Session-3

CNN UNDER THE HOOD

March 15, 2024

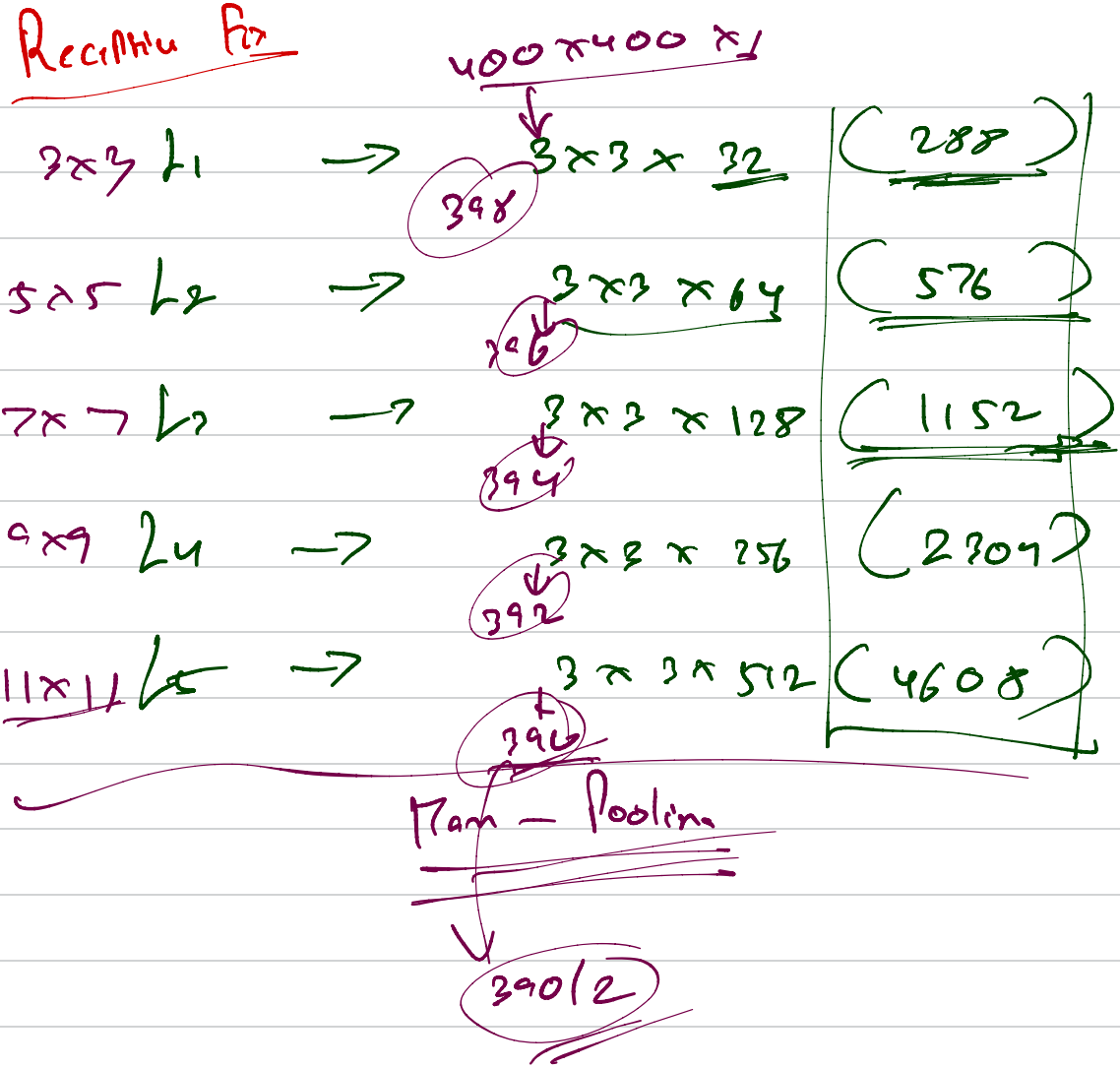
Me using Image Augmentation to increase the training data

*Flips image horizontally *

Deep learning model:

AGENDA

1. Intuition behind multi-channel Convolution
2. Global Average Pooling
3. One x One Convolution
4. Dilated Convolution
5. Image Augmentation.


