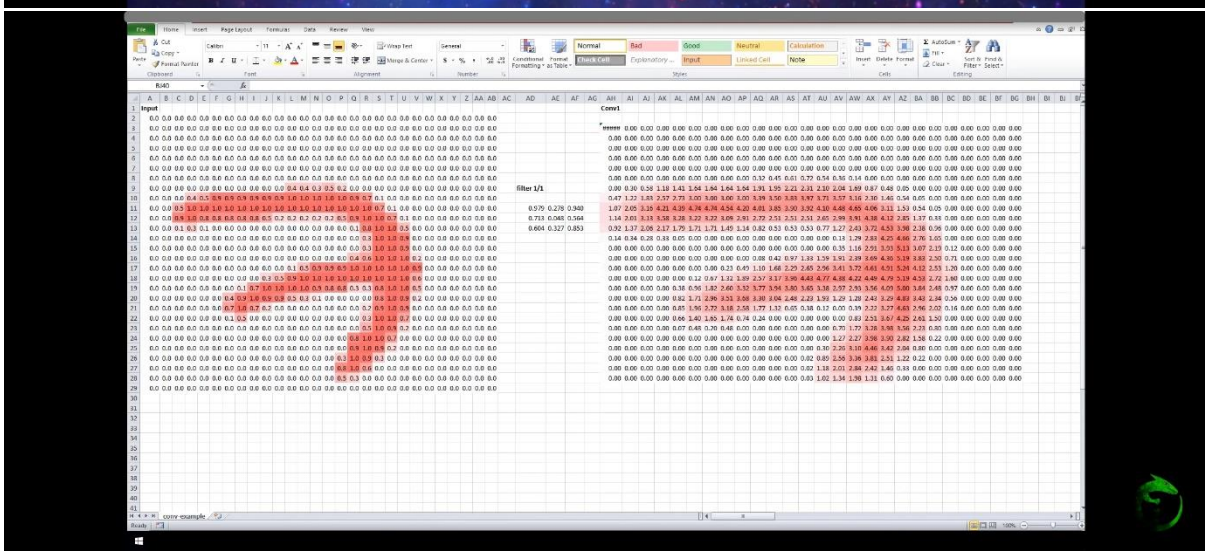
Size of OutPut after using filter- (WITHOUT PADDING AND POOLING)

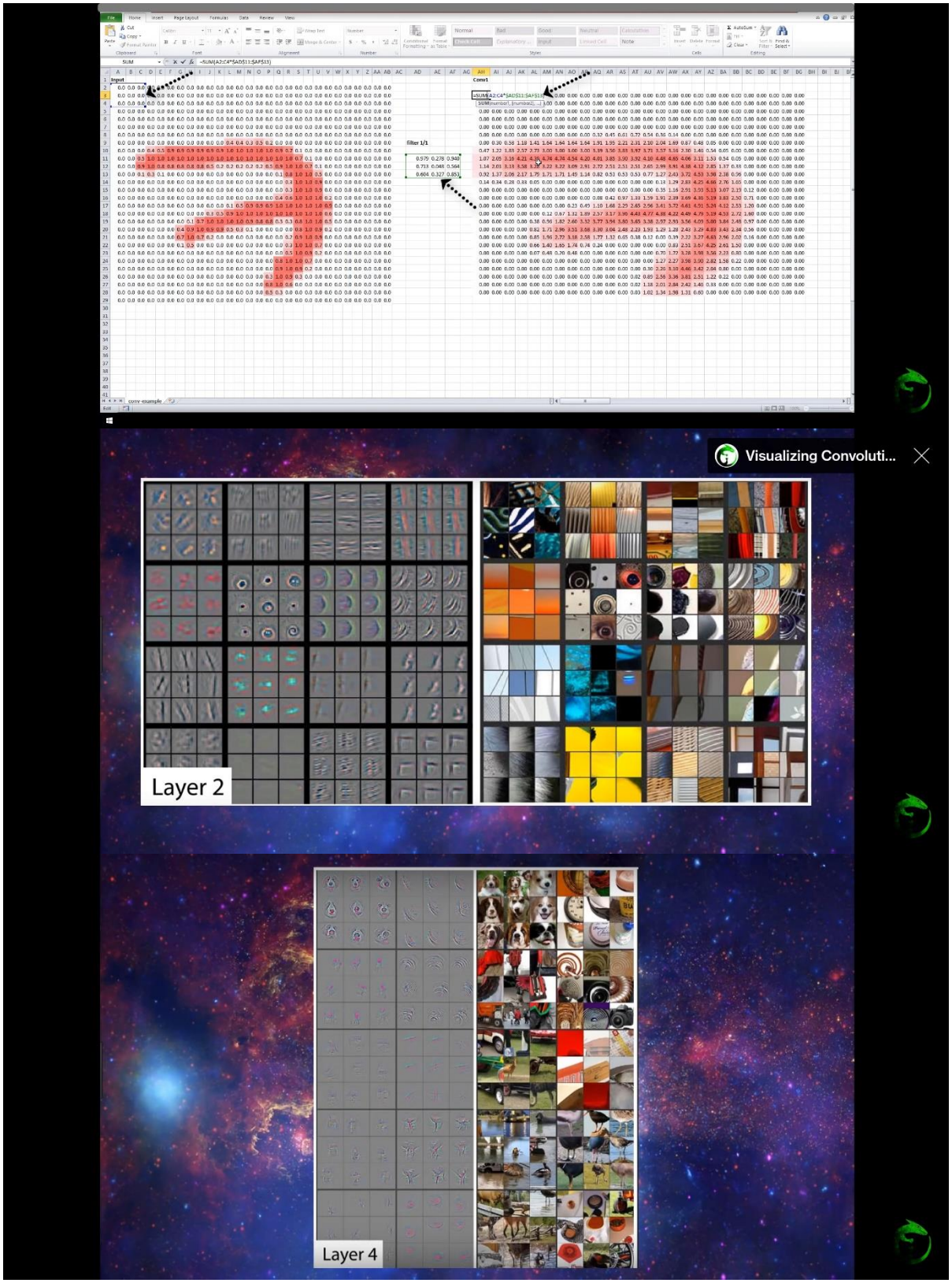
O=[imput-kernal]/s + 1

S is the stride (how much up or down it will go)

Flattening-






Visualizing Convoluti...
Layer 2
Layer 4