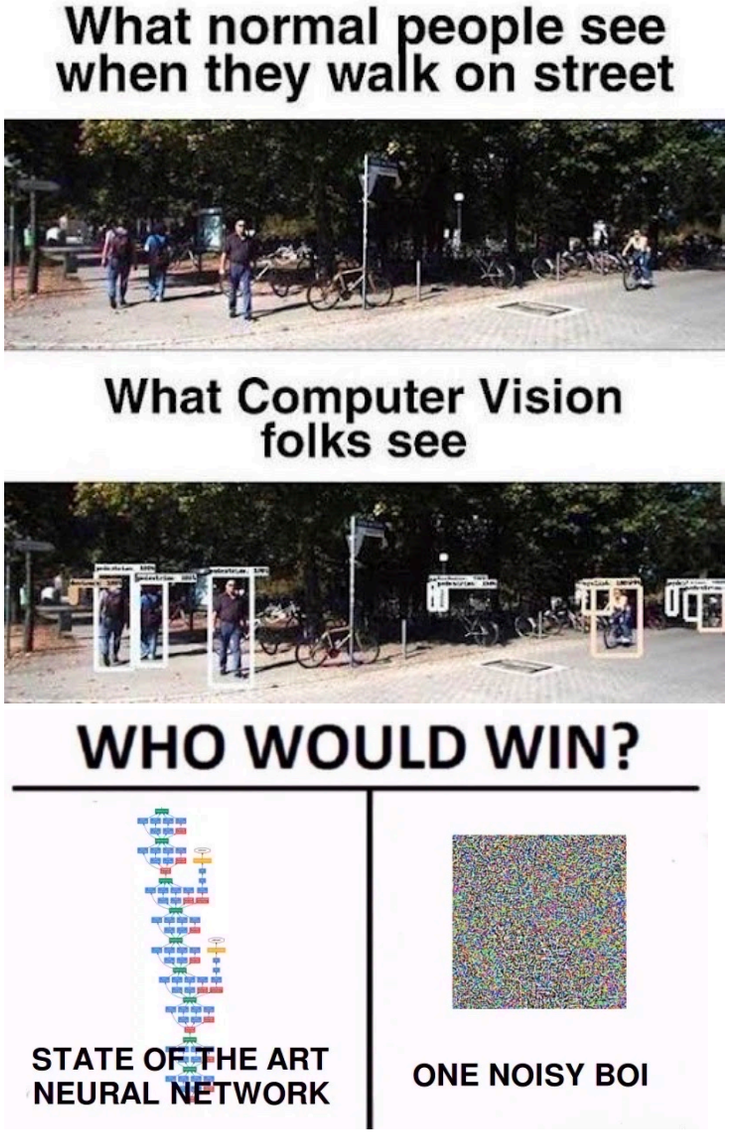
April 5, 2023

DSML: Computer Vision.

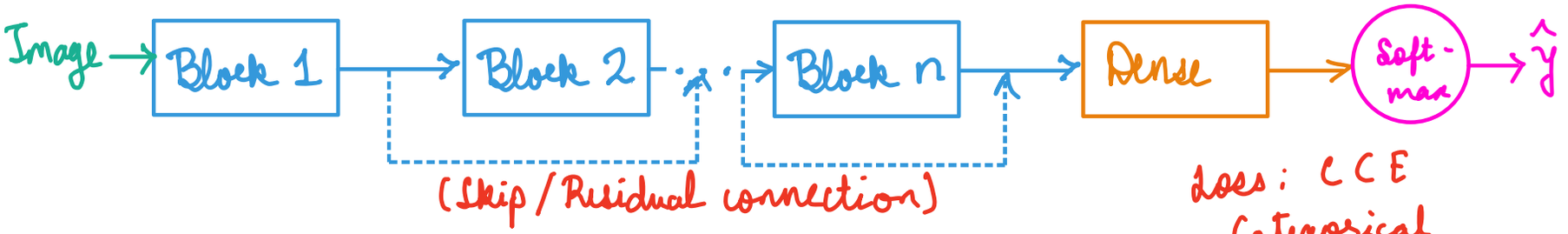
Object Localization and Detection - 1

Class starts @ 9:05 pm.



Recap:

* Basic CNN architecture for Image classification:



loss: CCE
Categorical
Cross-Entropy.

