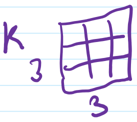
29_CNN_Architectures_1

09 October 2024 12:07

Convolution Layer :-

Accepts a volume of size $\underline{W1} \times \underline{H1} \times \underline{D1}$



K → # of filters

S → stride

P → Amount of padding

F → spacial extent

Produces an output of size $W2 \times H2 \times \underbrace{D2}_{K}$

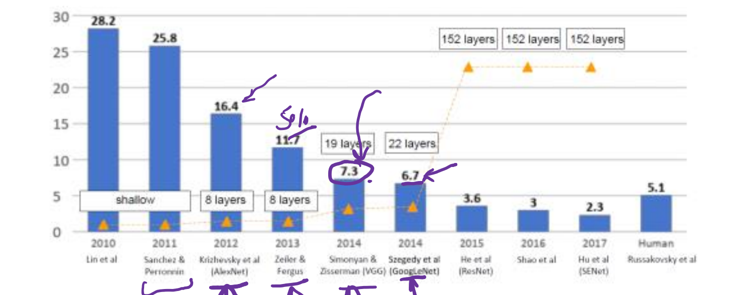
$$W2 = \frac{W1 - F + 2P}{S} + 1$$

$$H2 = \frac{H1 - F + 2P}{S} + 1$$

Weights per filter :-

$F \times F \times D1$

ImageNet Dataset (ILSVRC)



ReLU ↑ 60 M

ZF Net 61.5m parameters

ALEXNET :-

$227 \times 227 \times 3$

1000

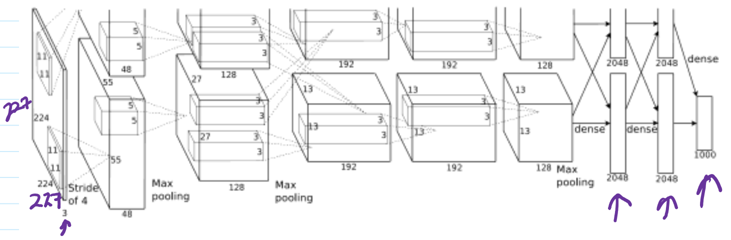


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

$$\frac{227 - 11 + 2 \times 0}{4} + 1$$

(227x227x3) → Input

(55x55x96) → CONV1 → 96 11x11 filters at stride 4, pad 0

(27x27x96) → MAX POOL1 3x3 filters at stride 2

(") → NORM1 ← Not common anymore

(27x27x256) → CONV2 → 256 5x5 filters stride 1, pad 2

(13×13×256) → MAXPOOL 2 → 3×3 stride 2
(") → NORM2
(13×13×384) → CONV3 → 384 3×3 filters, stride 1, pad 1
(13×13×384) → CONV4 → 384 3×3 filters, strid 1, pad 1
(13×13×256) → CONV 5 → 256 3×3 filters stride 1, pad 1
(6×6×256) → MAXPOOL5 → 3×3 filters stride 2

(6×6×256×4096)

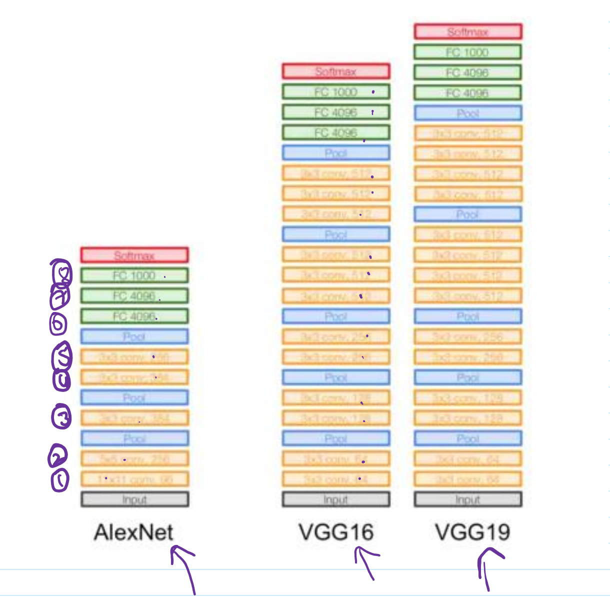
(4096) → FC 6 → 4096 neurons
(4096) → FC 7 → 4096 neurons
(1000) → FC 8 → 1000 neurons ← softmax

