Module 1 - Intro to Deep Learning

- 1.1 Introduction to Deep Learning
- 1.2 Deep Learning Pipeline
- 1.3 Graded Review Questions (2 Questions)

Learning Objectives

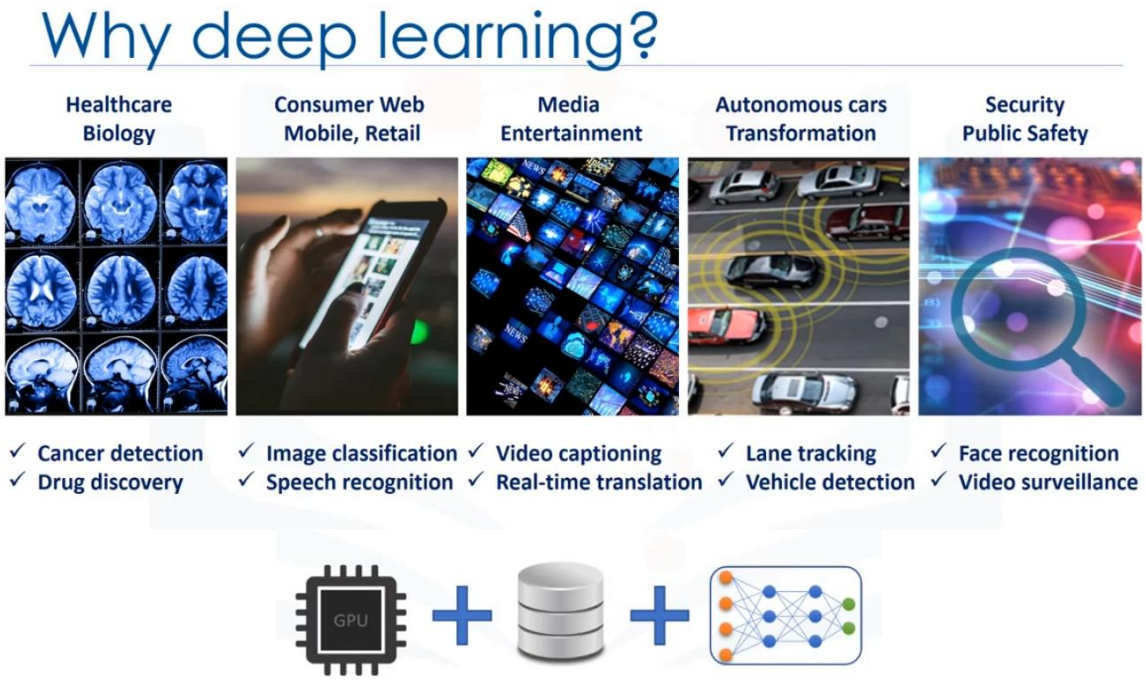
In this lesson you will learn about:

- Introduction to Deep Learning
- Why you should use hardware accelerators for Deep Learning

总结:

1.1 Introduction to Deep Learning

This video is an introduction to deep learning, its applications, and also the difference between the shallow neural networks and deep neural networks. Please skip this video if you have already taken an introductory course in deep learning, or you know about deep learning and its applications.

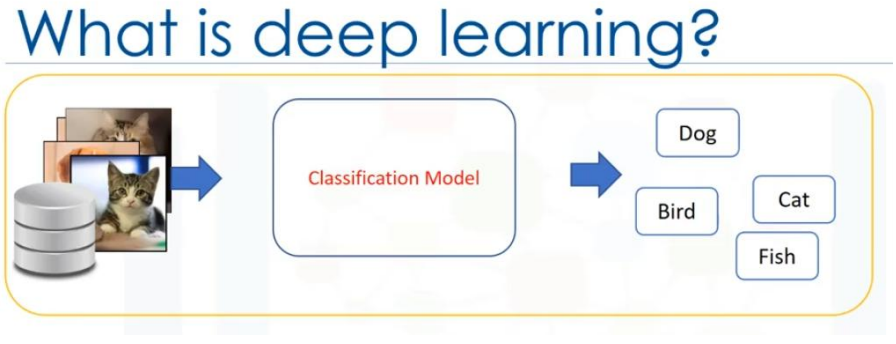


What is deep learning? And why is it such a hot topic today? In this video, we'll begin to answer those questions. We'll also look at the difference between traditional shallow neural networks and deep learning.

