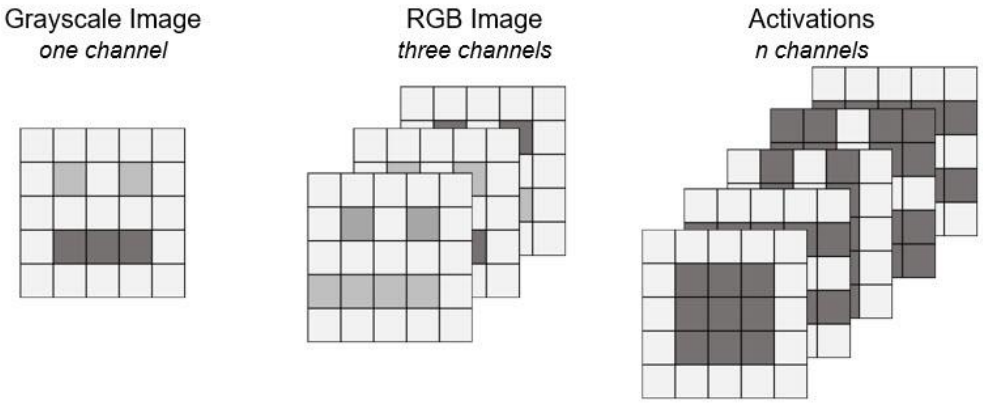
Viewing Activations

Convolution layers are made of weights and biases that are used to filter an input image. The output of a convolution layer is a set of filtered images. This output is called the *activations* of that layer.

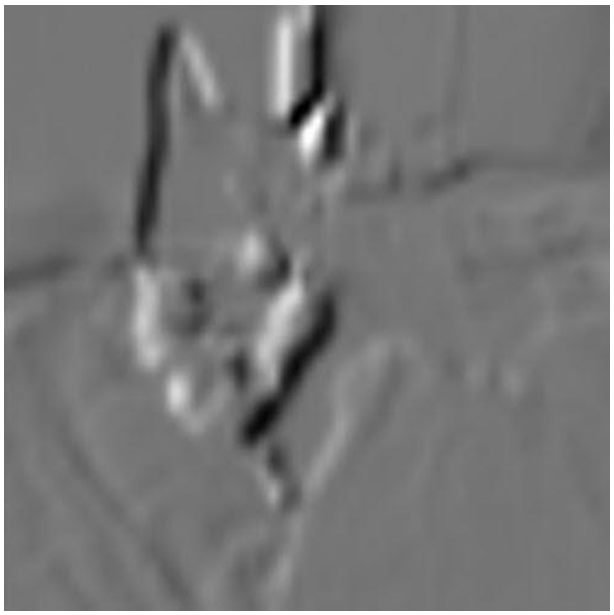
These activations are a 3-D array, where the third dimension is often called a *channel*.



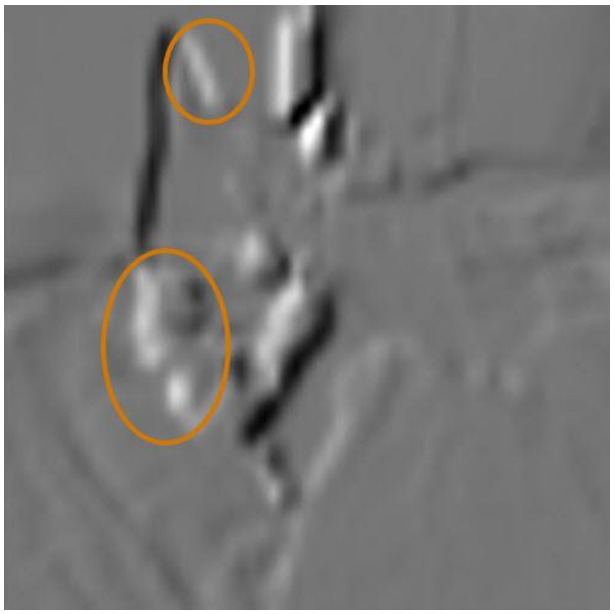
Many images have multiple channels. For example, RGB images have three (red, green, and blue). Each channel can be visualized together as a color image, or separately as grayscale images.

There is one output channel for each filter in a convolution layer. A convolution layer can have hundreds of filters, so each layer can create hundreds of channels. You will visualize each channel as a grayscale image.

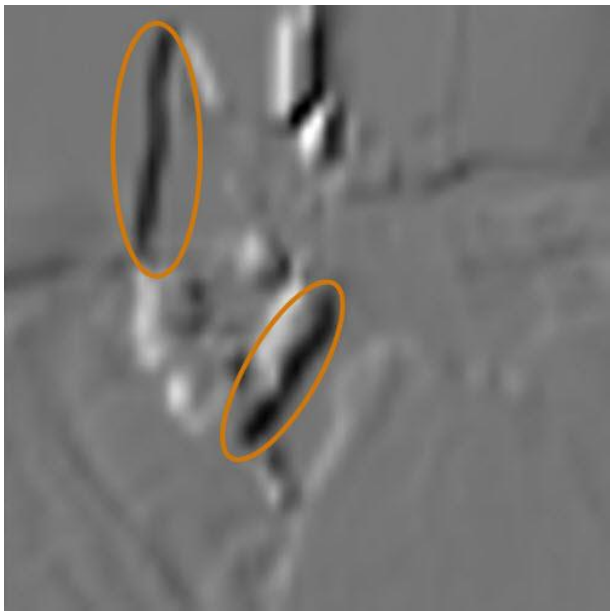
Consider this input image of a cat.



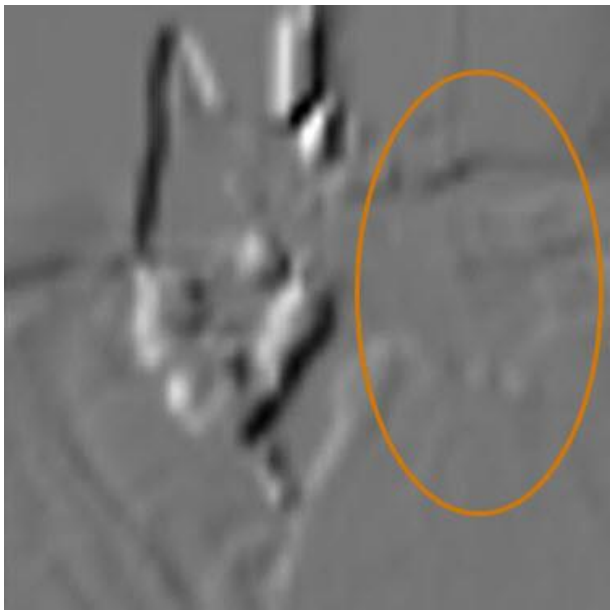
This grayscale image is one channel of the filtered cat image. The filter that produced this image finds edges.



Convolution layers process images from left to right. This filter finds edges from dark to light.
The positive activations are in places where the dark fur is left of the white wall or white fur.



The negative activations show the opposite. The white wall is left of the black fur.



Gray activations show nothing interesting. This filter did not find any edges in the right half of the image.