* Popular architectures and their performance:

Model Name	Number of params	Top 1 Acc	Top 5 Acc
EfficientnetB0	5.3M	77.3	93.5
MobileNet	2.3M	71.0	90.5
ResNet50	25.6M	83.2	96.5
Inception	22.9M	79.0	94.5
VGG16	138M	74.4	91.9
AlexNet	62M	63.3	84.6

Novelty

Automated scaling strategy.

→ Factorizing convolutions

Residual / Skip connections.

Inception module, 1×1 conv

3×3 kernels, blocks.

First method to use GPUs, SGD.

Recap:

* Major CNN related concepts:

(a) Transfer learning: Reuse an already trained CNN for image classification.

→ Helpful when we have a small dataset.

→ As long as images are similar to Imagenet, better than training from scratch.

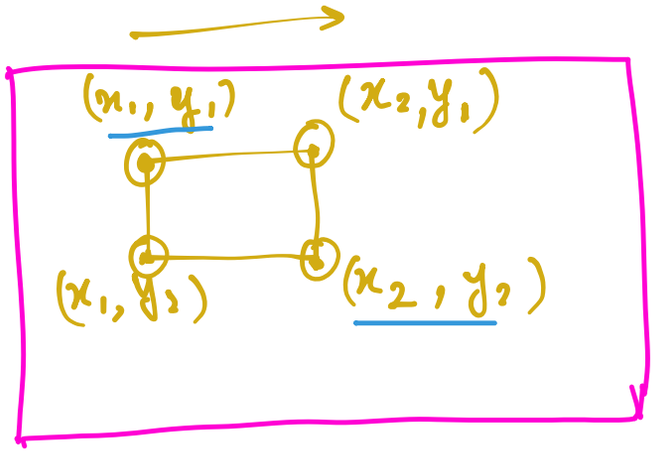
(b) Image similarity: Reuse of an already trained CNN for unsupervised similarity search.

→ Select a convolutional representation. This is referred to as "activations" or "embedding"

→ Store these image representations in a database optimized for KNN.

→ Similar images have "activations" which are close in the Euclidean sense.

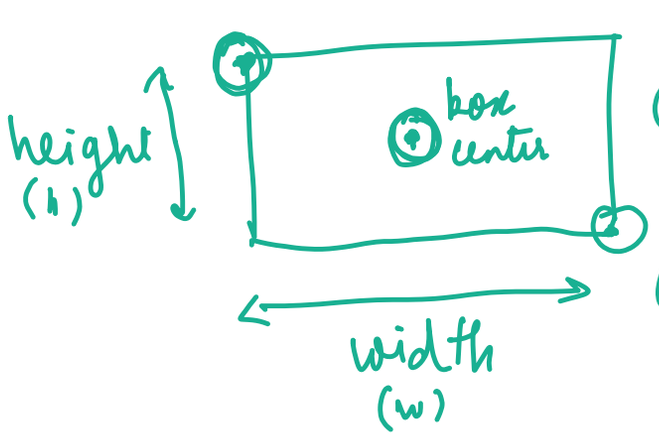
Discussion : Bounding boxes



$h \times w \times 3$

(x_1, y_1) (x_2, y_2) ⟵

$(x_c - \frac{w}{2}, y_c - \frac{h}{2})$



$(x_c + \frac{w}{2}, y_c + \frac{h}{2})$

For our dataset:

Annotations: contain 5 numbers.

