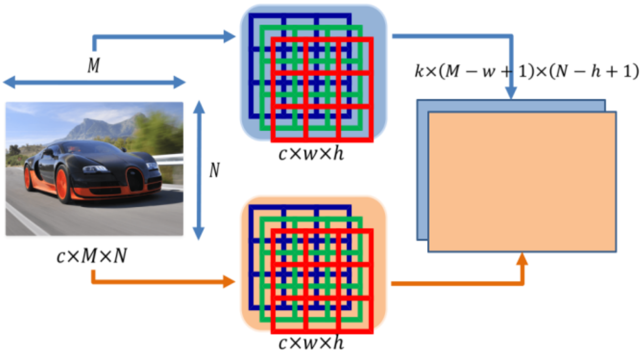
(Solved) Worksheet 2 for AI61002 Spring 2020
Name:
Roll no.:

Calculate the number of parameters and Flops for each layer of AlexNet like network for an input of size $3 \times 227 \times 227$. The network operates in inference mode, i.e., in the software library we set model.eval().

S.No	Layer	#Params	#Flops
1	Conv2d: 64c 11w 4s 2p	23,296	7,30,56,256
2	ReLU	0	4,01,408
3	MaxPool2d: 2w 2s 0p	0	1,50,528
4	Conv2d: 192c 5w 1s 2p	3,07,392	24,09,95,328
5	ReLU	0	3,01,056
6	MaxPool2d: 2w 2s 0p	0	1,12,896
7	Conv2d: 384c 3w 1s 1p	6,63,936	13,01,31,456
8	ReLU	0	1,50,528
9	Conv2d: 256c 3w 1s 1p	8,84,992	17,34,58,432
10	ReLU	0	1,00,352
11	Conv2d: 256c 3w 1s 1p	5,90,080	11,56,55,680
12	ReLU	0	1,00,352
13	MaxPool2d: 4w 2s 0p	0	1,38,240
14	Dropout	0	0
15	Linear: 9216->4096	3,77,52,832	3,77,52,832
16	ReLU	0	8,192
17	Dropout	0	0
18	Linear: 4096->4096	1,67,81,312	1,67,81,312
19	ReLU	0	8,192
20	Linear: 4096->1000	40,97,000	40,97,000
Total		61,100,840	79,34,00,040

2D Convolution Layer (conv) in a network



Width of input tensor = M
Height of input tensor = N
Number of Channels in input tensor = c

Width of 2D convolution kernel = w
Height of 2D convolution kernel = h
Channels of 2D convolution kernel = c

Padding on width of input tensor = p_w
Padding on height of input tensor = p_h

Stride along width of input tensor = s_w

Stride along height of input tensor = s_h

Number of 2D convolution kernels = k

Width of output tensor = $\frac{M-w+2p_w}{s_w}+1$

Height of output tensor = $\frac{N-h+2p_h}{s_h}+1$

Number of channels in output tensor = k
Number of multiplications per output channel per location = cwh
Number of additions per output channel per location (with bias) = cwh
Number of ops. using Multiply-Accumulate (MAC) blocks per channel per location = $cwh+1$
Number of operations per output channel = $(cwh+1)\left(\frac{M-w+2p_w}{s_w}+1\right)\left(\frac{N-h+2p_h}{s_h}+1\right)$
Total number of operations in layer (with bias) = $k(cwh+1)\left(\frac{M-w+2p_w}{s_w}+1\right)\left(\frac{N-h+2p_h}{s_h}+1\right)$

ReLU Transfer Function
Width of input (output) tensor = M
Height of input (output) tensor = N
Channels in input (output) tensor = c

Total number of comparison operations = cMN
Total number of assignment operations = cMN
Total number of operations = $2cMN$

Fully-connected (Linear) Layer
Number of input nodes = n
Number of output nodes = k
Number of multiplication per output node = n
Number of additions per output node (with bias) = $n+1$
Number of ops. using MAC blocks per output node (with bias) = $n+1$
Total number of operations = $(n+1)k$

2D Max-pooling Layer
Width of input (output) tensor = M
Height of input (output) tensor = N
Channels in input (output) tensor = c
Width of 2D pooling kernel = w
Height of 2D pooling kernel = h
Total number of comparison operations per location = $wh-1$
Total number of operations in layer = $c(wh-1)\left(\frac{M-w+2p_w}{s_w}+1\right)\left(\frac{N-h+2p_h}{s_h}+1\right)$

Worked out detailed solution

S. No. 1: (Input) $3 \times 227 \times 227$→(Conv2d)$64c\ 11w\ 4s\ 2p$→(Output)$64 \times 56 \times 56$

$$M = N = 227, c = 3, w = h = 11, k = 64, p_w = p_h = 2, s_w = s_h = 4$$

\# Params = $(cwh + 1)k = \big((3 \times 11 \times 11) + 1\big) \times 64 = \mathbf{23,296}$

