Introducing Deep Learning for Computer Vision

Introduction

You have probably heard a lot about deep learning. The term appears all over the place and seems to apply to everything. In reality, deep learning is a subset of machine learning, which in turn is a subset of artificial intelligence.

If you are just getting into this field then it can be really difficult to cut through the media hype about Deep Learning. In this article, we will understand what deep learning truly is, and introduce some key concepts of deep learning, starting all the way from the basics of artificial intelligence and machine learning. Finally, we will dive into what deep learning really is. By the end of this article, you will have a solid foundation which will help you navigate through your new learning chapter.

Why should you read this article

Having found it difficult to get started with AI and Deep Learning, I aim to make this article path for all new learners to tackle this field, and make complex topics easier to understand. This will also help you explore deep learning and see if it interests you. We also plan to follow up with additional articles that dive deeper into certain aspects of deep learning.

What is Deep Learning

Before we learn about Deep Learning let's take a look into AI and ML. Artificial Intelligence [AI] consists of a wide range of technologies that enable computers to perform various tasks that require human intelligence (Human Intelligence comprises skills that allow us humans to learn, solve problems, understand new complex ideas, and adapt to our ever-changing environment). This could be recognizing images, understanding speech, etc.

Machine Learning [ML], a subset of AI focuses on the development of algorithms that allow computers to learn and make decisions. Deep learning is a type of learning that is inspired by the structure of the human brain. Deep learning implements this concept with the help of artificial neural networks. Let us take an example of a sorting problem where we need to sort between images of cricket balls and footballs, if done using machine learning we will have to tell the machine the features based on which the two can be differentiated. These features could be the size and color of the ball. With Deep Learning, the features are picked out by the neural network without any human intervention!

Image classification

Since we are reading about deep learning for computer vision let's take a look into image classification. Image Classification is a core computer vision task where we analyse data, which is in the form of images and assign an image to one of a fixed set of categories.

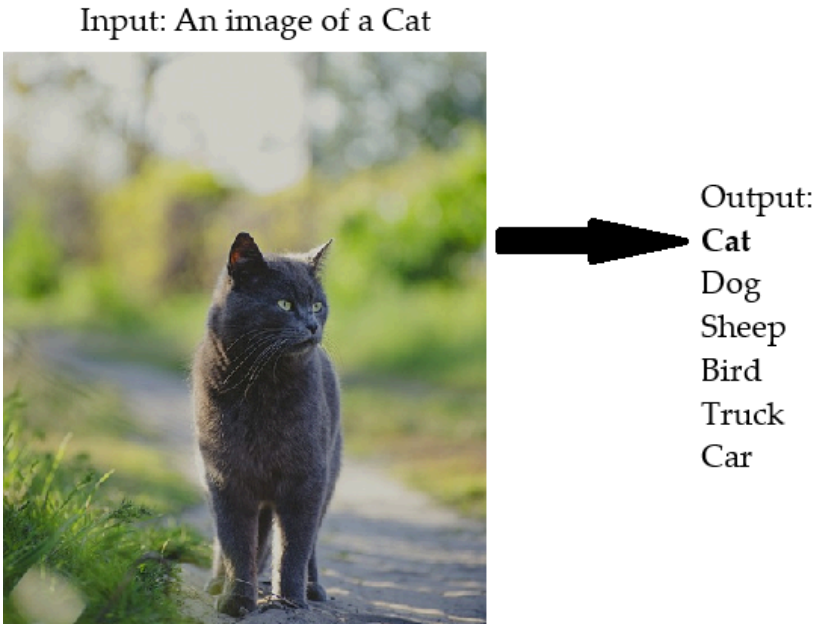


Image classification plays an important role in various applications in a lot of sectors including healthcare (identifying benign and malignant tumours) , security, autonomous vehicles (recognizing objects, pedestrians and road signs), social media, etc.

Before we go in depth into image classification lets understand image labels. Labels are words that describe the content of an image, they are used to categorise images. For example, if you have a collection of animal photos, the labels could be “dog” or “cat” or “bird” etc, or if you have a group of traffic signs the labels could be “stop”, “yield”, “speed limit”, “no parking”, etc.

Generally, image classification involves certain steps that include data collection, preprocessing, model training and prediction. We will discuss all of these things in detail in future articles. This field of image classification is not simple and has its own limitations and challenges. Previously we saw how an ideal image detection model would work, where we provided an image of a cat and the model returned the correct class of the image being cat. We can face several issues while

trying to implement an image classification model. I have tried to highlight some of the issues that we might face with the help of cats below

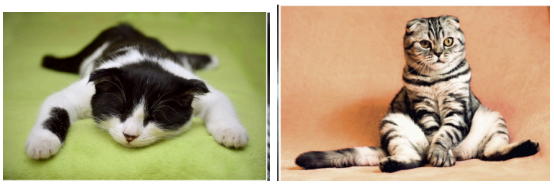
Challenge 1: Change in viewpoint of the camera

Changes in the camera angle is something that we have to take into consideration while training our image detection model

Challenge 2: Changes in lighting

Different lighting conditions can result in models giving out the wrong prediction for classes, clearly in the above images, we can see why some images might be harder for image classification model to classify

Challenge 3: Deformation



The above images of cats sleeping/sitting are semantically different for a computer which can only see pixel values of the images

Challenge 4: Obstructions

If you notice the first image, we notice the paws sticking out of a blanket. We humans, with the help of our experience, know cats love burying themselves in blankets, and hence are able to deduce that there is a cat sleeping inside the blanket. Now imagine a computer trying to do that with arbitrary pixel values and no prior knowledge about the behaviour of cats

Challenge 5: camouflage and Background disturbance

The white cat in the image almost merges into its snowy environment, differentiating the cat from the snow can be a challenge for our image classifier

---

To overcome such challenges there have been several image classification models that have been developed over the years, now let us look at some of these image classification models. In our article series of Deep Learning for Computer

Vision we will stick more towards these image classification models and methods. Now let's look at what image classifiers are and their different categories

An Image Classifier

Image classifiers are algorithms or models that can categorise images into predefined images/ labels. We can categorise these into two types based on their underlying methodologies, these are our traditional machine learning classifiers and deep learning classifiers.

