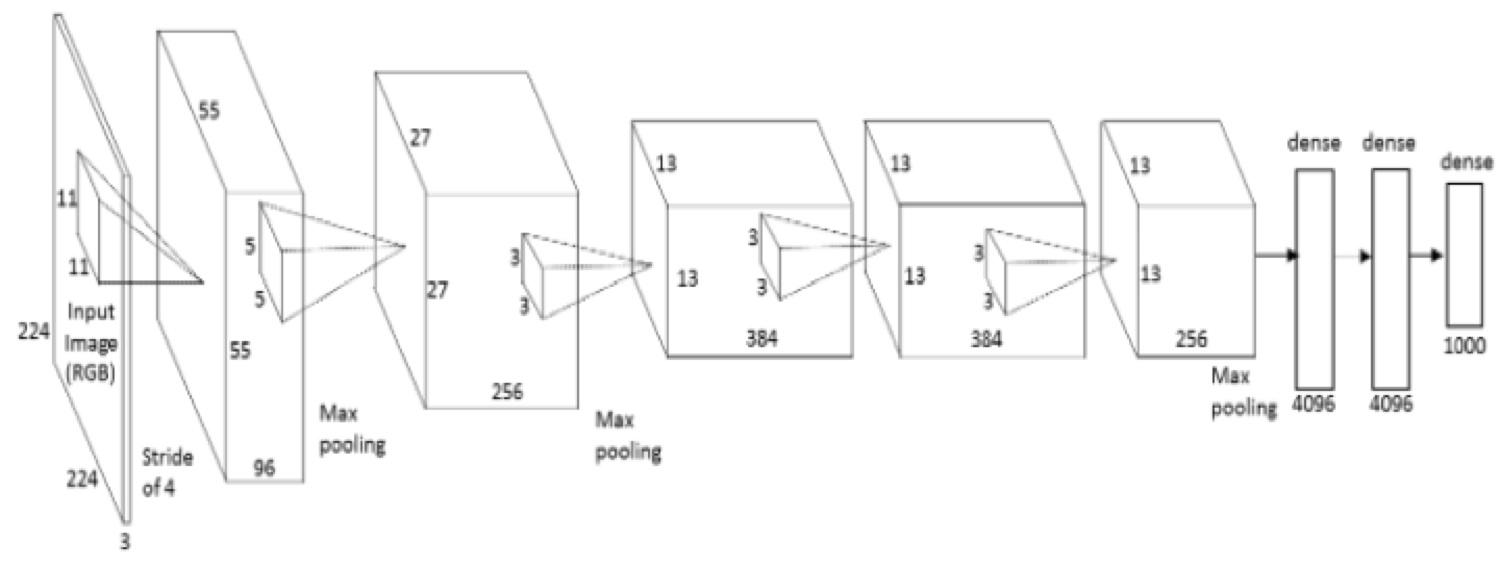
Tools for Machine Learning:

# AlexNet (2012)

The one that started it all (Though some may say that Yann LeCun’s [paper](http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf) in 1998 was the real pioneering publication). This paper, titled “ImageNet Classification with Deep Convolutional Networks”, has been cited a total of 6,184 times and is widely regarded as one of the most influential publications in the field. Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton created a “large, deep convolutional neural network” that was used to win the 2012 ILSVRC (ImageNet Large-Scale Visual Recognition Challenge). For those that aren’t familiar, this competition can be thought of as the annual Olympics of computer vision, where teams from across the world compete to see who has the best computer vision model for tasks such as classification, localization, detection, and more. 2012 marked the first year where a CNN was used to achieve a top 5 test error rate of 15.4% (Top 5 error is the rate at which, given an image, the model does not output the correct label with its top 5 predictions). The next best entry achieved an error of 26.2%, which was an astounding improvement that pretty much shocked the computer vision community. Safe to say, CNNs became household names in the competition from then on out.

In the paper, the group discussed the architecture of the network (which was called AlexNet). They used a relatively simple layout, compared to modern architectures. The network was made up of 5 conv layers, max-pooling layers, dropout layers, and 3 fully connected layers. The network they designed was used for classification with 1000 possible categories.



**Main Points**

* Trained the network on ImageNet data, which contained over 15 million annotated images from a total of over 22,000 categories.
* Used ReLU for the nonlinearity functions (Found to decrease training time as ReLUs are several times faster than the conventional tanh function).
* Used data augmentation techniques that consisted of image translations, horizontal reflections, and patch extractions.
* Implemented dropout layers in order to combat the problem of overfitting to the training data.
* Trained the model using batch stochastic gradient descent, with specific values for momentum and weight decay.
* Trained on two GTX 580 GPUs for **five to six days**.

**Why It’s Important**

The neural network developed by Krizhevsky, Sutskever, and Hinton in 2012 was the coming out party for CNNs in the computer vision community. This was the first time a model performed so well on a historically difficult ImageNet dataset. Utilizing techniques that are still used today, such as data augmentation and dropout, this paper really illustrated the benefits of CNNs and backed them up with record breaking performance in the competition.