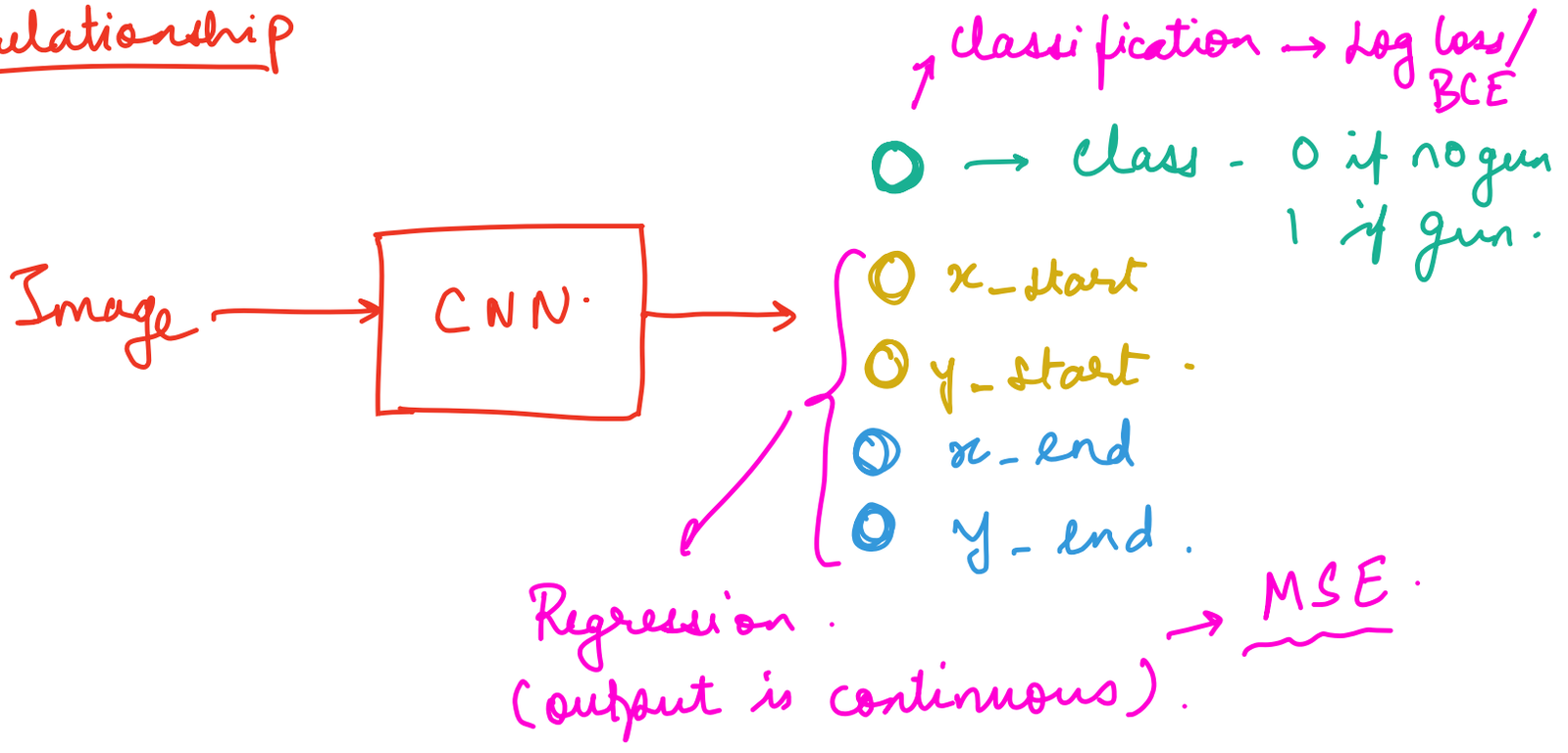
1	2 3	4 5
↓	x_c, y_c	w, h } normalized.

