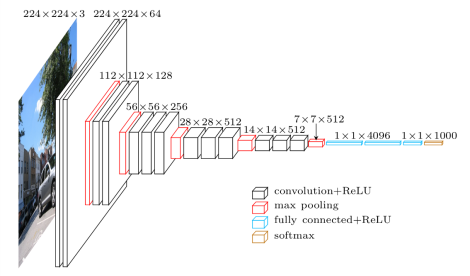
**SSD MULTIBOX — REAL-TIME OBJECT DETECTION IN DEEP LEARNING**

In the field of computer vision, convolution neural networks excel at image classification, which consists of categorising images, given a set of classes and having the network determine the strongest class present in the image.

1. **Single Shot:**This means that the tasks of object localization and classification are done in a single forward pass of the network. As you can see from the diagram above, SSD’s architecture builds on the venerable VGG-16 architecture, but discards the fully connected layers. The reason VGG-16 was used as the base network is because of its strong performance in high quality image classification tasks and its popularity for problems where transfer learning helps in improving results. Instead of the original VGG fully connected layers, a set of auxiliary convolutional layers (from conv6 onwards) were added, thus enabling to extract features at multiple scales and progressively decrease the size of the input to each subsequent layer.



VGG architecture (input is 224x224x3)

## **MultiBox**

MultiBox starts with the priors as predictions and attempt to regress closer to the ground truth bounding boxes. More default boxes results in more accurate detection, although there is an impact on speed. Having MultiBox on multiple layers results in better detection as well, due to the detector running on features at multiple resolutions. MultiBox’s loss function also combined two critical components that made their way into SSD:

* **Confidence Loss**: this measures how confident the network is of the *objectness*of the computed bounding box. Categorical [cross-entropy](https://rdipietro.github.io/friendly-intro-to-cross-entropy-loss/#cross-entropy) is used to compute this loss.
* **Location Loss:**this measures how *far away*the network’s predicted bounding boxes are from the ground truth ones from the training set. [L2-Norm](https://rorasa.wordpress.com/2012/05/13/l0-norm-l1-norm-l2-norm-l-infinity-norm/) is used here.