L1	L2	L3	L4	L5	L6	L7	L8	
35k	307k	884k	663k.	442k	37M	16M	4M	→ 60M

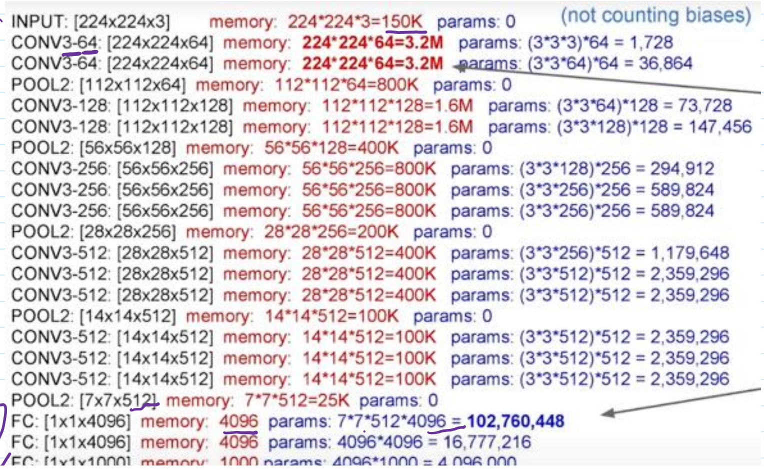
Details:-

1. First use of ReLU
2. Heavy data augmentation
3. Dropout 0.5 ←
4. Batch size 128
5. SGD Momentum 0.9.
6. LR, 1e-2, reduced manually where val. accuracy had plateaus.

VGG Net :- (Visual Geometry Group) → Oxford University



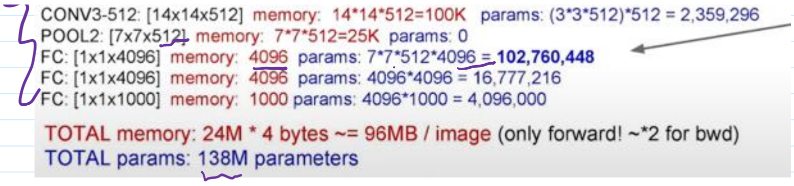
Initial Layers



```
INPUT: [224x224x3]        memory:  224*224*3=150K   params: 0          (not counting biases)
CONV3-64: [224x224x64]  memory:  224*224*64=3.2M   params: (3*3*3)*64 = 1,728
CONV3-64: [224x224x64]  memory:  224*224*64=3.2M   params: (3*3*64)*64 = 36,864
POOL2: [112x112x64]  memory:  112*112*64=800K   params: 0
CONV3-128: [112x112x128]  memory:  112*112*128=1.6M   params: (3*3*64)*128 = 73,728
CONV3-128: [112x112x128]  memory:  112*112*128=1.6M   params: (3*3*128)*128 = 147,456
POOL2: [56x56x128]  memory:  56*56*128=400K   params: 0
CONV3-256: [56x56x256]  memory:  56*56*256=800K   params: (3*3*128)*256 = 294,912
CONV3-256: [56x56x256]  memory:  56*56*256=800K   params: (3*3*256)*256 = 589,824
CONV3-256: [56x56x256]  memory:  56*56*256=800K   params: (3*3*256)*256 = 589,824
POOL2: [28x28x256]  memory:  28*28*256=200K   params: 0
CONV3-512: [28x28x512]  memory:  28*28*512=400K   params: (3*3*256)*512 = 1,179,648
CONV3-512: [28x28x512]  memory:  28*28*512=400K   params: (3*3*512)*512 = 2,359,296
CONV3-512: [28x28x512]  memory:  28*28*512=400K   params: (3*3*512)*512 = 2,359,296
POOL2: [14x14x512]  memory:  14*14*512=100K   params: 0
CONV3-512: [14x14x512]  memory:  14*14*512=100K   params: (3*3*512)*512 = 2,359,296
CONV3-512: [14x14x512]  memory:  14*14*512=100K   params: (3*3*512)*512 = 2,359,296
CONV3-512: [14x14x512]  memory:  14*14*512=100K   params: (3*3*512)*512 = 2,359,296
POOL2: [7x7x512]  memory:  7*7*512=25K  params: 0
FC: [1x1x4096]  memory:  4096  params: 7*7*512*4096 = 102,760,448
FC: [1x1x4096]  memory:  4096  params: 4096*4096 = 16,777,216
FC: [1x1x1000]  memory:  1000 params: 4096*1000 = 4,096,000
```

Later

Later



CONV3-512: [14x14x512] memory: 14*14*512=100K params: (3*3*512)*512 = 2,359,296
POOL2: [7x7x512] memory: 7*7*512=25K params: 0
FC: [1x1x4096] memory: 4096 params: 7*7*512*4096 = **102,760,448**
FC: [1x1x4096] memory: 4096 params: 4096*4096 = 16,777,216
FC: [1x1x1000] memory: 1000 params: 4096*1000 = 4,096,000

TOTAL memory: 24M * 4 bytes ~= 96MB / image (only forward! ~*2 for bwd)
TOTAL params: 138M parameters