ECE/CS/ME 539 – Fall 2024 — Activity Solution 24

(a) Output Size of the First Conv2D Layer

The input tensor has shape $N \times 3 \times 32 \times 32$. After applying the first Conv2D layer with a 5×5 kernel, stride of 1, and no padding, the output dimensions are calculated as follows:

$$\text{Output Height/Width} = \frac{\text{Input Size} - \text{Kernel Size} + 2 \times \text{Padding}}{\text{Stride}} + 1$$

Substituting the values:

$$\text{Output Height} = \text{Output Width} = \frac{32 - 5 + 0}{1} + 1 = 28$$

Thus, the output shape after the first Conv2D layer is $N \times 32 \times 28 \times 28$.

(b) Trainable Parameters in the First Conv2D Layer

The number of trainable parameters in the first Conv2D layer includes both weights and biases.

- **Weight Parameters:** There are 32 filters, each with a 5×5 kernel over 3 input channels, giving:

$$3 \times 32 \times 5 \times 5 = 2400$$

- **Bias Parameters:** Each filter has one bias, so:

$$32$$

The total number of trainable parameters is:

$$2400 + 32 = 2432$$

(c) FLOPs for the First Conv2D Layer

The costliest operation is multiplication. To compute the output for the first layer:

- Each application of the kernel (shape $3 \times 32 \times 5 \times 5$) to a single position requires 2400 multiplications.
- The output dimensions after applying the convolution are 28×28, so the total number of multiplications is:

$$2400 \times 784 = 1,881,600$$

Therefore, the model performs $1,881,600$ multiplications for the first layer.

(d) Filters in the Second Conv2D Layer

In the second Conv2D layer:

- **Number of Filters:** 64
- **Kernel Size:** Each kernel has a shape of $3 \times 3 \times 32$.

(e) Output of the Flatten Layer

To find the output shape after the Flatten layer, we compute the activation shapes for each layer:

- **After Conv1:** $N \times 32 \times 28 \times 28$
- **After Pool1:** $N \times 32 \times 14 \times 14$
- **After Conv2:** $N \times 64 \times 12 \times 12$
- **After Pool2:** $N \times 64 \times 6 \times 6$
- **After Conv3:** $N \times 64 \times 4 \times 4$

  The Flatten layer reshapes $N \times 64 \times 4 \times 4$ into a vector of length:

$$4 \times 4 \times 64 = 1024$$