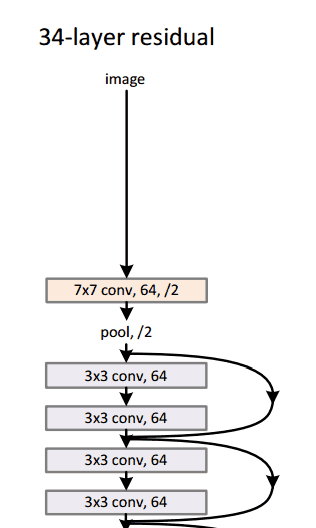
Tools for Machine Learning:

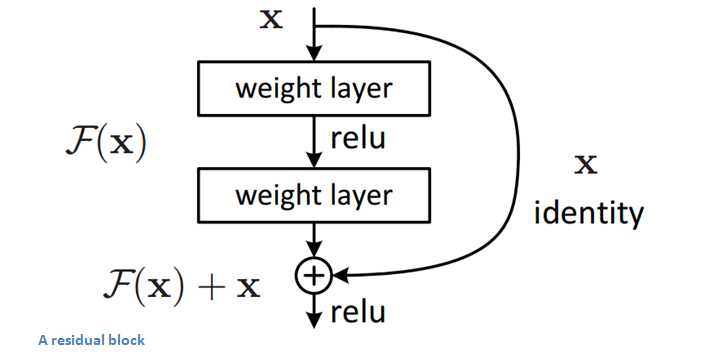
# Microsoft ResNet (2015)

Imagine a deep CNN architecture. Take that, double the number of layers, add a couple more, and it still probably isn’t as deep as the ResNet architecture that Microsoft Research Asia came up with in late 2015. ResNet is a new 152 layer network architecture that set new records in classification, detection, and localization through one incredible architecture. Aside from the new record in terms of number of layers, ResNet won ILSVRC 2015 with an incredible error rate of 3.6% (Depending on their skill and expertise, humans generally hover around a 5-10% error rate. See Andrej Karpathy’s [great post](http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/) on his experiences with competing against ConvNets on the ImageNet challenge).



**Residual Block**

The idea behind a residual block is that you have your input x go through conv-relu-conv series. This will give you some F(x). That result is then added to the original input x. Let’s call that H(x) = F(x) + x. In traditional CNNs, your H(x) would just be equal to F(x) right? So, instead of just computing that transformation (straight from x to F(x)), we’re computing the term that you have to *add*, F(x), to your input, x. Basically, the mini module shown below is computing a “delta” or a slight change to the original input x to get a slightly altered representation (When we think of traditional CNNs, we go from x to F(x) which is a completely new representation that doesn’t keep any information about the original x). The authors believe that “it is easier to optimize the residual mapping than to optimize the original, unreferenced mapping”.



Another reason for why this residual block might be effective is that during the backward pass of backpropagation, the gradient will flow easily through the effective because we have addition operations, which distributes the gradient.

**Main Points**

* “Ultra-deep” – Yann LeCun.
* 152 layers…
* Interesting note that after only the *first 2* layers, the spatial size gets compressed from an input volume of 224x224 to a 56x56 volume.
* Authors claim that a naïve increase of layers in plain nets result in higher training and test error (Figure 1 in the [paper](https://arxiv.org/pdf/1512.03385v1.pdf)).
* The group tried a 1202-layer network, but got a lower test accuracy, presumably due to overfitting.
* Trained on an 8 GPU machine for **two to three weeks**.

**Why It’s Important**

3.6% error rate. That itself should be enough to convince you. The ResNet model is the best CNN architecture that we currently have and is a great innovation for the idea of residual learning. With error rates dropping every year since 2012, I’m skeptical about whether or not they will go down for ILSVRC 2016. I believe we’ve gotten to the point where stacking more layers on top of each other isn’t going to result in a substantial performance boost. There would definitely have to be creative new architectures like we’ve seen the last 2 years. On September 16th, the results for this year’s competition will be released. Mark your calendar.