Adding Fire

Add fire modules to the model and learn about delayed downsampling.

Chapter Goals:

- Add a multi-fire module block to the model
- Learn about delayed downsampling

A. Delayed downsampling

While reducing the data dimensions via max pooling can help our model training, strategic placement of max pooling layers can improve the SqueezeNet model's accuracy. Rather than applying max pooling within the first fire module (e.g. after the squeeze layer), we instead apply it after the first two fire modules.

Placing the max pooling layers later in the model structure is referred to as delayed downsampling. We wait until later in the model structure to downsample the data (i.e. reduce height/width dimensions through pooling), so the earlier convolution layers can have a larger dimension input. The creators of SqueezeNet showed that having larger dimension inputs for more layers in the model can help improve its accuracy.

Time to Code!

In this chapter, we'll continue building the model_layers function by adding the first multi-fire module.

We'll start by setting the list of parameter tuples to describe our fire modules. The list will contain two tuples representing the fire_module arguments. For both fire modules, we use 32 as the squeeze depth and 64 as the expand depth. The first fire module will be named 'fire1' while the second will be named 'fire2'.

Set fire_params1 equal to a list of two tuples with the specified values.

Now we apply the multi-fire module block with our list of tuple parameters.

Set multi_fire1 equal to self.multi_fire_module with pool1 as the first argument and fire_params1 as the second argument.

We perform max pooling on the output of the multi-fire module block.

Set pool2 equal to self.custom_max_pooling2d applied with multi_fire1 as the first argument and 'pool2' as the name.

```
import tensorflow as tf
    class SqueezeNetModel(object):
        # Model Initialization
        def __init__(self, original
            self.original_dim = or:
            self.resize dim = resi;
            self.output_size = outp
        # Convolution layer wrapper
        def custom_conv2d(self, in
11
            return tf.layers.conv2
12
13
                inputs=inputs,
                filters=filters,
                kernel_size=kernel
                activation=tf.nn.re
                padding='same',
                name=name)
        # Max pooling layer wrappe
        def custom_max_pooling2d(se
            return tf.layers.max_po
                inputs=inputs,
                pool_size=[2, 2],
24
                strides=2,
                name=name)
        def fire_module(self, input
            with tf.variable scope
                squeezed inputs
                                                                                   []
```