Receptive Field
400 x 400 x 1
3x3 L1
3x3x32
288
398
5x5 L2
3x3x64
576
196
7x7 L3
3x3x128
1152
394
9x9 L4
3x3x256
2304
392
11x11 L5
3x3x512
4608
396
Max - Pooling
3901 2

Intuition - Multi Channel Conv

Case-I Multi spectral Imaging in Agriculture.

→ Capture Image data at specific wavelength across Electromagnetic spectrum.

→ Eg. Following spectrum

→ RGB spectrum

* For visual Information what humans see

→ Infrared Spectrum

* Healthy plants reflect more near infrared (NIR) light

NDVI (Normalized Diff- Vegetation Index)

→ Shortwave Infrared (SWIR)

* For water content in soil.
* Water absorbs SWIR signals strongly so it appears darker.

Hawks → Zone Plan

← Hardware

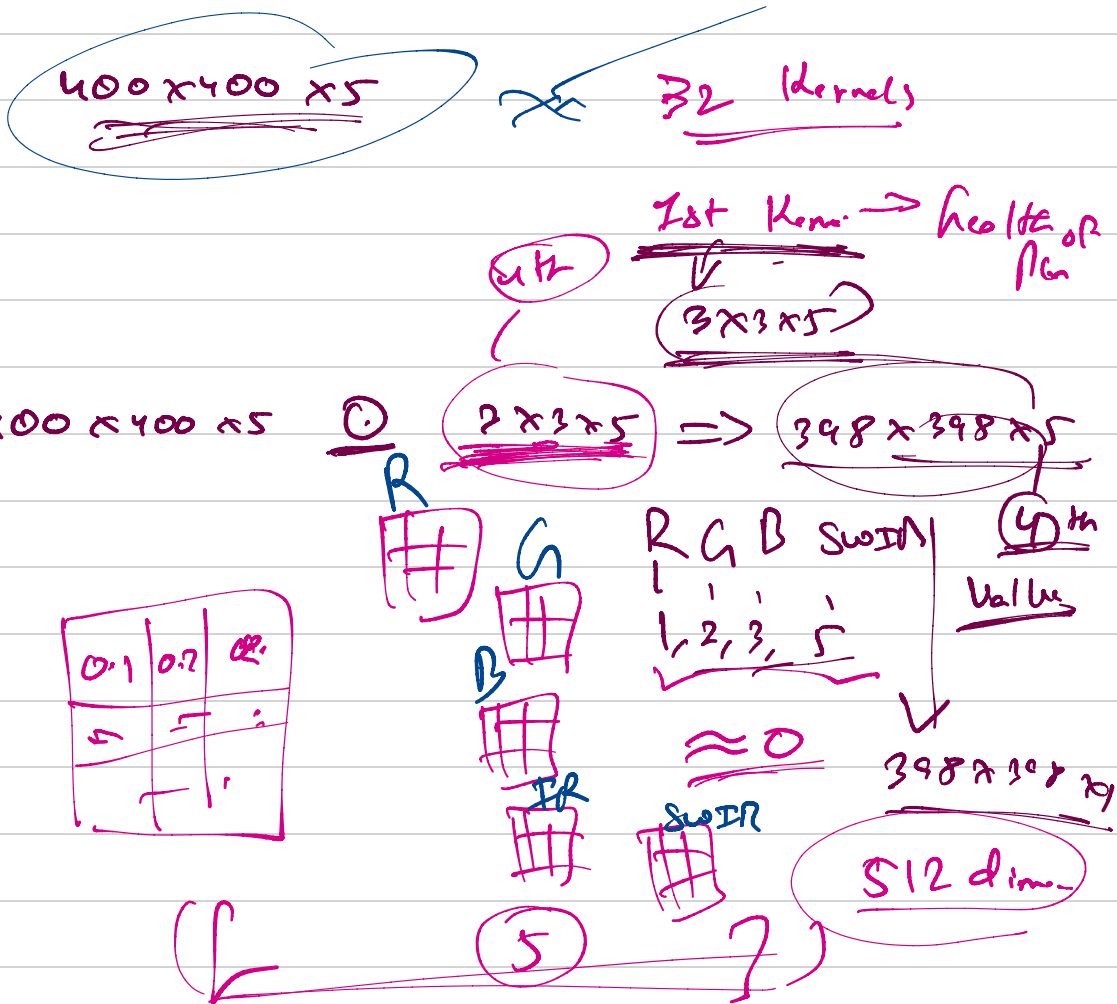
→ For what Kernels might look For.