I'm sure most of you have already heard about the usage of deep learning. It's not overstating it to say that the use of Deep Learning is becoming increasingly prevalent across all industries. For example, Deep Learning is trying to help the health care industry for tasks such as cancer detection and drug discovery. In the internet service and mobile phone industries, we can see various apps which are using deep learning for image/video classification and speech recognition. For example, Google Voice, Apple Siri, Microsoft Skype, and so on. In media, entertainment, and news, we can see applications such as video captioning, real-time translation and personalization, or recommendation systems such as Netflix. In autonomous cars, Deep Learning is trying to overcome key concerns, such as sign and passenger detection or for lane tracking. In Security, Deep Learning is used for face recognition and video surveillance. Deep Learning is used in many other fields and domains as well.

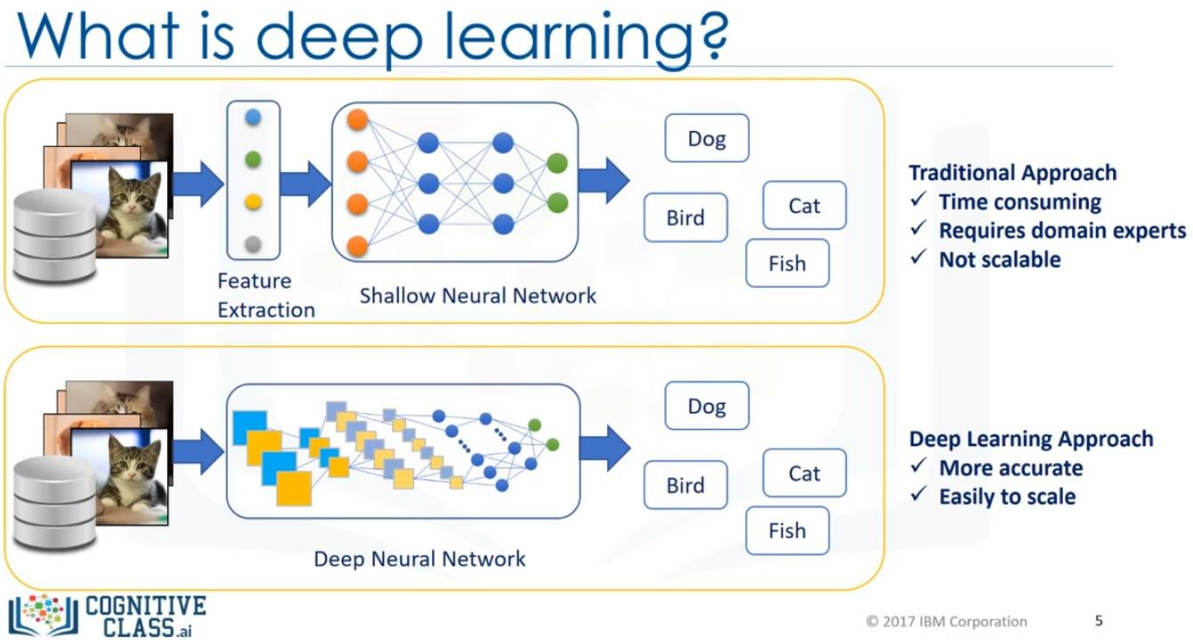
Looking at all of these industries, we can see that the increasing popularity of Deep Learning today is due to three reasons:

- first, in the dramatic increases in computer processing capabilities;
- second, in the availability of massive amounts of data for training computer systems;
- and third, in the advances in machine learning algorithms and research.



Now, let's take a closer look at deep learning and see why it's such a hot topic today?
Assume that you have a dataset of images of animals, such as cats and dogs, and you want to build a model that can recognize and differentiate them. This model is supposed to look at the sample set of images, learn from the images, and get trained. Later, given an image, we should be able to use this trained model to recognize the image as either a dog or a cat. Though this looks like a simple task, even for a very young child, computers often have difficulties in overcoming this task.

总结:



So, let's see how computers address these difficulties.

Traditionally, your first step in building such a model would be "feature extraction and feature selection." That is, to choose the best features from your images, and then to use those features in a classification algorithm, such as a shallow Neural Network. Doing this would result in a model that, given an image, could predict "cat" or "dog." Those chosen features could simply be the color, object edges, pixel location, or countless other features that could be extracted from the images. Of course, the better and more effective you are at finding feature sets, the more accurate and efficient you can become at image-classification. In fact, in the last two decades, there has been a lot of scientific research in image processing related to finding the best feature sets within images for the purposes of classification. However, as you can imagine, the process of selecting and using the best features is a tremendously time-consuming task and is often ineffective. Furthermore, extending the features to other types of images becomes an even greater problem – because the features you've used to discriminate cats and dogs, cannot be easily generalized to things like recognizing hand-written digits, for example. Therefore, the importance of effectively and accurately selecting features can't be overstated.