0 - no gun

or

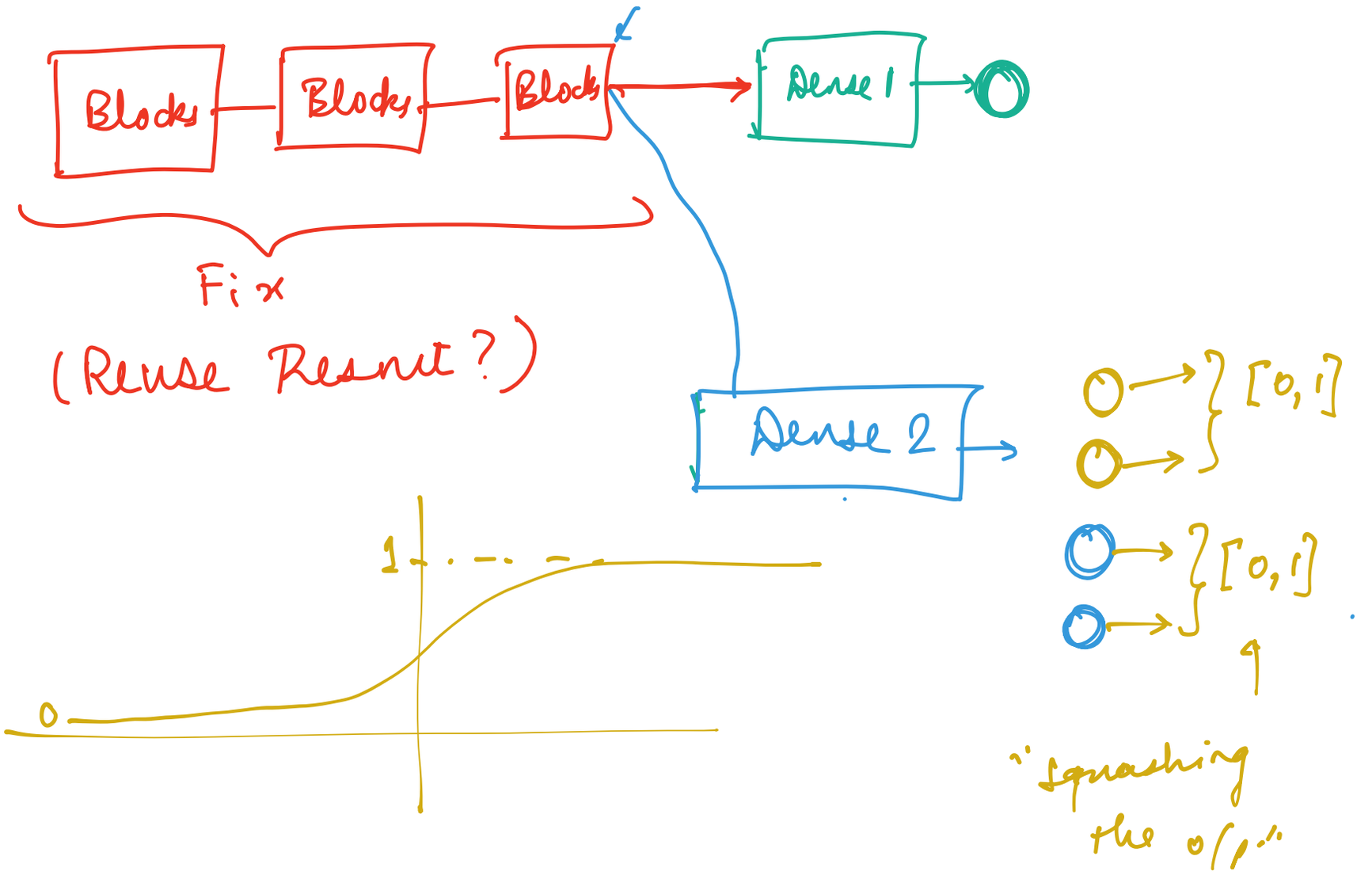
1 - gun

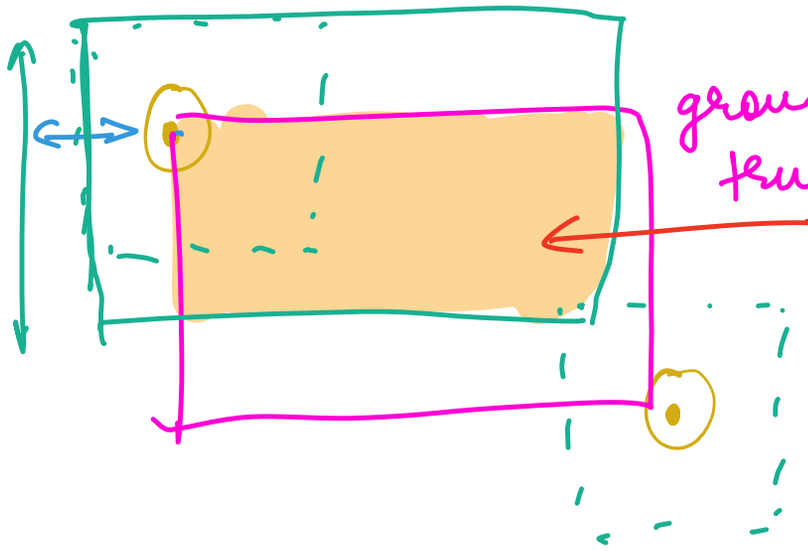
I/O relationship



Issue → only 1 gun per image !!

Architecture for the CNN.





Interpreting 5% error.

~~0.05 × height~~

0.05 × width.

