Chapter 10: Convolutional networks

Problem 10.1

Show that the operation in equation 10.4 is equivariant with respect to translation.

$$h_i = \mathrm{a}[\beta + \omega_1 x_{i-1} + \omega_2 x_i + \omega_3 x_{i+1}] \tag{10.4}$$

A function is equivariant to a transformation $t(x)$ if $f(t(x)) = t(f(x))$. In this case, the transformation is translation, so $t(x) = x + \delta$. Let's denote the transformed sequence as $\{x_i'\}$ and the function as h_i':

$$h_i' = \mathrm{a}[\beta + \omega_1 x_{i-1}' + \omega_2 x_i' + \omega_3 x_{i+1}']$$

Substitute $x_i' = x_i + \delta$:

$$h_{i+\delta} = \mathrm{a}[\beta + \omega_1 x_{i+\delta-1} + \omega_2 x_{i+\delta} + \omega_3 x_{i+\delta+1}]$$

Which shows that $h_i' = h_{i+\delta}$, demonstrating that this operation is is equivariant with respect to translation by any integer δ.

Problem 10.2

Equation 10.3 defines 1D convolution with a kernel size of three, stride of one, and dilation one. Write out the equivalent equation for the 1D convolution with a kernel size of three and a stride of two as pictured in figure 10.3a-b.

$$z_i = \omega_1 x_{i-1} + \omega_2 x_i + \omega_3 x_{i+1} \tag{10.3}$$

Figure 10.3a shows $0, x1$, and x_3 corresponding to weightes w_1, w_2, and w_3 respectively, giving output z_1. The next output z_2 is computed using x_2, x_3, and x_4, corresponding to weights w_1, w_2, and w_3 respectively. The equations for this are therefore:

$$\begin{aligned} z_1 &= \omega_1 x_0 + \omega_2 x_1 + \omega_3 x_2 \\ z_2 &= \omega_1 x_2 + \omega_2 x_3 + \omega_3 x_4 \end{aligned}$$

Which generalises to:

$$z_i = \omega_1 x_{2i-2} + \omega_2 x_{2i-i} + \omega_3 x_{2i}$$

This equation shows that as you increase i, you skip every second element of i to compute the next output, consistent with a stride of two.

Problem 10.3

> Write out the equation for the 1D dilated convolution with a kernel size of three and a dilation rate of two, as pictured in figure 10.3d.

The equation for the 1D dilated convolution with a kernel size of three, stride of one, and a dilation rate of two is shown in figure 10.3d is:

$$z_5 = \omega_1 x_3 + \omega_2 x_5 + \omega_3 x_7$$

This generalises to:

$$z_i = \omega_1 x_{i-2} + \omega_2 x_i + \omega_3 x_{i+2} \tag{1}$$

Problem 10.4

> Write out the equation for a 1D convolution with kernel size of seven, a dilation rate of three, and a stride of three

The equation for the 1D convolution with a kernel size of seven is:

$$z_i = \omega_1 x_{i-3} + \omega_2 x_{i-2} + \omega_3 x_{i-1} + \omega_4 x_i + \omega_5 x_{i+1} + \omega_6 x_{i+2} + \omega_7 x_{i+3}$$

Adding a dilation rate of three, the equation becomes:

$$z_i = \omega_1 x_{i-9} + \omega_2 x_{i-6} + \omega_3 x_{i-3} + \omega_4 x_i + \omega_5 x_{i+3} + \omega_6 x_{i+6} + \omega_7 x_{i+9}$$

Adding a stride of three, the equation becomes:

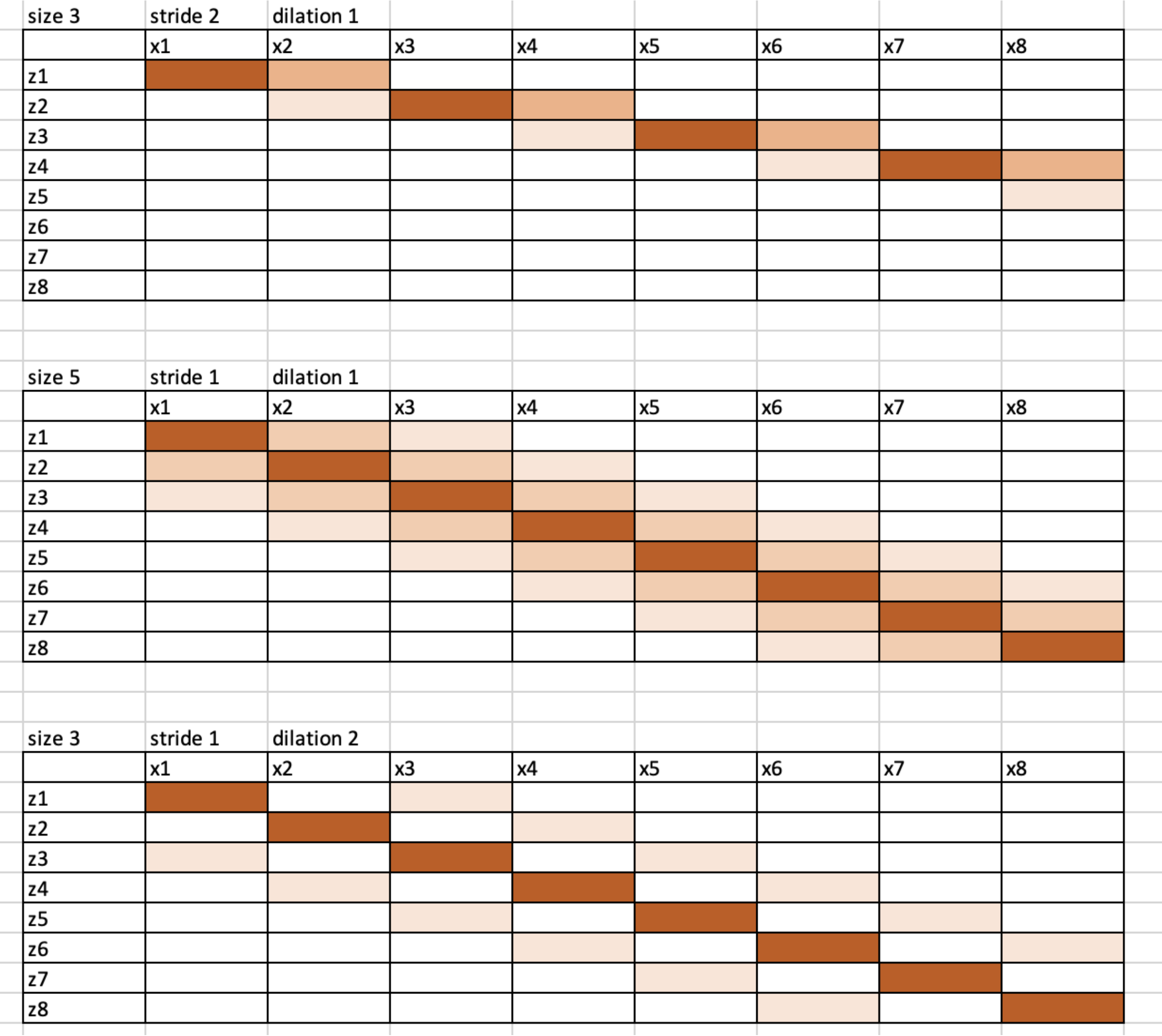
$$z_i = \omega_1 x_{3i-6} + \omega_2 x_{3i-3} + \omega_3 x_{3i} + \omega_4 x_{3i+3} + \omega_5 x_{3i+6} + \omega_6 x_{3i+9} + \omega_7 x_{3i+12}$$

i.e., When the dilation rate is d and the kernel size is k, the generic formula for the index of the input element contributing to the output z_i at position j within the kernel is:

$$\text{Input index} = \text{stride} \times (i-1) + d \times (j-1)$$

Problem 10.5

Draw weight matrices in the style of figure 10.4d for (i) the strided convolution in figure 10.3a–b, (ii) the convolution with kernel size 5 in figure 10.3c, and (iii) the dilated convolution in figure 10.3d



size 3	stride 2	dilation 1						
	x1	x2	x3	x4	x5	x6	x7	x8
z1	■	▨						
z2		░	■	▨				
z3				░	■	▨		
z4						░	■	▨
z5								░
z6								
z7								
z8								

size 5	stride 1	dilation 1						
	x1	x2	x3	x4	x5	x6	x7	x8
z1	■	▨	░					
z2	▨	■	▨	░				
z3	░	▨	■	▨	░			
z4		░	▨	■	▨	░		
z5			░	▨	■	▨	░	
z6				░	▨	■	▨	░
z7					░	▨	■	▨
z8						░	▨	■