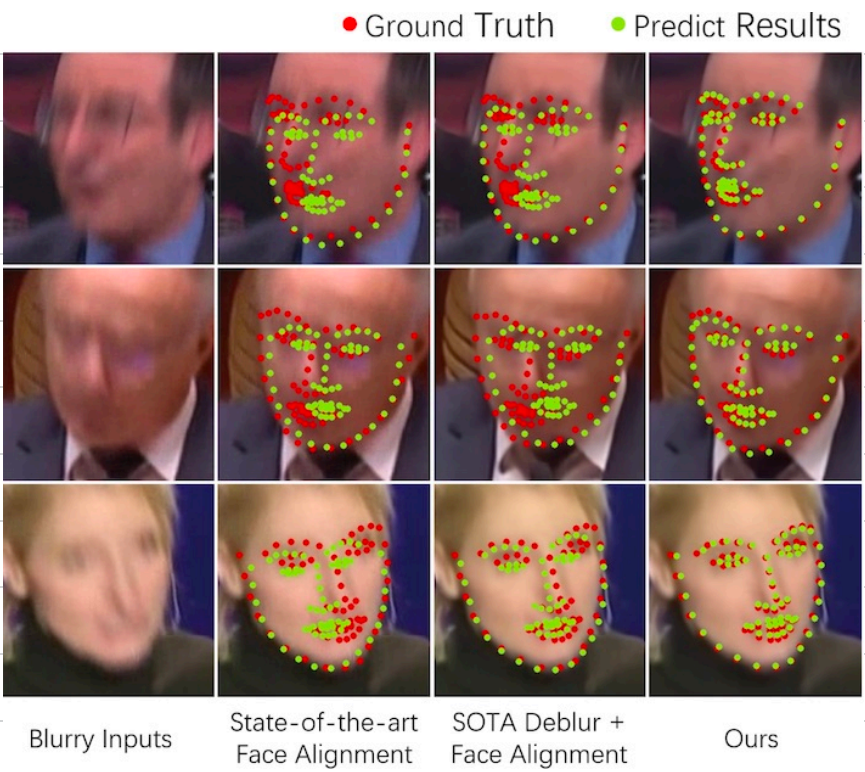
(i) Signs of Stress

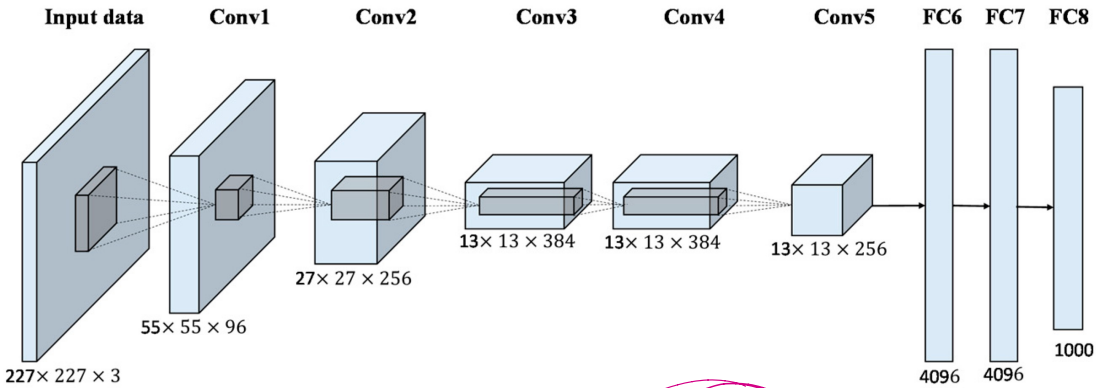
(ii) Optimizing water use

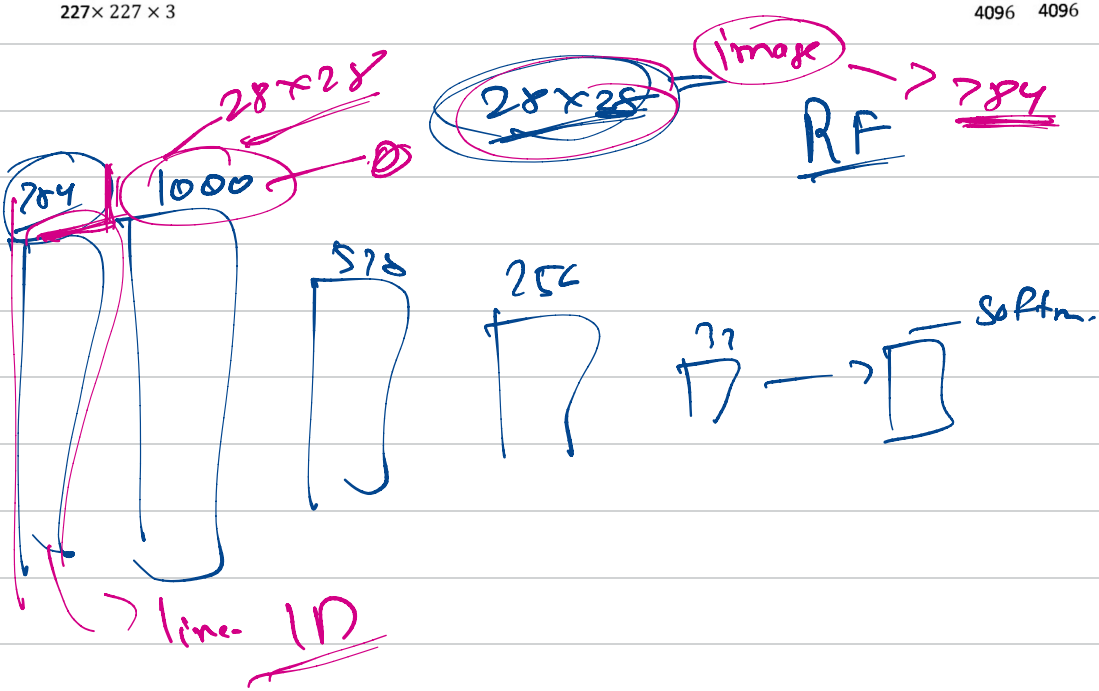
(iii) Crop health

(iv) Yield prediction.




Ground Truth
Predict Results
Blurry Inputs
State-of-the-art Face Alignment
SOTA Deblur + Face Alignment
Ours


Input data
Conv1
Conv2
Conv3
Conv4
Conv5
FC6
FC7
FC8
227× 227 × 3
55× 55 × 96
27× 27 × 256
13× 13 × 384
13× 13 × 384
13× 13 × 256
4096
4096
1000


28×28
28×28
image
784
RF
784
1000
578
256
Softm.
line ID

Global Average Pooling

Model: "sequential"

224 x 224 x 3

Layer (type)	Output Shape	Param #
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)

7x7x512

+ (ons)

25088 x 1024

Dog Vs Cat

1st → Dog no

2nd → Cat's no

Dog's Pic

7 × 7 × 512 (Dog's image)

→ Cat's nose

For Cat's Features Kernel ?! → 0

7 × 7 × 512 → 1 × 1 × 512

→ 0

Dog's nose → Dog → 0

[5, 6, 78, 0.1, 0.2, 0.01]

512

↓

Dense(10, softmax)