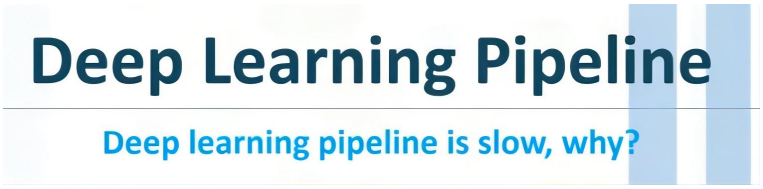
Enter "deep neural networks" – such as Convolutional Neural Networks. Suddenly, without having to find or select features, this network automatically and effectively finds the best features for you. So instead of you choosing what image features to classify dogs versus cats, Convolutional Neural Networks can automatically find those features and classify the images for you.

So, we can say Deep Learning is an algorithm that learns directly from samples much better than traditional approaches.

Thanks for watching this video.

总结:

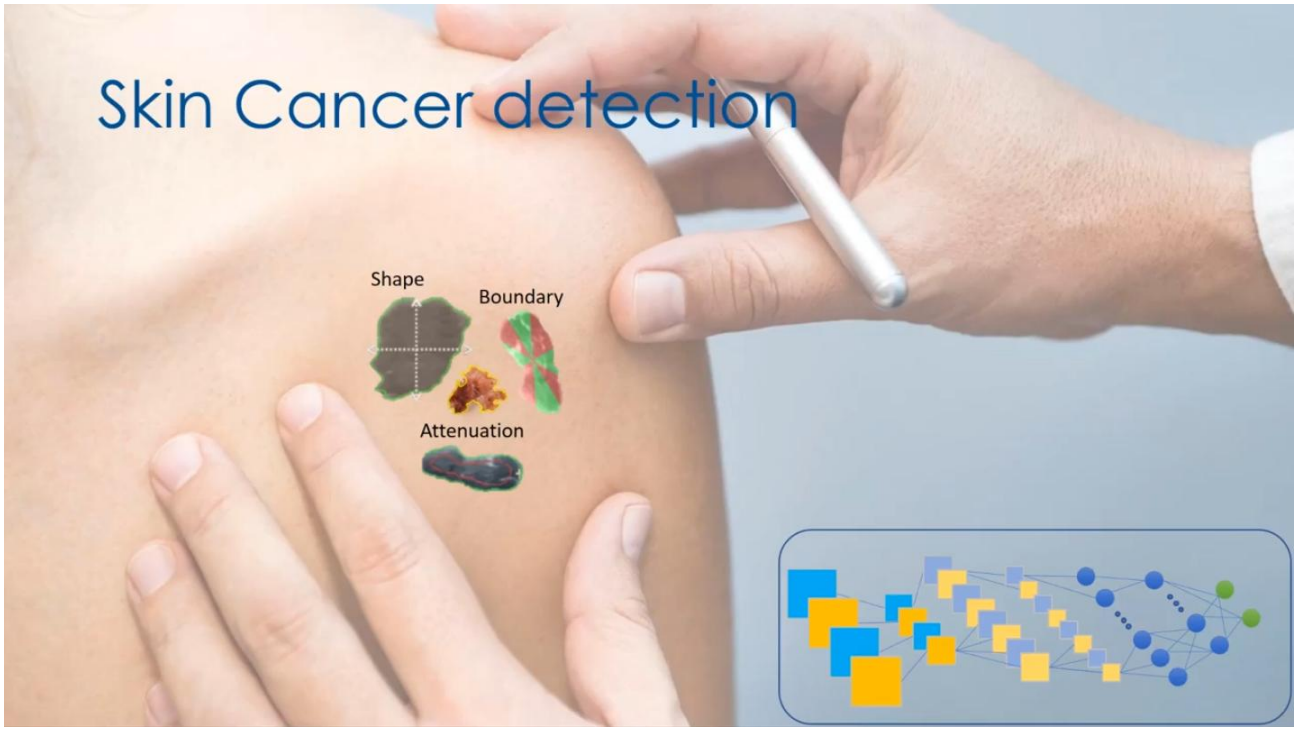
1.2 Deep Learning Pipeline



In this video, we will look at real world problems that deep learning tries to solve, as well as the deep learning pipeline. We'll also investigate why the deep learning pipeline is slow.

To be honest, the problems that most data scientists are trying to solve when using deep learning revolve around things that are much more serious than recognizing cats or dogs from images. Indeed, Deep Learning is being used to solve major issues in the health care industry, such as skin cancer, which is the most commonly diagnosed cancer in the United States.
Over five million cases are diagnosed each year, costing the U.S. healthcare system over $8 billion. More than 100,000 of those cases involve melanoma, the deadliest form of skin cancer, which leads to over 9,000 deaths a year.