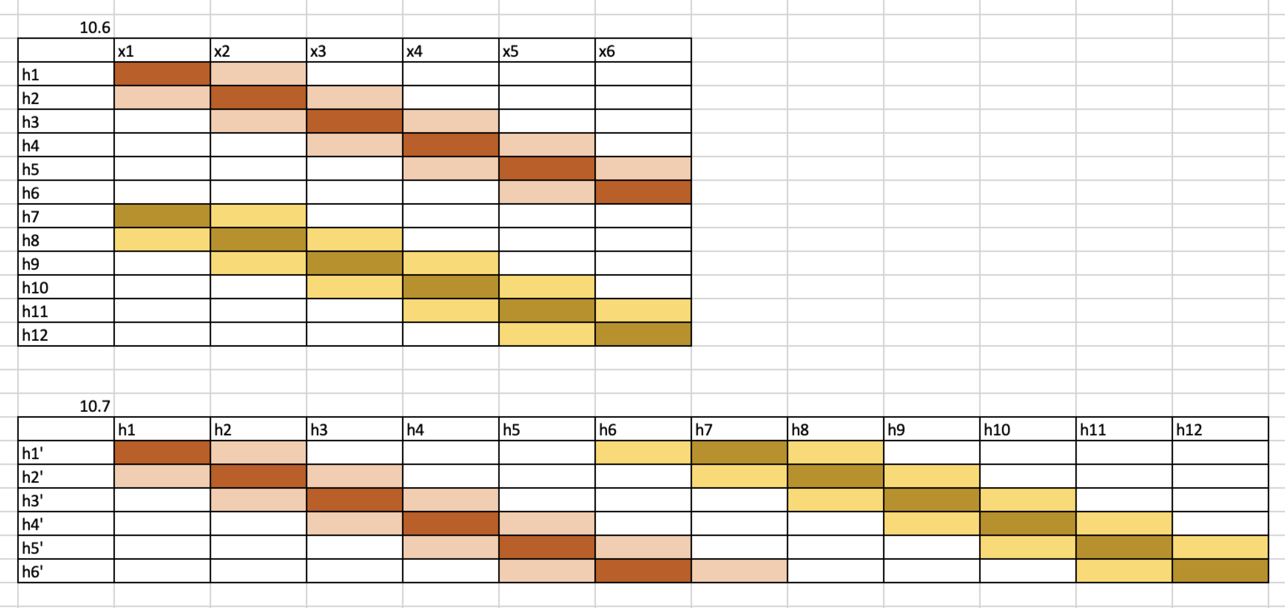
size 3	stride 1	dilation 2						
	x1	x2	x3	x4	x5	x6	x7	x8
z1	■		░					
z2		■		░				
z3	░		■		░			
z4		░		■		░		
z5			░		■		░	
z6				░		■		░
z7					░		■	
z8						░		■

Problem 10.6

Draw a 6×12 weight matrix in the style of figure 10.4d relating the inputs $x_1, \ldots, x_6$ to the outputs $h_1, \ldots, h_{12}$ in the multi-chanel convolution as depicted in figures 10.5a-b.

Problem 10.7

> Draw a 12×6 weight matrix in the style of figure 10.4d relating the inputs $h_1, \ldots, h_{12}$ to the outputs $h'_1, \ldots, h'_6$ in the multi-chanel convolution as depicted in figures 10.5c.



