PadhAI: CNN Architectures

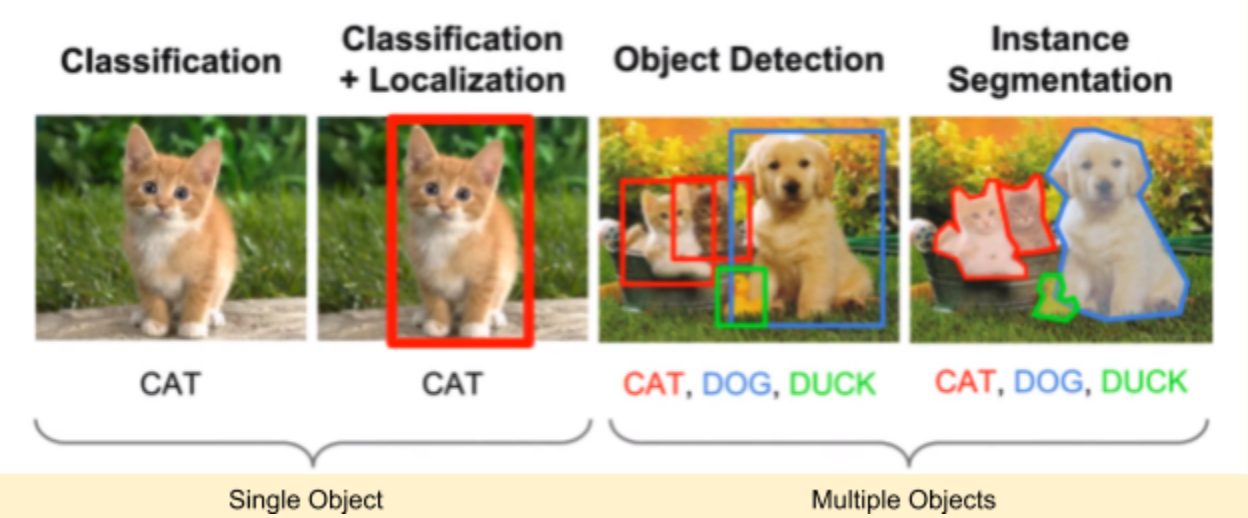
One Fourth Labs

CNN Architectures I

Setting the context

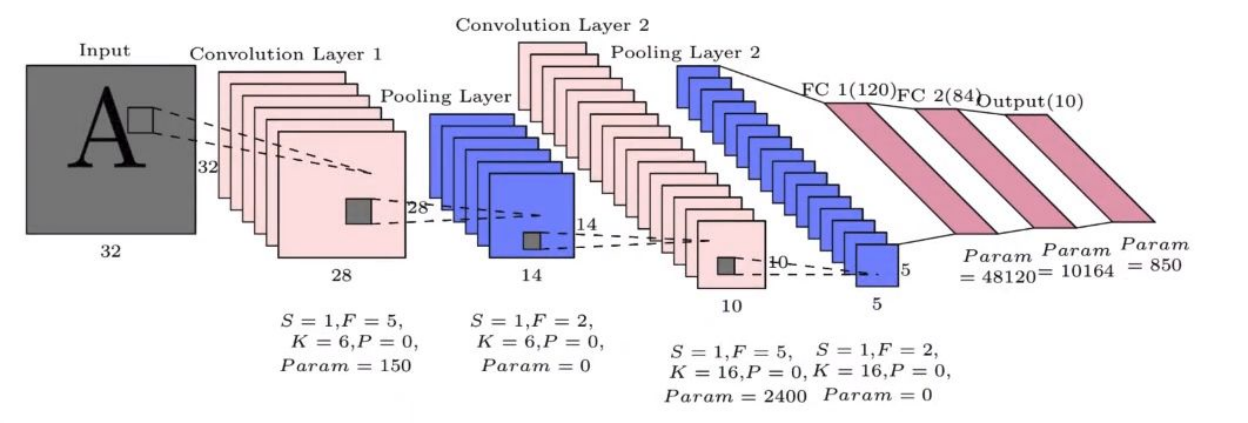
What are CNNs used for?

1. First, let's see **what kind of tasks** are CNNs used for
   a. Consider the following image for an overview of the tasks that CNNs are used for



   b. Using popular datasets like Imagenet, which contains 1000 classes of objects, ranging from vehicles and sceneries, up to the different dog & cat breeds.
   c. **Classification**: In the first image, the task at hand is to correctly predict the class to which the object belongs to. This example uses an "Iconic Photo", i.e. where the class-object of the photo occupies most of the photograph area.
   d. **Classification + Localization**: In the second image, we are predicting the class of the object and also precisely where it is located in the image. Given a starting point outside the object, we detect the width and height of the bounding box enclosing the object. This is both a classification problem and a regression problem.
   e. **Object Detection**: In the third image, we have multiple objects. Therefore, we must correctly detect each object and classify them respectively. It involves multiple Classification + Localization operations on the same image.
   f. **Instance Segmentation**: In the fourth image, we are moving a step further from object detection. Here, we are identifying the precise bounding area around each of the objects present in the image. This is commonly performed in autonomous driving etc, where we have to detect the presence of multiple objects at any given point in time.
   g. These are the four applications that commonly use CNNs. In fact, even our capstone-project of Character-detection and recognition also requires the above mentioned processes.
2. A typical recipe for solving image-related tasks is as follows
   a. First passing the input images through as series of convolutional layer
   b. We obtain a 3D tensor which we flatten to a single dimensional vector
   c. We pass the vector through a number of fully connected layers, which culminate in output prediction
   d. In these cases, we either perform Classification or Regression
3. Now, if we are going to use CNNs for these tasks, we have to make a few choices with regards to the design of the CNN layout.

4. What are some of the decisions that need to be taken?
   a. Let's look at a sample CNN architecture



   b. Some of the factors under our control are as follows:
   c. Number of layers
   d. Number of filters in each layer
   e. Filter Size
   f. Max pooling
5. With the amount of choice we have, designing a CNN architecture could become a very messy process.
6. So a **standard practice is to use tried and tested architectures**.
7. In this chapter, we will be looking at the most popular CNN architectures.