Who

* Researcher
* Audience

What

* Study how depth of a CNN affects a model in image classification

Where & When

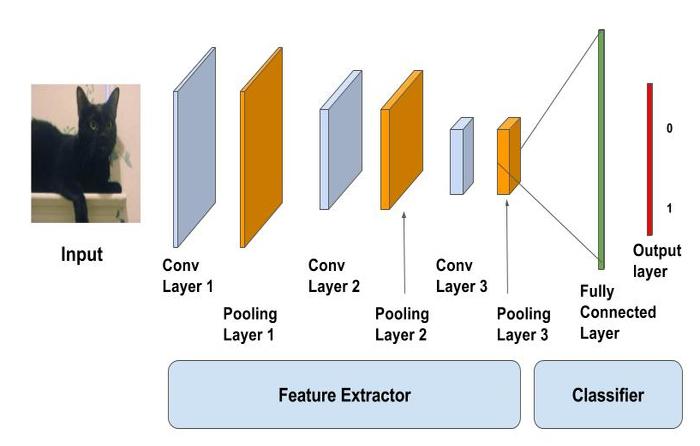
* Framing by explaining what is deep learning, and it’s utility in image classification
* Explain “deep” cuz it helps to explain depth of a network later (they are the sub-WHATs):
  + What is “deep learning”
  + What is image classification
  + What is CNN

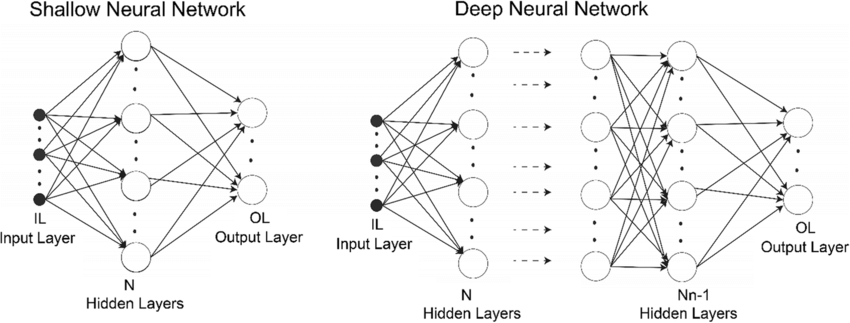
How

* Systematically add more layers. And observe.

Why

* To better answer “How deep is too deep?”





Imagine your coworker slides you a photo of their daughter. You could’ve offended your coworker if you said “Oh, your daughter looks adorable in that blue collar.” But you won’t // because luckily, your mind is able to tell that on the right – it’s furry, it’s on all fours, it has triangular ears: it’s a dog -- while on the left – it’s standing on 2 legs, significantly less hairy, no triangle ears – it’s a human child. That process of picking apart the defining features of an object is second nature for almost everyone, but not so for machines.

Computer Scientists have long since been developing machine learning algorithms for “Image Classification”. If we look deeper into machine learning, there’s a subfield called “Deep Learning” which uses Artificial neural networks or ANNs to learn and become better at classifying images. ANNs consists of the “Input”, “Hidden” and “Output Layers”. And if they have more than 1 hidden layer, it becomes a deep ANN. One of the types of ANN best suited for image classification is the Convolutional Neural Network, or CNN.

As a type of ANN, the Convolutional Neural Network also has 3 layers but within its hidden layer, there’s specialized sub-layers that carries out a 2-step routine for image classification: the Feature Extraction Step and the Classification step. Coming back to your coworker’s photo, if your brain that identifies 4 legs, furry face and triangle ears that’s the Feature Extraction Step while if your brain that tells you that (4 legs + furry face + triangle ears = DOG), it’s the Classification Step. The sublayer responsible for feature extraction is a “convolution” layer and the classification Step is done by the “Fully-Connected” layers.

A convolution layer extracts an image’s distinguishing features using many “filters, while Fully-Connected layer, has “neurons” that determine how much each feature from the previous step / contributes to identifying an image. ~~For example, 4 legs, furry face & triangle ears is easier to identify as a dog compared to 4 legs, furry face and it has eyes.~~

In general, more filters = more features, which is commonly a good thing. However, research warned against too many convolutional layers as it would lead to dramatically longer training times and are at risk of overfitting. Now, Overfitting occurs when during validation, a **white** dog is classified as a dog, but during testing a **black** dog is classified as a horse. Therefore, in this project, we seek to answer the question “How deep is too deep?”

A systematic approach was set-up. Starting with a baseline model with only 1 convolutional layer, we incrementally add one more conv layer until we have a maximum of 4 layers.Then we observed how good the model is at predicting an input image’s class.

What I found was interesting, the models follow the general trend that the more convolutional layers, the more accurate the model’s predictions for both the training and testing dataset. Going into this project, it was expected that by 4 convolutional layers, overfitting would occur. Most likely it didn’t because the dataset are large and varied enough to include objects or animals of different colours, sizes, orientation, and types.

So have we found the answer to “How deep is too deep?” Well, from the findings of this project, deep enough to consistently make correct predictions but not too deep that it takes forever to train or end up calling a dog, a child.

6:06 minutes

4:15 minutes

3:30 minutes

3:12 minutes

2:56 minutes