\# Flops = $k(cwh + 1)\left(\frac{M-w+2p_w}{s_w} + 1\right)\left(\frac{N-h+2p_h}{s_h} + 1\right)$

$$= 64 \times \big((3 \times 11 \times 11) + 1\big)\left(\frac{227 - 11 + 2(2)}{4} + 1\right)\left(\frac{227 - 11 + 2(2)}{4} + 1\right) = \mathbf{7,30,56,256}$$

S. No. 2: (Input) $64 \times 56 \times 56$→(ReLU)→(Output) $64 \times 56 \times 56$

$$M = N = 56, c = 64$$

\# Params = $\mathbf{0}$

\# Flops = $2cMN = 2 \times 64 \times 56 \times 56 = \mathbf{4,01,408}$

S. No. 3: (Input) $64 \times 56 \times 56$ → (MaxPool2d) $2w\ 2s\ 0p$ → (Output) $64 \times 28 \times 28$

$$M = N = 56, c = 64, w = h = 2, p_w = p_h = 0, s_w = s_h = 2$$

\# Params = $\mathbf{0}$

\# Flops = $c(wh - 1)\left(\frac{M-w+2p_w}{s_w} + 1\right)\left(\frac{N-h+2p_h}{s_h} + 1\right)$

$$= 64 \times \big((2 \times 2) - 1\big)\left(\frac{56 - 2 + 2(0)}{2} + 1\right)\left(\frac{56 - 2 + 2(0)}{2} + 1\right) = \mathbf{1,50,528}$$

S. No. 4: (Input) $64 \times 28 \times 28$ → (Conv2d)$192c\ 5w\ 1s\ 2p$→(Output) 192$\times\ 28 \times 28$

$$M = N = 28, c = 64, w = h = 5, k = 192, p_w = p_h = 2, s_w = s_h = 1$$

\# Params = $(cwh + 1)k = \big((64 \times 5 \times 5) + 1\big) \times 192 = \mathbf{3,07,392}$

\# Flops = $k(cwh + 1)\left(\frac{M-w+2p_w}{s_w} + 1\right)\left(\frac{N-h+2p_h}{s_h} + 1\right)$

$$= 192((64 \times 5 \times 5) + 1)\left(\frac{28 - 5 + 2(2)}{1} + 1\right)\left(\frac{28 - 5 + 2(2)}{1} + 1\right) = \mathbf{24,09,95,328}$$

S. No. 5: (Input) 192 × 28 × 28 → (ReLU) → (Output) 192 × 28 × 28

$$M = N = 28, c = 192$$

\# Params = $\mathbf{0}$

\# Flops = $2cMN = 2 \times 192 \times 28 \times 28 = \mathbf{3,01,056}$

S. No. 6: (Input) 192 × 28 × 28 → (MaxPool2d) $2w\ 2s\ 0p$ → (Output) 192$\times\ 14 \times 14$

$$M = N = 28, c = 192, w = h = 2, p_w = p_h = 0, s_w = s_h = 2$$

\# Params = $\mathbf{0}$

\# Flops = $c(wh - 1)\left(\frac{M-w+2p_w}{s_w} + 1\right)\left(\frac{N-h+2p_h}{s_h} + 1\right)$

$$= 192 \times \big((2 \times 2) - 1\big)\left(\frac{28 - 2 + 2(0)}{2} + 1\right)\left(\frac{28 - 2 + 2(0)}{2} + 1\right) = \mathbf{1,12,896}$$

S. No. 7: (Input) 192$\times\ 14 \times 14$ → (Conv2d) 384$c\ 3w\ 1s\ 1p$ → (Output) 384$\times\ 14 \times 14$

$$M = N = 14, c = 192, w = h = 3, k = 384, p_w = p_h = 1, s_w = s_h = 1$$

\# Params = $(cwh + 1)k = \big((192 \times 3 \times 3) + 1\big) \times 384 = \mathbf{6,63,936}$

Flops = $k(cwh+1)\left(\frac{M-w+2p_w}{s_w}+1\right)\left(\frac{N-h+2p_h}{s_h}+1\right)$

$$= 384((192\times 3\times 3)+1)\left(\frac{14-3+2(1)}{1}+1\right)\left(\frac{14-3+2(1)}{1}+1\right) = \mathbf{13,01,31,456}$$

S. No. 8: (Input) $384\times 14\times 14 \rightarrow$ (ReLU) $\rightarrow$ (Output) $384\times 14\times 14$

$$M = N = 14, c = 384$$

Params = **0**
Flops = $2cMN = 2\times 384\times 14\times 14 = \mathbf{1,50,528}$