No ← softmax feature

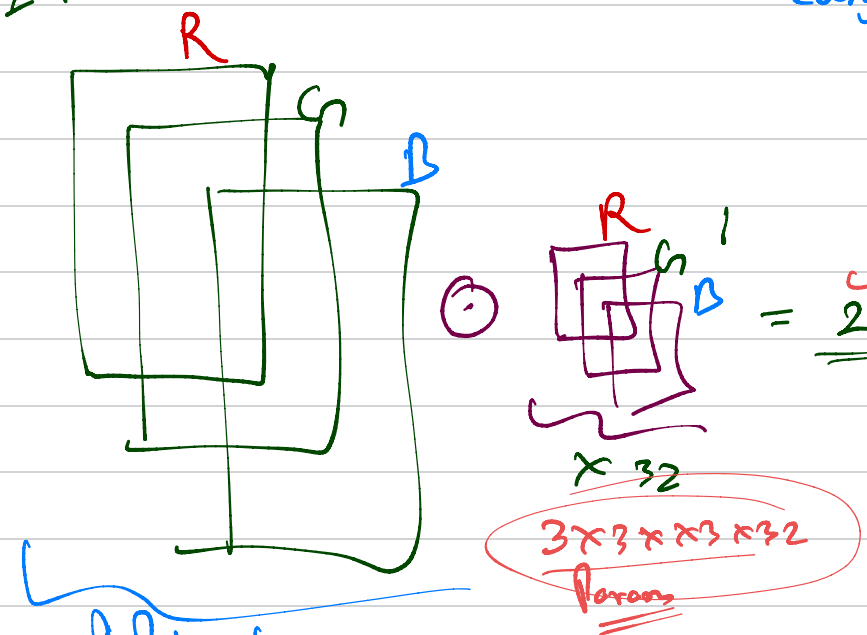
[a, b, c, ...,] 512

H W # of Channels

Image = 224, 224, 3

Every Ch. -> Channel

2 im.



400 x 400 x 3 (3x3x3) x32

↓

398 x 398 x 32

32, 32

↓

4, 6 / Y, B

One x One Convolution

400x400x3 I (3x3x3)x32 I 398x398x32 RF of 3x3

398x398x32 I (3x3x32)x64 I 396x396x64 RF of 5X5

396x396x64 I (3x3x64)x128 I 394x394x128 RF of 7X7

394x394x128 I (3x3x128)x256 I 392x392x256 RF of 9X9

392x392x256 I (3x3x256)x512 I 390x390x512 RF of 11X11

MaxPooling

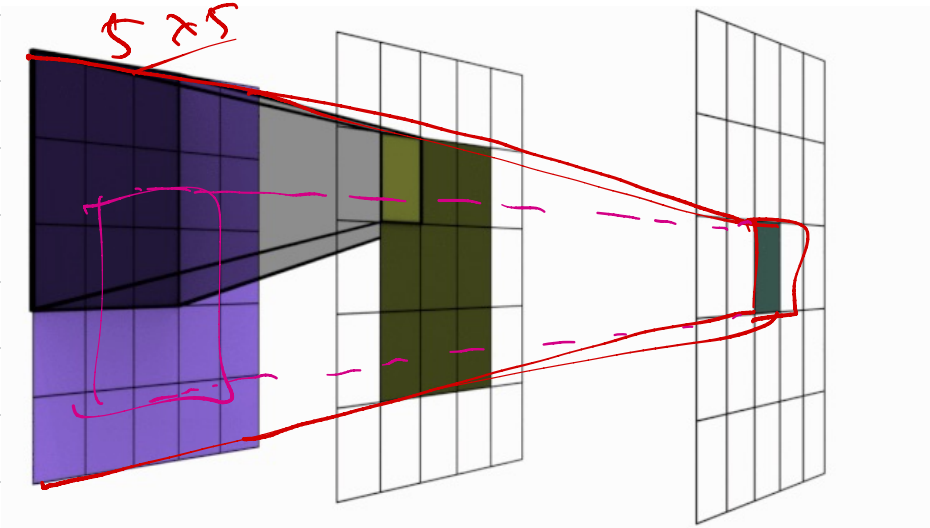
195x195x512 I (?x?x512)x32 I ?x?x32 RF of 22x22

