Session-6

CNN For Medical Diagnosis

March 22, 2024

Agenda

1. Recap
2. Business - Case - Pneumonia
3. Mobile Net
4. Grad Cam
5. Efficient Net

Why GAP makes Sense.

conv2d (Conv2D)	(None, 224, 224, 32)	896
conv2d_1 (Conv2D)	(None, 224, 224, 32)	9248
max_pooling2d (MaxPooling2D)	(None, 112, 112, 32)	0
conv2d_2 (Conv2D)	(None, 112, 112, 64)	18496
conv2d_3 (Conv2D)	(None, 112, 112, 64)	36928
max_pooling2d_1 (MaxPooling2D)	(None, 56, 56, 64)	0
conv2d_4 (Conv2D)	(None, 56, 56, 128)	73856
conv2d_5 (Conv2D)	(None, 56, 56, 128)	147584
max_pooling2d_2 (MaxPooling2D)	(None, 28, 28, 128)	0
conv2d_6 (Conv2D)	(None, 28, 28, 256)	295168
conv2d_7 (Conv2D)	(None, 28, 28, 256)	590080
max_pooling2d_3 (MaxPooling2D)	(None, 14, 14, 256)	0
conv2d_8 (Conv2D)	(None, 14, 14, 512)	1180160
conv2d_9 (Conv2D)	(None, 14, 14, 512)	2359808
max_pooling2d_4 (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 1024)	25691136
dense_1 (Dense)	(None, 1024)	1049600
dense_2 (Dense)	(None, 10)	10250

Total params: 31463210 (120.02 MB)
Trainable params: 31463210 (120.02 MB)
Non-trainable params: 0 (0.00 Byte)

512 → containing a specific kind of inf

$224 \times 224 \times 3$

$\odot$

$(3 \times 3 \times 3) \times 32$

↓

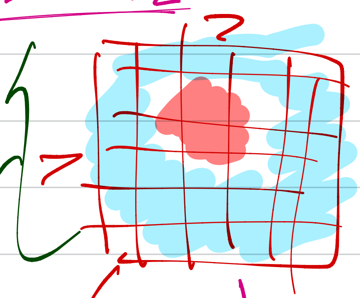
$224 \times 224 \times 32$

$\odot$

$(3 \times 3 \times 32) \times 32$

↓

$224 \times 224 \times 32$



Avg

GAP

Dog kernel chan

[[- - -]

4

Feat is present or not

≠ 0, 2, 3/

[0, 2.7, 5.8, 0, 0]