10.6

	x1	x2	x3	x4	x5	x6
h1	🟥	🟧				
h2	🟧	🟥	🟧			
h3		🟧	🟥	🟧		
h4			🟧	🟥	🟧	
h5				🟧	🟥	🟧
h6					🟧	🟥
h7	🟫	🟨				
h8	🟨	🟫	🟨			
h9		🟨	🟫	🟨		
h10			🟨	🟫	🟨	
h11				🟨	🟫	🟨
h12					🟨	🟫

10.7

	h1	h2	h3	h4	h5	h6	h7	h8	h9	h10	h11	h12
h1'	🟥	🟧				🟨	🟫	🟨				
h2'	🟧	🟥	🟧				🟨	🟫	🟨			
h3'		🟧	🟥	🟧				🟨	🟫	🟨		
h4'			🟧	🟥	🟧				🟨	🟫	🟨	
h5'				🟧	🟥	🟧				🟨	🟫	🟨
h6'					🟧	🟥	🟧				🟨	🟫

Problem 10.8

> Consider a 1D convolutional network where the input has three channels. The first hidden layer is computed using a kernel size of three and has four channels. The second hidden layer is computed using a kernel size of five and has ten channels. How many biases and how many weights are needed for each of these two convolutional layers?

For a 1D CNN where the input has three channels, $C_i = 3$, the first hidden layer has a kernel size of three and four channels, $K_1 = 3, C_1 = 4$, and the second hidden layer has a kernel size of five and ten channels, $K_2 = 5, C_2 = 10$, the number of weights and biases for each convolutional layer can be calculated as follows:

For the first layer, there are 3 input channels, therefore for a single kernel of size 3 there are $3 \times 3 = 9$ weights, giving a total of $9 \times 4 = 36$ weights. There is one bias per channel, giving a total of 4 biases.

For the second layer, there are 4 input channels (from the output size of layer one), therefore a single kernel of size 5 has $4 \times 5 = 20$ weights, giving a total of $20 \times 10 = 200$ weights in the second layer. There is one bias per output channel, giving a total of 10 biases.

Problem 10.9

A network consists of three 1D convolutional layers. At each layer, a zero-padded convolution with kernel size three, stride one, and dilation one is applied. What size is the receptive field of the hidden units in the third layer?

Assuming dilation of one means no dilation (as per latest edition of UDL).

The receptive field of layer one is 3, layer two is 5, and layer three is 7.

Problem 10.10

A network consists of three 1D convolutional layers. At each layer, a zero- padded convolution with kernel size seven, stride one, and dilation one is applied. What size is the receptive field of hidden units in the third layer?

The receptive field of layer one is 7, layer two is 13, and layer three is 19.

Problem 10.11

Consider a convolutional network with 1D input $\mathbf{x}$. The first hidden layer $\mathbf{H}_1$ is computed using a convolution with kernel size five, stride two, and a dilation rate of one. The second hidden layer $\mathbf{H}_2$ is computed using a convolution with kernel size three, stride one, and a dilation rate of one. The third hidden layer $\mathbf{H}_3$ is computed using a convolution with kernel size five, stride one, and a dilation rate of two. What are the receptive field sizes at each hidden layer?

The receptive field of $\mathbf{H}_1$ is 5 (= kernel size).

The receptive field of $\mathbf{H}_2$, which has a kernel size of 3 is equal to the receptive field of $\mathbf{H}_1$, plus one minus the kernel size of $\mathbf{H}_2$, multiplied by the stride of $\mathbf{H}_1$: $5 + (3 - 1) \times 2 = 9$.

If $\mathbf{H}_3$ had a dilation of 1, then to calculate the receptive field we would follow the same procedure described above: $9 + (5 - 1) \times 2 = 17$. However, $\mathbf{H}_3$ has a dilation of 2, so we multiply the above equation by 2 again to give $9 + (5 - 1) \times 2 \times 2 = 25$.

