Here is the flow:

Convolution relu pool

Every image could be represented as a 3-d matrix( and not 2-d as we expect, 3rd dimension is the channels , i.e. r g b , each consisting values between 0 -255)

Convolution is basically doing matrix dot multiplication.

Usually we don’t take the complete image and multiple it since it becomes way to computational expensive , so we use something that we learned in Computer graphics that’s the window… small piece of the image. And then we slide it across , just as like we are doing when cropping a photo say for instance.

What Pooling basically does is , that it reduces the size of the matrix we got after convolutional operation by taking the most important part of that part so that to reduce the size of matrix since its very big to analysis the matrix we get after convolution. Usually max pooling is used, i.e . taking the maximum out the the pool window.

ReLU : Rectified Linear Unit, so why we use ReLU is that to detect a feature the matrix we a multiplying with should only contain values which has some feature, right and the rest should contain a 0 so as to remove all other non requiring or non focused features after convolution.