.. 3x3x32x64 RF of 24x24

.. 3x3x64x128 RF of 26x26

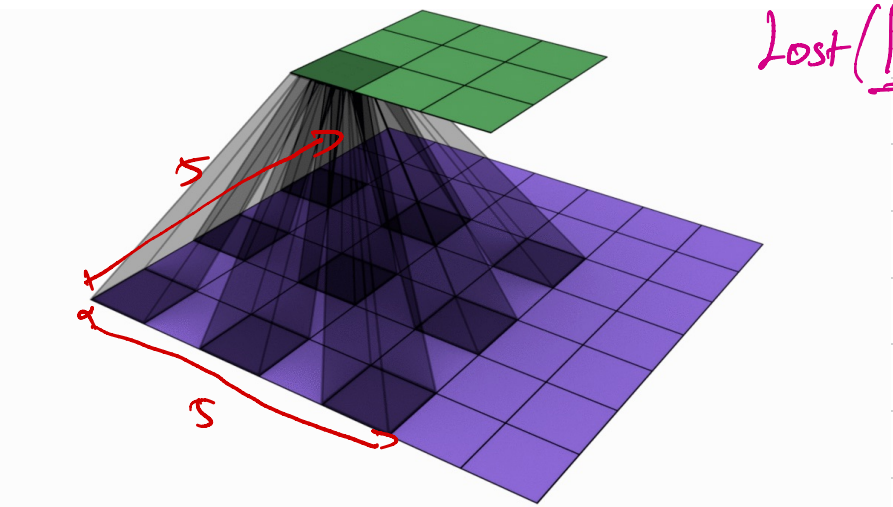
.. 3x3x128x256 RF of 38x28

.. 3x3x256x512 RF of 30x30


5 x 5

Dilated Convolution

Q. How to get RF of 5×5 with one 3×3 Kernel?



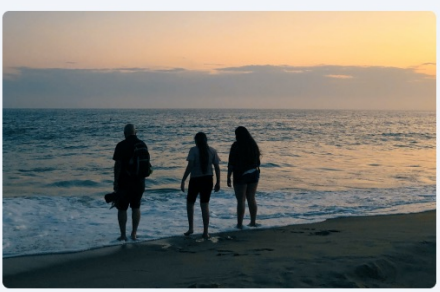
Lost (Filtrin

→ Dilated convolution increases receptive field of the network exponentially and Linear parameter accretion

→ It's used in application where we care more about integrating Knowledge of the broader content with less cost.

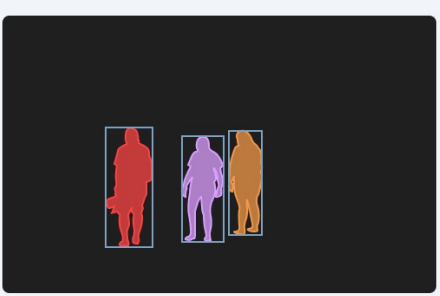
→ Key application the dilated conv authors have in mind is **dense prediction**: vision applications where predicted object has a similar size & structure to input image.