The traditional classification method relies on us to provide the model with features of each class with which it can classify the data, while the deep learning classifiers rely on neural networks to learn hierarchical features from raw pixel data.

Traditional Classifiers

Now let us take a look into what an ideal image classifier should look like. Let's say we have a hypothetical method that accepts an image as a parameter and returns the class/label of the image. Lets say, something like this.

```
def classify_image(image):
    # we classify the image
    return image_label
```

From our machine learning/ traditional learning approach we need to collect the data (images and labels) we need to use machine learning to train the images and evaluate our newly trained model on new images

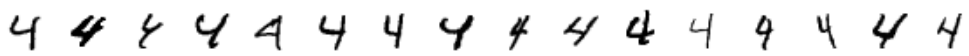
```
def train_image(image, label):
    # we train the model here
    return model

def classify_image(image):
    # we classify the image
    return image_label
```

Assume you have a large number of images, firstly let's segregate these images into training set and test set images. Training set images are images that have labels attached to them, these images are used to train the model. After training our model we use the images we have in the test set to see how well our image works by checking the predicted label to the correct label. An example of your traditional image classifier is the nearest neighbour classifier. Nearest neighbour classifier is the most simple image classifier, it decides what class something belongs to just by looking at the closest examples in the training set for that image. It checks which set

of examples resembles the test image the most and provides the most common category among them.

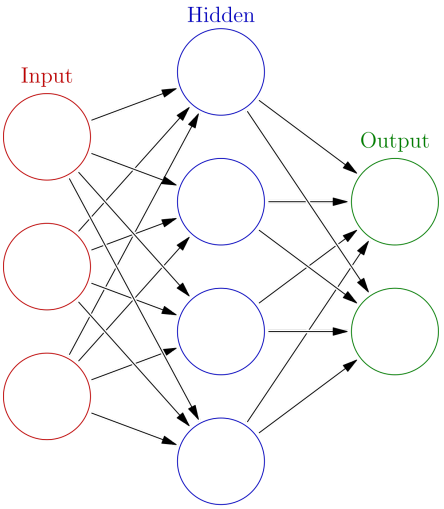
Neural Networks

The image below shows the handwritten digit "4", each digit is different from the conventional 4 as seen on our keyboard, but our brain is easily able to recognize these numbers. Now what if a computer had to perform the same task?



And how would a computer differentiate between  and  which are just different ways of writing the number 7?

This is where the concept of neural networks comes in. The image below shows an example of a neural network,



A neural network consists of layers, in our example, we have red, blue, and green layers. Each circle in the neural network is called a neuron, this is the fundamental unit of a neural network. This is where the information processing takes place.

The Red layer is called the input layer that contains information about the image, in our case, it is the example of the digit 7. On the other end, we have the Green layer which is called the output layer.

There can be multiple layers in between the input and the output layer and they are generally referred to as the hidden layers of the neural network

Information is transferred from one layer to another with the help of channels [represented with the help of arrows in the figure]. These channels have values attached to them and hence are called weighted channels. These are just the definitions of different parts of a neural network. Each neuron has a value assigned to it and is called the bias. With the help of these weights and bias we determine which neuron is activated with the help of an activation function. These now-activated neurons are used to calculate the activated neurons for the next layer. The weights and bias are continuously adjusted till we get the correct output (in our case, till the neural network is able to correctly identify the number)

Don't worry if this feels too overwhelming, we are just trying to provide you with definitions for the different parts of a neural network so you can understand things

easier. In our next article we will understand neurons, layers, weights, biases, activation function and many other things in details

Actionable Takeaways

What must you do after reading this article? If you found the field of Deep Learning interesting and want to continue learning, check out our upcoming article on Neural Networks in which we look at a particular dataset called the MNIST dataset which consists of grey scale images which are in a 28x28 pixel format. There we will continue to work on the number example which we provided in the neural network section of our article and build our first neural network to classify images from 0 to 9.

Conclusion

While Deep Learning is a powerful and transformative field, it can initially seem overwhelming due to its complexity and abundance of information available. In our article we tried to give you a strong foundation by breaking down what Deep Learning is all about. We tried to make this article accessible, ensuring that you will gain a foundational understanding of this concept even without prior knowledge about Python or TensorFlow and complete projects like digit recognition using the MNIST database. A lot of other guides require several prerequisites. We tried to make this article and its project as beginner friendly as possible so that you with your newly gained interest and confidence in Deep Learning can go and benefit from the vast amount of information on the Web. Yes, if you want to make projects of your own, having a good grasp of Python, TensorFlow and Pytorch will be extremely crucial. And so, you must keep exploring and learning, and while doing so, remember deep learning is a journey that requires patience and practice.

We hope this article has made the topic feel much more approachable and encouraged you to take your next step into your Deep Learning journey. Stay tuned for our next Article where we will explore more advanced neural networks architecture and their practical applications along with another project!

References and Citations

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
Goodfellow, I., Bengio, Y., & Courville, A. (2016).Deep Learning. MIT Press. Chapter 1
Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. Chapter 6