

---

## Computer Vision Assignment 1

### 1. What is a feature, exactly?

Answer:- In computer vision, A feature is a measurable piece of data in your image which is unique to this specific object it may be a Distinct color in an image or a specific shape such as a line, edge or an image segment. A good feature is used to distinguish objects from one another. Features are parts or patterns of an object in an image that help to identify it. For example — a square has 4 corners and 4 edges, they can be called features of the square, and they help us humans identify it's a square. Features include properties like corners, edges, regions of interest points, ridges, etc.

### 2. Create the convolutional kernel matrix for a top edge detector.

Answer:- Edge detection is a technique of image processing used to identify points in a digital image with discontinuities, simply to say, sharp changes in the image brightness. These points where the image brightness varies sharply are called the edges (or boundaries) of the image. It is one of the basic steps in image processing, pattern recognition in images and computer vision. When we process very high-resolution digital images, convolution techniques come to our rescue. Let us understand the convolution operation (represented in the below image using \*) using an example—

### 3. Explain how a 3x3 kernel performs a mathematical function on a single pixel in an image.

Answer:- We can think of an image as a 2D matrix containing pixel color values in the range of 0 to 255. Mathematically we can manipulate this matrix by applying various matrix operations. We will be using [OpenCV](#) (a flexible library for image processing), NumPy for matrix and array operations, and Matplotlib for plotting the images. We use `imread()` object to read the image. By default `cv2.imread()` reads the image in the format of Blue, Green, and Red. We need to convert it into Red, Blue, and Green format, that makes sense. We take matrix values of a GRAY scale image where each pixel contains values in between 0 and 255. The problem with the color image is that each pixel value is a combination of 3 values probably the form of  $[R, G, B]$  or  $[B, G, R]$  which can make the computation complicated. So, to keep things simple we take a GRAY scale image

### 4. What does adding a convolutional kernel to a 3x3 matrix of zeros mean?

Answer:- One of the reasons to prefer small kernel sizes over fully connected network is that it reduces computational costs and weight sharing that ultimately leads to lesser weights for back-propagation. In image processing, a kernel, convolution matrix, or mask is a small matrix. It is used for blurring, sharpening, embossing, edge detection, and more. This is accomplished by doing a convolution between a kernel and an image. What does convolutional input mean?

The first layer of a Convolutional Neural Network is always a Convolutional Layer. Convolutional layers apply a convolution operation to the input, passing the result to the next layer. A convolution converts all the pixels in its receptive field into a single value.

### 5. What is padding, exactly?

Answer:- Padding is a term relevant to convolutional neural networks as it refers to the amount of pixels added to

an image when it is being processed by the kernel of a CNN. For example, if the padding in a CNN is set to zero, then every pixel value that is added will be of value zero.

Padding basically extends the area of an image in which a convolutional neural network processes. The kernel/filter which moves across the image scans each pixel and converts the image into a smaller image. The padding type is called SAME because the output size is the same as the input size (when stride=1). Using 'SAME' ensures that the filter is applied to all the elements of the input. Normally, padding is set to "SAME" while training the model. Output size is mathematically convenient for further computation.

#### 6. What is the meaning of the term "stride"?

Answer:- Computer vision (and as the name suggests), a sliding window is a rectangular region of fixed width and height that "slides" across an image, such as in the following figure. Example of the sliding a window approach, where we slide a window from left-to-right and top-to-bottom. Low-level vision: process image for feature extraction (edge, corner, or optical flow). Middle-level vision: object recognition, motion analysis, and 3D reconstruction using features obtained from the low-level vision. Middle and low level vision tasks should be performed.

#### 7. What are the input and weight parameters for PyTorch's 2D convolution?

Answer:-

The convolutional layer also has a bias input value that also requires a weight that we will set to zero. The weights must be specified in a three-dimensional structure, in terms of rows, columns, and channels. The filter has a single row, three columns, and one channel.

#### 8. What is the definition of a channel?

Answer:- The term comes from the definition of "channel" that means a specific portion of a frequency spectrum. In this case, the red, green and blue components of a color are often referred to as "channels" (since red, green and blue light are portions of the visible light spectrum). What is a channel in an image?

In the context of a digital image, a channel is simply an array of values, one per pixel, that together specify one aspect or dimension of the image. For example, you can describe an image by specifying the red, green, and blue values for each of the pixels in the image. This is known as RGB (for red, green, blue).

#### 9. Can you explain how matrix multiplication and convolution are related?

Answer:- Convolution, for discrete-time sequences, is equivalent to polynomial multiplication which is not the same as the term-by-term multiplication. Convolution also requires a lot more calculation: typically  $N^2$  multiplications for sequences of length  $N$  instead of the  $N$  multiplications of the term-by-term multiplication.