Semantic Segmentation vs. Instance Segmentation vs. Panoptic Segmentation



(a) Image

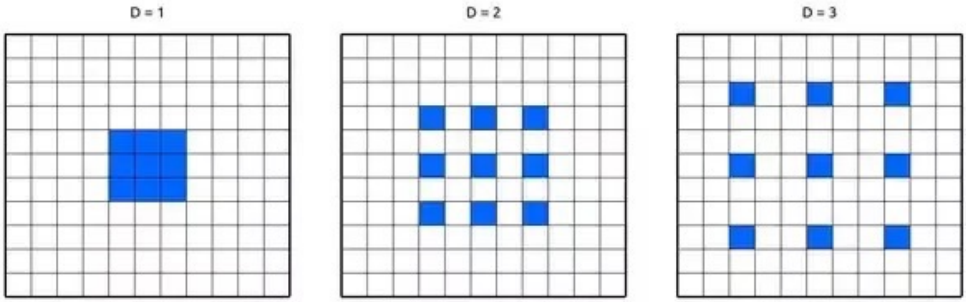
(b) Semantic Segmentation

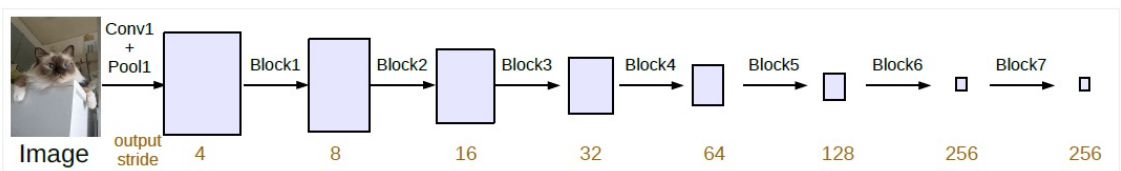


(c) Instance Segmentation

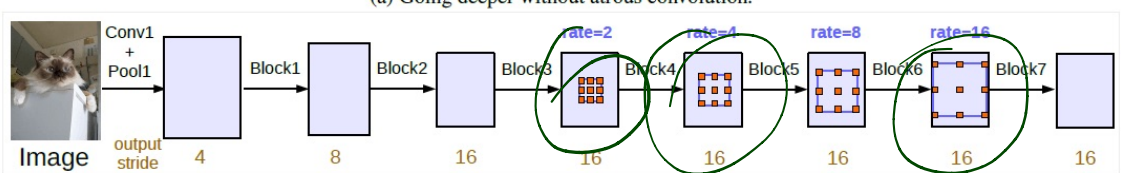
(d) Panoptic Segmentation

V7 Labs





(a) Going deeper without atrous convolution.



(b) Going deeper with atrous convolution. Atrous convolution with $rate > 1$ is applied after block3 when $output\_stride = 16$.