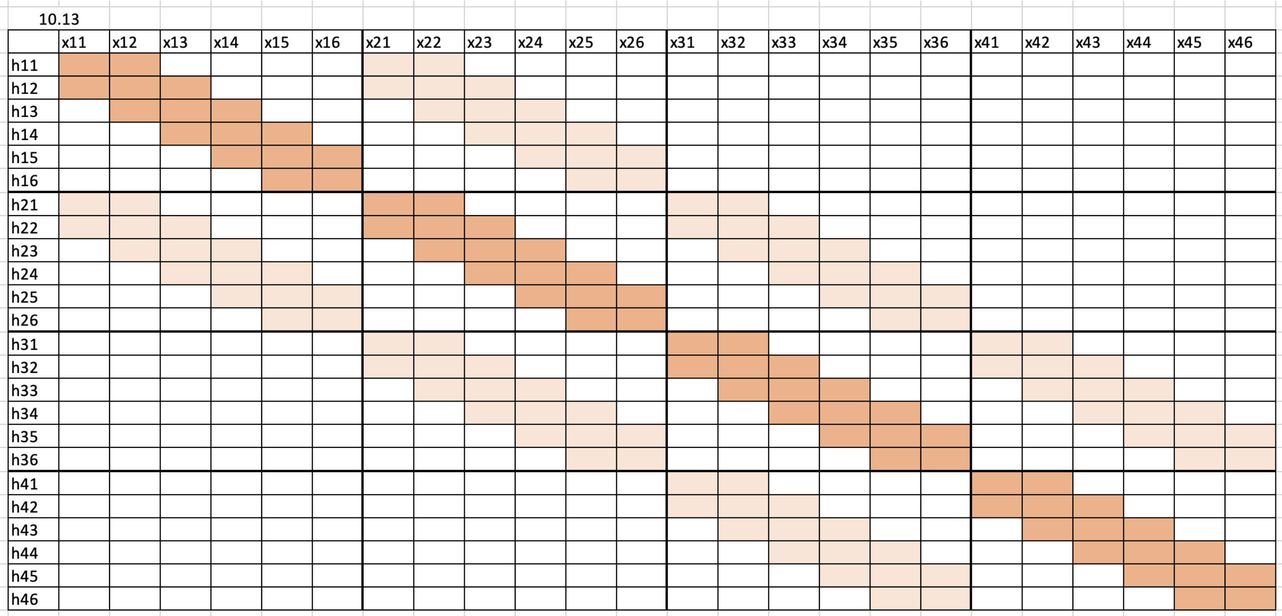
Problem 10.12

The 1D convolutional network in figure 10.7 was trained using stochastic gradient descent with a learning rate of 0.01 and a batch size of 100 on a training dataset of 4,000 examples for 100,000 steps. How many epochs was the network trained for?

An epoch consists of one full pass through the training dataset. Given the network was trained for 100,000 steps, on 4,000 examples with a batch size of 100, the number of steps per epoch is given by the number of examples divided by the batch size: $4000/100 = 40$. The number of epochs is therefore $100000/40 = 2500$.

Problem 10.13

Draw a weight matrix in the style of figure 10.4d that shows the relationship between the 24 inputs and the 24 outputs in figure 10.9.



10.13

	x11	x12	x13	x14	x15	x16	x21	x22	x23	x24	x25	x26	x31	x32	x33	x34	x35	x36	x41	x42	x43	x44	x45	x46
h11	▓	▓					░	░																
h12	▓	▓	▓				░	░	░															
h13		▓	▓	▓				░	░	░														
h14			▓	▓	▓				░	░	░													
h15				▓	▓	▓				░	░	░												
h16					▓	▓					░	░												
h21	░	░					▓	▓					░	░										
h22	░	░	░				▓	▓	▓				░	░	░									
h23		░	░	░				▓	▓	▓				░	░	░								
h24			░	░	░				▓	▓	▓				░	░	░							
h25				░	░	░				▓	▓	▓				░	░	░						
h26					░	░					▓	▓					░	░						
h31							░	░					▓	▓					░	░				
h32							░	░	░				▓	▓	▓				░	░	░			
h33								░	░	░				▓	▓	▓				░	░	░		
h34									░	░	░				▓	▓	▓				░	░	░	
h35										░	░	░				▓	▓	▓				░	░	░
h36											░	░					▓	▓					░	░
h41													░	░					▓	▓				
h42													░	░	░				▓	▓	▓			
h43														░	░	░				▓	▓	▓		
h44															░	░	░				▓	▓	▓	
h45																░	░	░				▓	▓	▓
h46																	░	░					▓	▓

Problem 10.14

Consider a 2D convolutional layer with kernel size 5×5 that takes 3 input channels and returns 10 output channels. How many convolutional weights are there? How many biases?

With three input channels, a 5×5 kernel would have $5 \times 5 \times 3$ weights for one output. Therefore for 10 outputs, there are $5 \times 5 \times 3 \times 10 = 750$ weights. There is one bias per output channel, giving a total of 10 biases.