Discussion: Naïve approach:

1. Our model gives only 1 gun per image.
   - more suggestions.
2. Our model's performance is quite bad; only 12% overlap.

How to fix issue 1 ?

* At max, we can have 10 guns.

Con: Not scalable at all !!

50 outputs in o/p layer.

Suggestions :

1] Sliding window over image

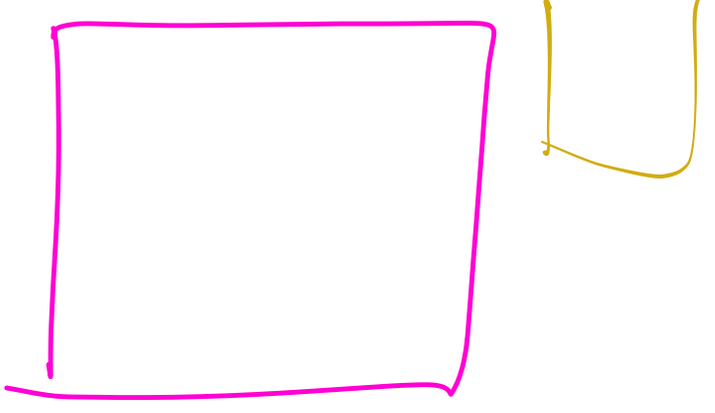
↓

predict if gun —No→ 0.

↓ yes

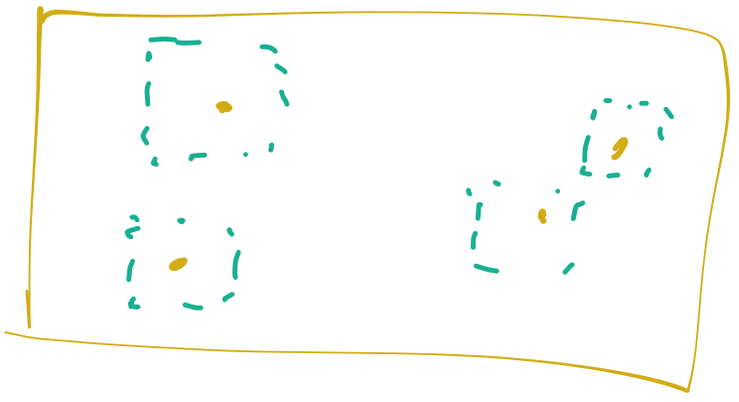
give BB co-ordinates.

2]