MOBILE - NET

$32 \times 32 \times 3$ ⊙ $3 \times 3 \times 3$

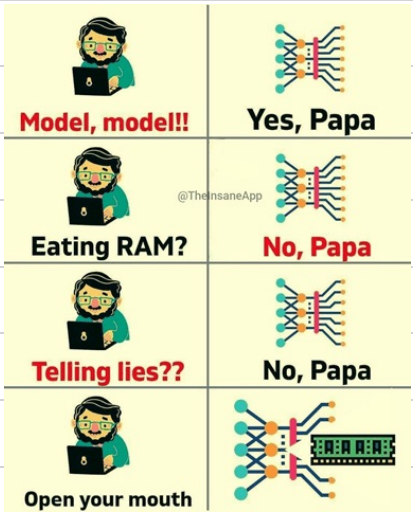
size of channel

no of channel

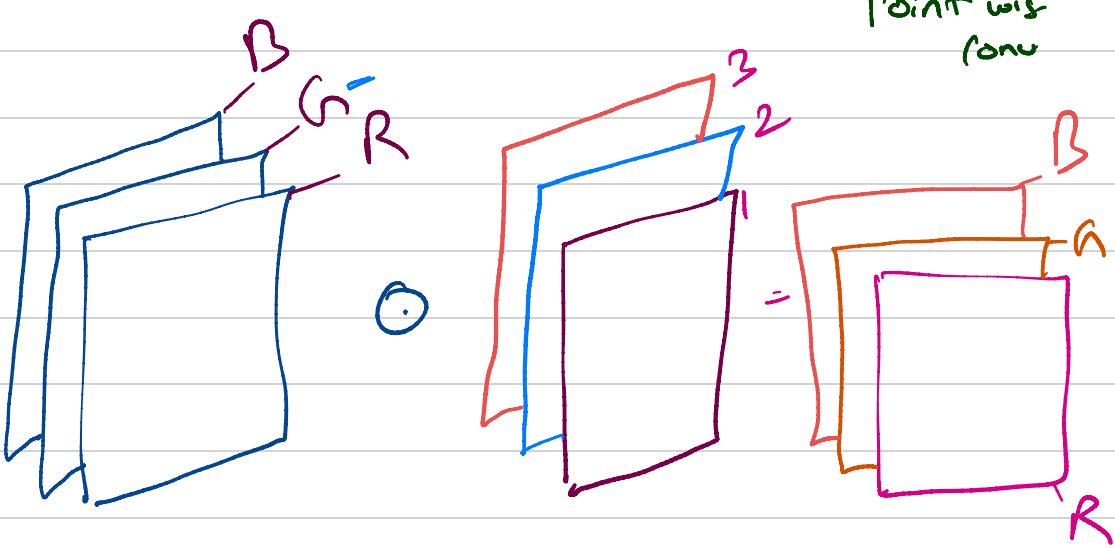
↳ $30 \times 30 \times 3$

↳ Element wise (+)

↳ 30×30



Mobile net uses Depthwise conv + Point wise conv

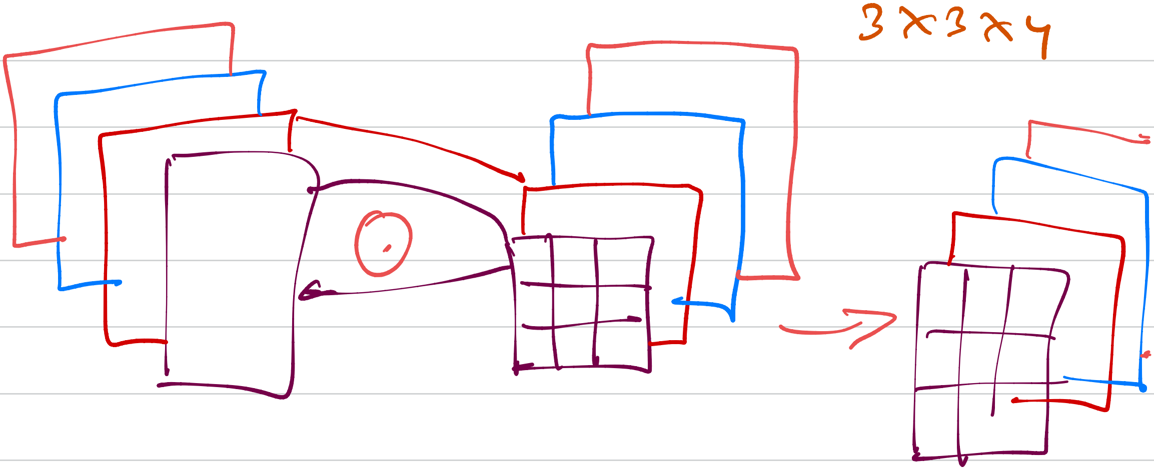


$224 \times 224 \times 32 \odot (3 \times 3 \times 1) \times 32$

$\rightarrow 222 \times 222 \times 32$

$5 \times 5 \times 4 \odot (3 \times 3 \times 1) \times 4$

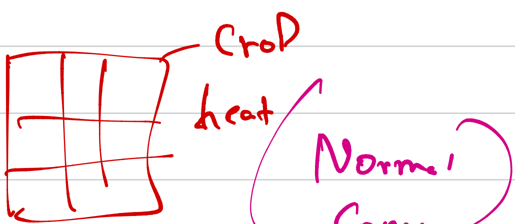
$3 \times 3 \times 4$



$14 \times 14 \times 124 \odot (3 \times 3 \times 1) \times 124$

$\rightarrow 12 \times 12 \times 124$

R, G, B, In, SWIR



MobileNet Body Architecture

Type / Stride	Filter Shape	Input Size
Conv / s2	3 × 3 × 3 × 32	224 × 224 × 3
Conv dw / s1	3 × 3 × 32 dw	112 × 112 × 32
Conv / s1	1 × 1 × 32 × 64	112 × 112 × 32
Conv dw / s2	3 × 3 × 64 dw	112 × 112 × 64
Conv / s1	1 × 1 × 64 × 128	56 × 56 × 64
Conv dw / s1	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 128	56 × 56 × 128
Conv dw / s2	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 256	28 × 28 × 128
Conv dw / s1	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 256	28 × 28 × 256
Conv dw / s2	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 512	14 × 14 × 256
5× Conv dw / s1	3 × 3 × 512 dw	14 × 14 × 512
5× Conv / s1	1 × 1 × 512 × 512	14 × 14 × 512
Conv dw / s2	3 × 3 × 512 dw	14 × 14 × 512
Conv / s1	1 × 1 × 512 × 1024	7 × 7 × 512
Conv dw / s2	3 × 3 × 1024 dw	7 × 7 × 1024
Conv / s1	1 × 1 × 1024 × 1024	7 × 7 × 1024
Avg Pool / s1	Pool 7 × 7	7 × 7 × 1024
FC / s1	1024 × 1000	1 × 1 × 1024
Softmax / s1	Classifier	1 × 1 × 1000