Problem 10.15

Draw a weight matrix in the style of figure 10.4d that samples every other variable in a 1D input (i.e., the 1D analog of figure 10.11a). Show that the weight matrix for 1D convolution with kernel size and stride two is equivalent to composing the matrices for 1D convolution with kernel size one and this sampling matrix.

Unsure about this one...

Problem 10.16

Consider the AlexNet network (figure 10.16). How many parameters are used in each convolutional and fully connected layer? What is the total number of parameters?

The layers in the AlexNet are as follows:

- Conv1: 11×11 kernel, 96 output channels, 3 input channels, stride 4. Number of parameters for this layer: $(11 \times 11 \times 3 + 1) \times 96 = 34,944$.
- Conv2 (+ max pooling): 5×5 kernel, 256 output channels, 96 input channels, stride 1. Number of parameters for this layer: $(5 \times 5 \times 96 + 1) \times 256 = 614,656$.
- Conv3: 3×3 kernel, 384 output channels, 256 input channels, stride 1. Number of parameters for this layer: $(3 \times 3 \times 256 + 1) \times 384 = 885,120$.
- Conv4: 3×3 kernel, 384 output channels, 384 input channels, stride 1. Number of parameters for this layer: $(3 \times 3 \times 384 + 1) \times 384 = 1,327,488$.
- Conv5: 3×3 kernel, 256 output channels, 384 input channels, stride 1. Number of parameters for this layer: $(3 \times 3 \times 384 + 1) \times 256 = 884,992$.
- Max pooling and FC layer 1: 4096 hidden units, input dimensions are $13 \times 13 \times 256$. Assuming the maxpooling kernel size is 3×3, the output dimensions are $6 \times 6 \times 256$. Flattened, this gives $6 \times 6 \times 256 = 9216$ input units. Therefore, the parameters for this layer are $(9216 + 1) \times 4096 = 37,752,832$. (+1 accounts for bias)
- FC layer 2: 4096 hidden units. Number of parameters = (input + 1) × output = $(4096 + 1) \times 4096 = 16,781,312$.
- FC layer 3: 1000 hidden units. Number of parameters = (input + 1) × output = $(4096 + 1) \times 1000 = 4,097,000$.
- Output layer: 1000 classes

Therefore, for AlexNet, the total number of parameters is $34,944 + 614,656 + 885,120 + 1,327,488 + 884,992 + 37,752,832 + 16,781,312 + 4,097,000 = 61,788,344$. (assuming a maxpooling kernel size of 3×3 between Conv5 and FC1 layers).

Problem 10.17

What is the receptive field size at each of the first three layers of AlexNet (figure 10.16)?

The first layer (Conv1) has a kernel size of 11×11 and a stride of 4, the receptive field size is therefore 11. Assuming the maxpooling kernel is 3×3 and stride is 1, the receptive field is $11 + (3 - 1) \times 4 = 19$. For Conv2, the receptive field is $19 + (5 - 1) \times 1 = 35$ (as the kernel size for this layer is 5×5). For Conv3, the receptive field is $35 + (3 - 1) \times 4 = 43$.

Problem 10.18

How many weights and biases are there at each convolutional layer and fully connected layer in the VGG architecture (figure 10.17)?

Skipped.

Problem 10.19

Consider two hidden layers of size 224×224 with C1 and C2 channels, respectively, connected by a 3×3 convolutional layer. Describe how to initialize the weights using He initialization.

He initialisation sets the weights of the network from a normal distribution with a mean of 0 and a variance of $\frac{2}{D_h}$, where D_h is the dimension of the original layer to which the weights were applied. This choice of variance in He initialisation ensures that the weights are scaled in such a way as to not diminish or blow up the gradients during forward and backward propagation.

D_h for the first layer is $3 \times 3 \times C_1$, and for the second layer is $3 \times 3 \times C_2$. Therefore, the weights for the convolutional layer are initialised from a normal distribution with a mean of 0 and a variance of $\frac{2}{3 \times 3 \times C_1}$. Similarly, for the second layer, the weights are initialised from a normal distribution with a mean of 0 and a variance of $\frac{2}{3 \times 3 \times C_2}$.