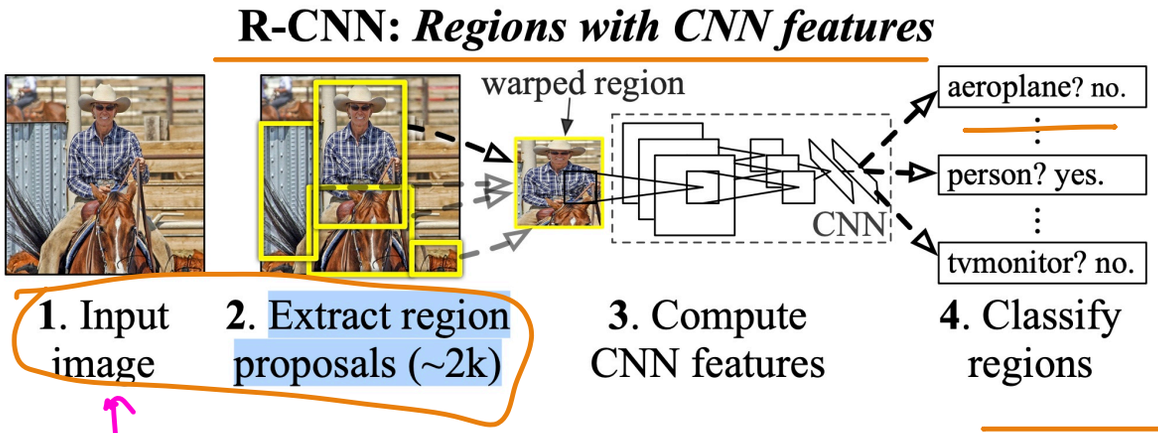


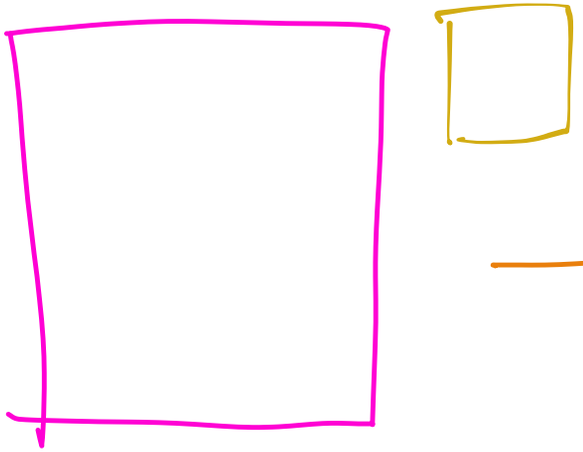
3] Use GradCAM to find regions of interest & then give B.B.

9) Predict multiple centers and fit B.B. around it.



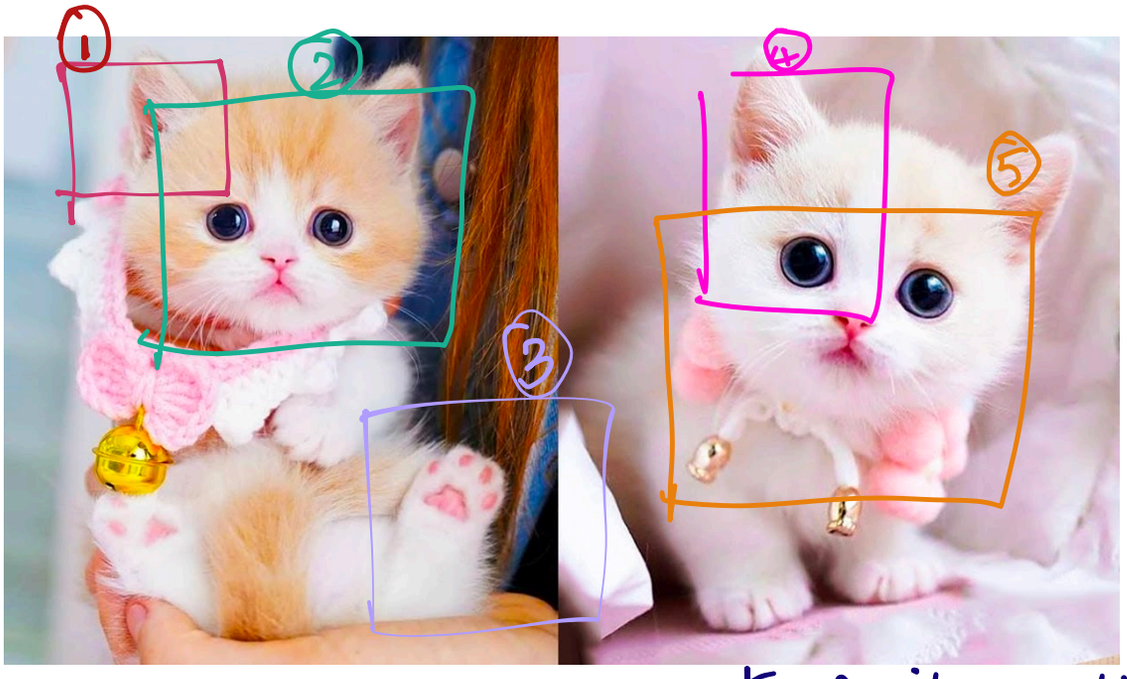
Solution 1 :





→ Extract patches.

Pass each patch through an image classifier → get the prob.



Keep if confidence > 70%.

1 – 10%.