S=2 (Sampling)

normal Conv (3×3×32)×32

112×112×32 ⊙ (3×3×1)×32

→ 112×112×32

⊙ normal Conv (1×1×32)×64

Each Kern = 112×112×64

112×112×64

⊙ (1×1×32)×32/64

→ 112×112×32

↓

elem > 112×112×32/64

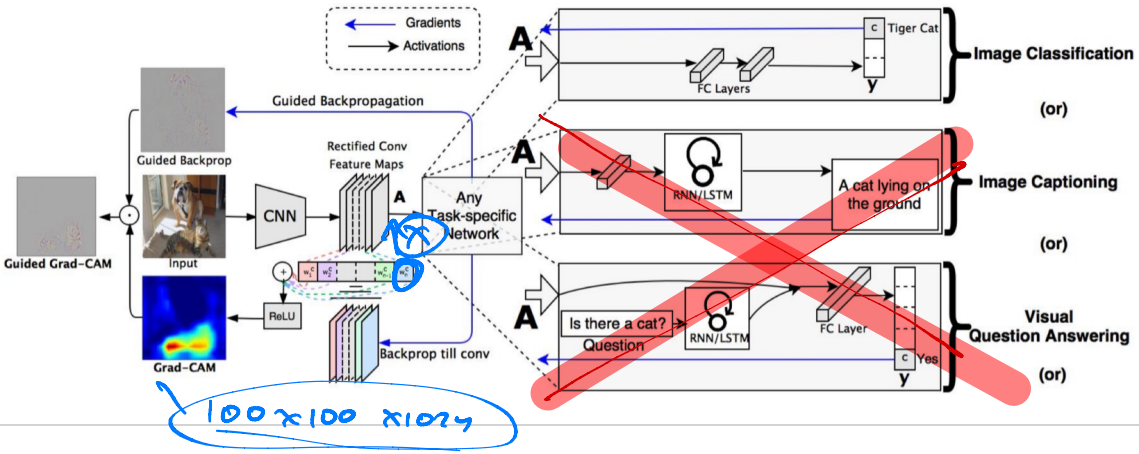
5 x 5 x 3  (3 x 3 x 3) x 32 → 864

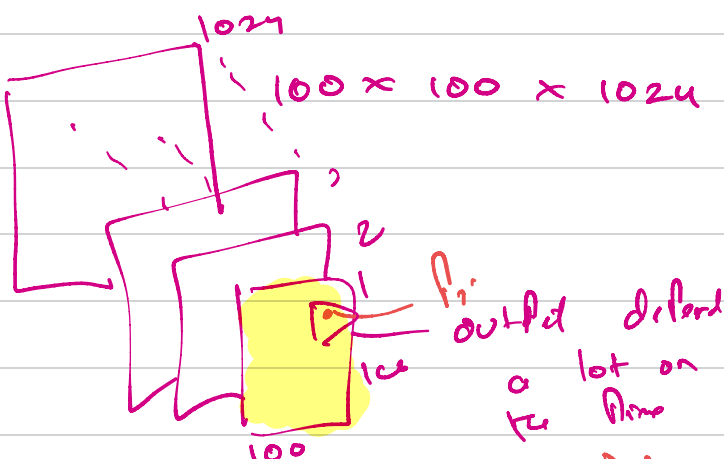
norm

depth

(3 x 3 x 1) x 3 → 27

GRADCAM


Gradients
Activations
Guided Backpropagation
Guided Backprop
Guided Grad-CAM
Input
CNN
Rectified Conv Feature Maps
A
Any Task-specific Network
ReLU
Grad-CAM
Backprop till conv
100x100 x1024
Tiger Cat
FC Layers
y
Image Classification
(or)
RNN/LSTM
A cat lying on the ground
Image Captioning
(or)
Is there a cat?
Question
FC Layer
Yes
Visual Question Answering
(or)


1024
100 x 100 x 1024
GAP
Dense (softmax)
Tiger
2
1
100
f_i
output depends a lot on f_i here
$\frac{\partial L}{\partial f_i}$
what does it mean if this gradient is high?
gradient map
avg of this
imp of channel

1. Extract layer before GAP → 50 x 50 x 512
2. Compute gradient matrix wrt. loss for all channels of step 1
3. Compute average (GAP) → 1 x 512 → 50 x 50 x 512
4. Multiply (1, 512) with step 1 matrix

↓

50 x 50 x 512

↓

RELU

↓

Output → Scaled → Resize + Original image → Output

Efficient Net