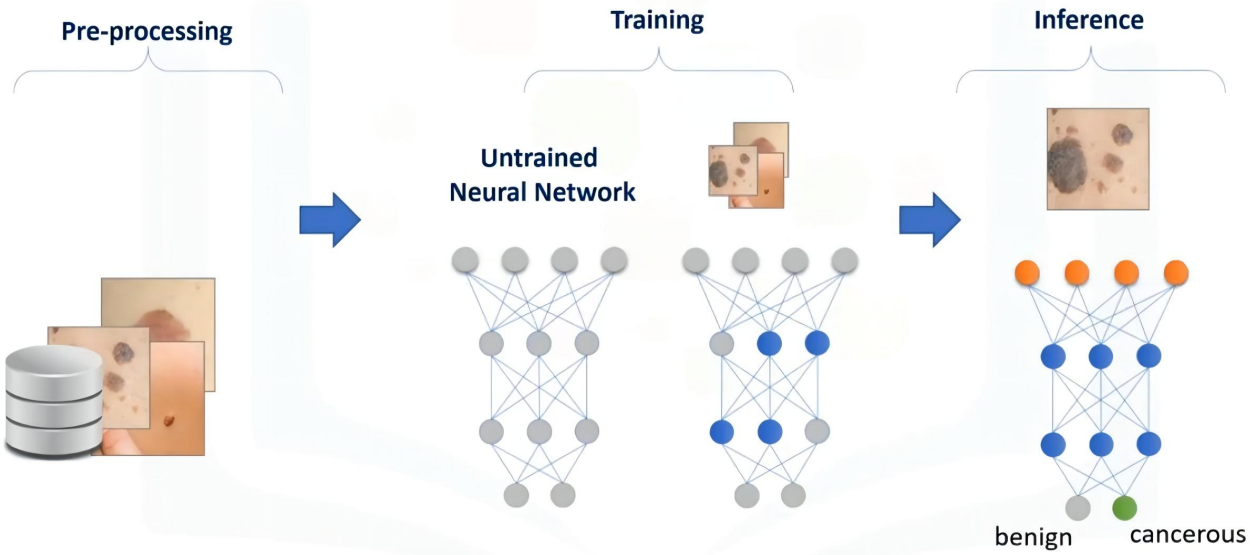
The key question becomes, "How can the use of computers solve issues that even leading physicians can't overcome?" While diagnosing melanoma early is the key to a patient's survival, there are a number of challenges in doing so. First, there is a limited supply of doctors specializing in melanoma, making them costly to visit or difficult to access in many geographic regions. Second, to the naked eye, there might be no obvious difference between a healthy mole and life-threatening melanoma. Even trained general practitioners, dermatologists, and surgeons struggle to interpret complex lesion morphology. Third, it can sometimes takes weeks for lesions to be examined, biopsied and sent to pathology for diagnosis.
So, what is the solution here?



The solution is a computer system, smart enough and quick enough, to "see" cancer before a doctor with a microscope can recognize it. **A Deep Neural Networks model might be able to solve this problem**. A network that can say, with high accuracy, "this image is showing a cancer" or "this image is showing a benign tumour" almost the same way that a classification model can say "this is a dog" or "this is a cat."
In this case, we can build a **convolutional neural network** (or **CNN**) to detect edges, boundaries, shapes, and attenuation of moles as features and use it to detect if a mole is 'cancerous' or 'not cancerous'. In this Deep Neural Networks model, lower layers detect features like texture, irregular boundary structures, and so on. Intermediate layers may recognize entire lesions of different shapes and sizes.
Ok, all of this is very good. So now let's put this all together to see what we should have.

总结:

Training and Inference



Basically, if we look at the pipeline of our deep learning model, we can see the following phases:

- First, pre-processing input data.
- Second, training the deep learning model.
- And third, inference and deployment of the model.