2 – 90% +

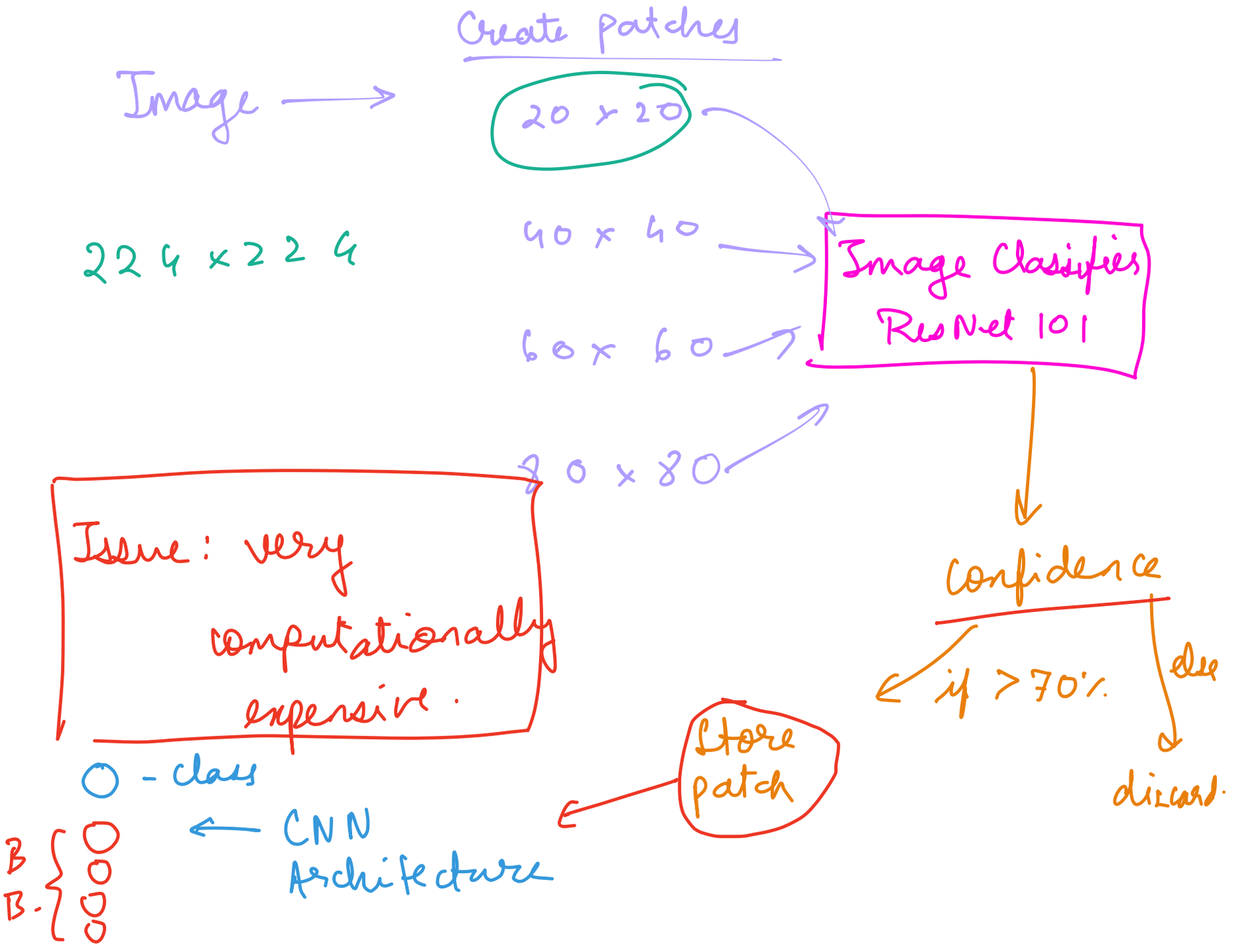
3 – 15%.

4 – 45%.