micrasoft. normalcy Telescope

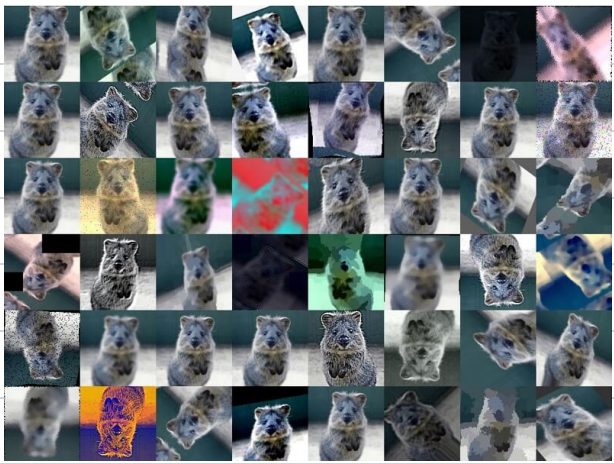
less non-linear combination of feature

In + 2x0

Image Augmentation

Lower Level Augmentations

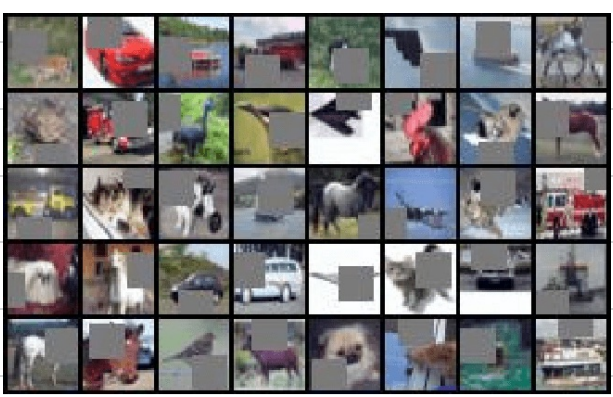
most of it is in built in Keras

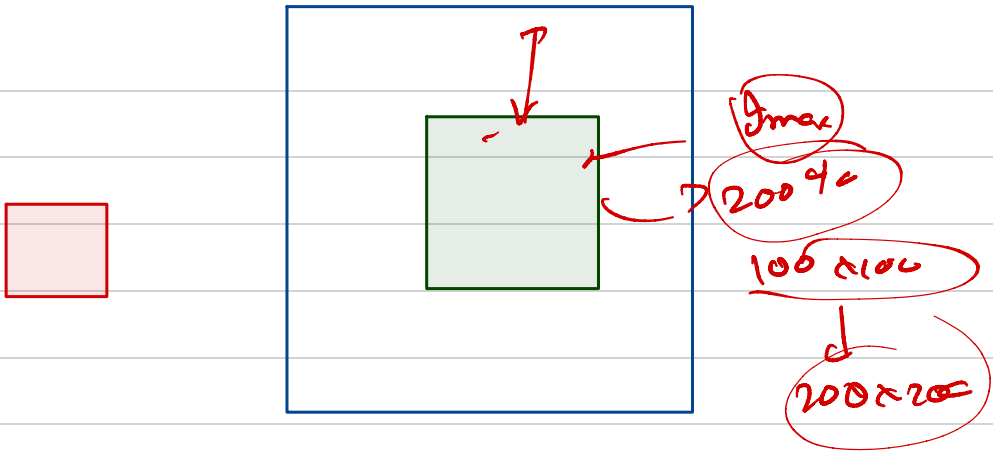


Cut Off — Higher Levels

V.V. Effective.

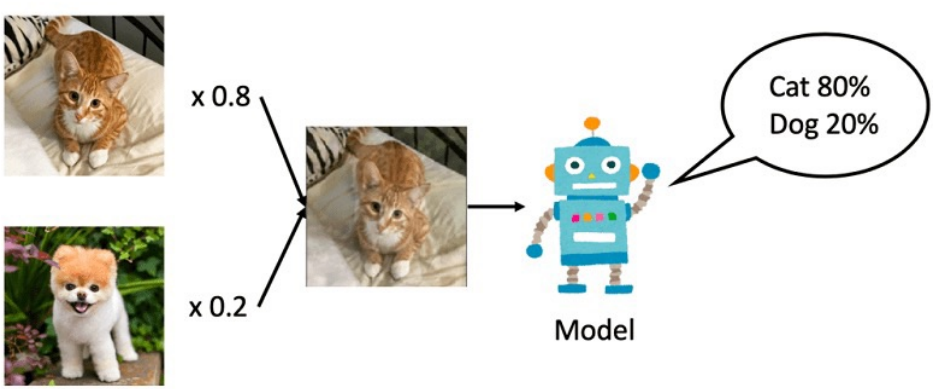
Albumentation





(Scientist with more facial Expressions)

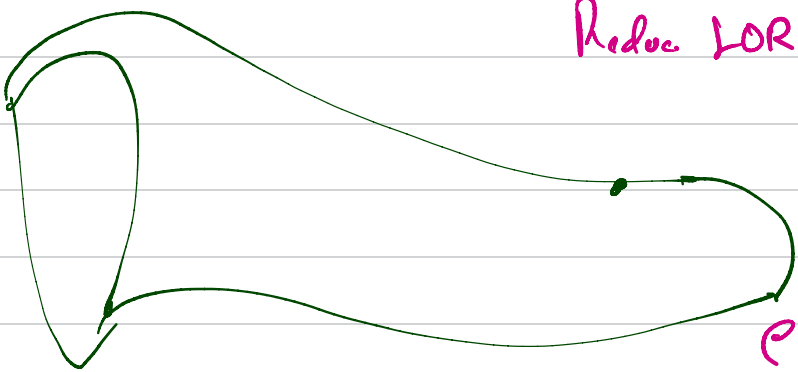
Mix UP – Higher Levels



(custom Implementation needed.

Tough but Effective.

Adam



Reduc LOR on Plat

5 epochs

% val_loss ?!

epoch = 6

LR = LR x 0.3