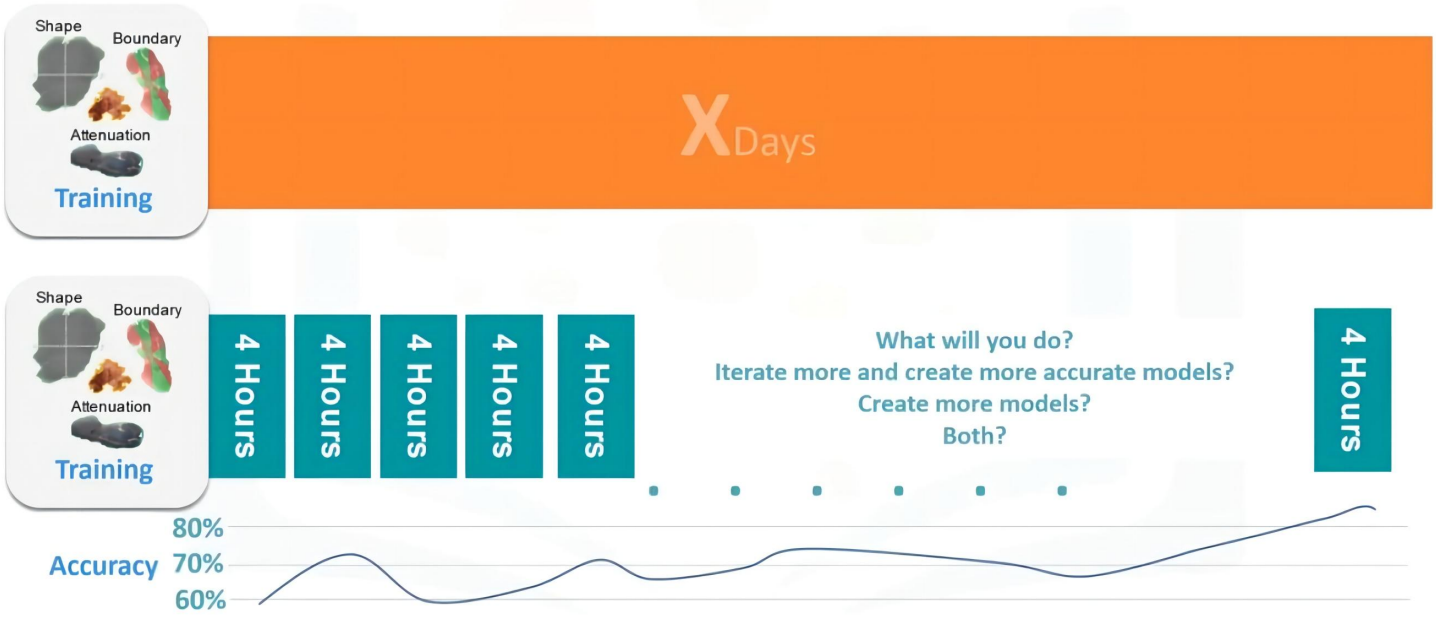
First we have to convert the images into a readable and proper format for our network.
Then, an untrained network is fed with a big dataset of images in the Training phase, prompting the network to learn.
Finally, we use the trained model in the Inference phase, which classifies a new image by inferring its similarity to the trained model. This model can be deployed and used as a melanoma detector.

But we should consider that, in general, **this pipeline is very slow for 3 reasons**:

- First, training a deep neural network is basically a slow process.
- Second, building a deep neural network is an iterative process for data scientists, that is, it needs optimization and tuning, and data scientists need to run it many times to make it ready to be used.
- And third, the trained model needs to get updated periodically, because new data needs to be added to the training set.

Because of these factors, the process is generally very slow.

Acceleration training .... days become hours



总结:

This leads to the question, **“How slow is the pipeline?”** As a rule of thumb: the smaller your system, the more time you’ll need to get a trained model that performs well enough.

For example, building a recognition model might take multiple days with a simple Intel x86 architecture. And that’s not surprising, given that we’re working with a training set of a billion pieces of data, including radiological and photographic images of melanomas and carcinogenic lesions and tumors. Given these volumes, it’s not surprising that it would take a few days to train the neural network, each time a data scientist wants to incorporate new data. Again, **this wouldn’t be so bad if it we were a do it once. But the process is an iterative one, in which we’re trying to set and reset parameters, and run it again and again trying to make it better**.

Now, think about a system that can complete this process in just 4 hours. Imagine what you could do with your model then. You might decide to train your model more frequently, which would make it much more accurate. Or you can **run the model many times and build two models that will merge to become an ‘ensemble’ model** (which is data science talk for when people mix multiple Machine Learning nets into a higher level one).

So, **training time is one of the key components to productivity, and indeed a key metric for deep learning**. But, how can we accomplish this? We’ll answer this question in the next video. Thanks for watching this video.

1.3 Graded Review Questions (2 Questions)

Which are applications of deep learning in the industry?

- ○ In Security: face recognition and video surveillance
- ○ In Media: entertainment and news
- ○ In Communications: internet service and mobile phones industries
- ◉ All of the above

✔

Which is **NOT** one of the main phases of a deep learning pipeline?

- ○ Preprocessing input data
- ◉ Feature selection
- ○ Training the deep learning model
- ○ Inference and Deployment of the model

✔

总结：