S. No. 9: (Input) $384\times 14\times 14 \rightarrow$ (Conv2d) 256 $3w\ 1s\ 1p \rightarrow$ (Output) $256\times 14\times 14$

$$M = N = 14, c = 384, w = h = 3, k = 256, p_w = p_h = 1, s_w = s_h = 1$$

Params = $(cwh+1)k = \big((384\times 3\times 3)+1\big)\times 256 = \mathbf{8,84,992}$
Flops = $k(cwh+1)\left(\frac{M-w+2p_w}{s_w}+1\right)\left(\frac{N-h+2p_h}{s_h}+1\right)$

$$= 256((384\times 3\times 3)+1)\left(\frac{14-3+2(1)}{1}+1\right)\left(\frac{14-3+2(1)}{1}+1\right) = \mathbf{17,34,58,432}$$

S. No. 10: (Input) $256\times 14\times 14 \rightarrow$ (ReLU) $\rightarrow$ (Output) $256\times 14\times 14$

$$M = N = 14, c = 256$$

Params = **0**
Flops = $2cMN = 2\times 256\times 14\times 14 = \mathbf{1,00,352}$

S. No. 11: (Input) $256\times 14\times 14 \rightarrow$ (Conv2d) 256 $3w\ 1s\ 1p \rightarrow$ (Output) $256\times 14\times 14$

$$M = N = 14, c = 256, w = h = 3, k = 256, p_w = p_h = 1, s_w = s_h = 1$$

Params = $(cwh+1)k = \big((256\times 3\times 3)+1\big)\times 256 = \mathbf{5,90,080}$
Flops = $k(cwh+1)\left(\frac{M-w+2p_w}{s_w}+1\right)\left(\frac{N-h+2p_h}{s_h}+1\right)$

$$= 256((256\times 3\times 3)+1)\left(\frac{14-3+2(1)}{1}+1\right)\left(\frac{14-3+2(1)}{1}+1\right) = \mathbf{11,56,55,680}$$

S. No. 12: (Input) $256\times 14\times 14 \rightarrow$ (ReLU) $\rightarrow$ (Output) $256\times 14\times 14$

$$M = N = 14, c = 256$$

Params = **0**
Flops = $2cMN = 2\times 256\times 14\times 14 = \mathbf{1,00,352}$

S. No. 13: (Input) $256\times 14\times 14 \rightarrow$ (MaxPool2d) $4w\ 2s\ 0p \rightarrow$ (Output) $256\times 6\times 6$

$$M = N = 14, c = 256, w = h = 4, p_w = p_h = 0, s_w = s_h = 2$$

Params = **0**
Flops = $c(wh-1)\left(\frac{M-w+2p_w}{s_w}+1\right)\left(\frac{N-h+2p_h}{s_h}+1\right)$

$$= 256\times\big((4\times 4)-1\big)\left(\frac{14-4+2(0)}{2}+1\right)\left(\frac{14-4+2(0)}{2}+1\right) = \mathbf{1,38,240}$$

S. No. 14: (Input) $256\times 6\times 6 \rightarrow$ Dropout $\rightarrow$ (Output) $256\times 6\times 6$
Params = **0**
Flops = **0** (model in evaluation mode)

S. No. 15: (Input) $9216 \rightarrow$ Linear $\rightarrow 4096$

$$n = 9216, k = 4096$$

Params = $(n + 1) \times k = (9216 + 1) \times 4096 = \mathbf{3,77,52,832}$

Flops = $(n + 1) \times k = (9216 + 1) \times 4096 = \mathbf{3,77,52,832}$

S. No. 16: (Input) $4096 \rightarrow$ (ReLU) $\rightarrow$ (Output) 4096

$$n = 4096$$

Params = $\mathbf{0}$

Flops = $2n = 2 \times 4096 = \mathbf{8,192}$

S. No. 17: (Input) $4096 \rightarrow$ Dropout $\rightarrow$ (Output) 4096

Params = $\mathbf{0}$

Flops = **0** (model in evaluation mode)

S. No. 18: (Input) $4096 \rightarrow$ Linear $\rightarrow$(Output) 4096

$$n = 4096, k = 4096$$

Params = $(n + 1) \times k = (4096 + 1) \times 4096 = \mathbf{1,67,81,312}$

Flops = $(n + 1) \times k = (4096 + 1) \times 4096 = \mathbf{1,67,81,312}$

S. No. 19: (Input) $4096 \rightarrow$ (ReLU) $\rightarrow$ (Output) 4096

$$n = 4096$$

Params = $\mathbf{0}$

Flops = $2n = 2 \times 4096 = \mathbf{8,192}$

S. No. 20: (Input) $4096 \rightarrow$ Linear $\rightarrow$ (Output) 1000

$$n = 4096, k = 1000$$

Params = $(n + 1) \times k = (4096 + 1) \times 1000 = \mathbf{40,97,000}$

Flops = $(n + 1) \times k = (4096 + 1) \times 1000 = \mathbf{40,97,000}$