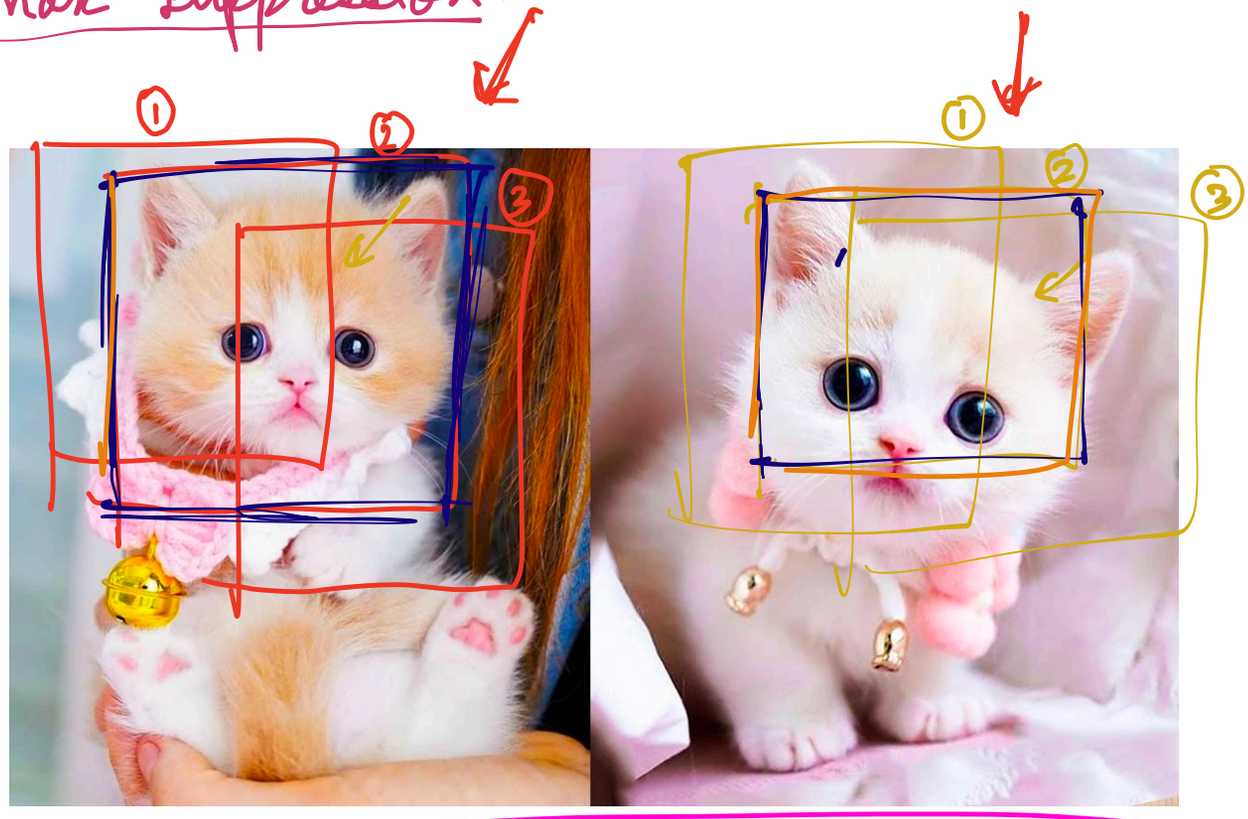
5 – 90%

2

5


Create patches
Image
20 x 20
224 x 224
40 x 40
60 x 60
80 x 80
Image Classifier
ResNet 101
Issue: very computationally expensive.
Confidence
if > 70%
else
Store patch
discard.
- class
CNN Architecture
B
B.

Non max suppression:



Step 1: Select the max.

Step 2: Eliminate windows which have overlap.

IOU threshold. 50%.

Step 3: Keep repeating 1 & 2 till array is empty

→ ① ~~70%~~ ✓

→ ✓

→ ③ ~~92%~~ ✓

Keep: ② 99% ② 92%

3 × 3 × 3 → Old way of doing convolutions.

3 × 3 → Depthwise convolution.

No FCN

If we want to make a CNN without a **fully connected layer** at the end, for the task of classification, What is the replacement for the '**Dense**' Layers used in traditional CNNs?

HINTS | Complete Solution

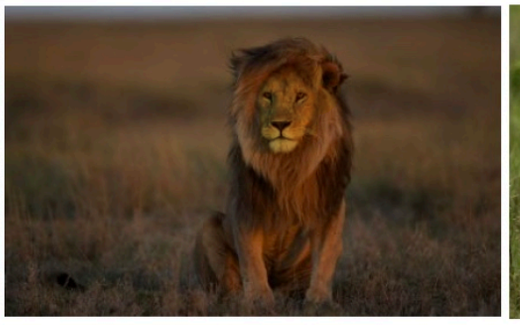
You will get full points if and only if you give CORRECT ANSWER in first attempt. All later attempts will get you ZERO score.

- Avg-Pooling Layers → correct answer.
- Adding Dropout Layers
- 1x1 convolutions Layers →

- Max-Pooling Layers

Question wrong.

Data augmentation effectiveness


Suppose you wish to train a neural network to locate lions anywhere in the images, and you use a training dataset that has images similar to the ones shown above. In this case, if we apply the data augmentation techniques, it will be _____ as there is _____ in the training data.

HINTS | Hint 1 | Complete Solution

You will get full points if and only if you give CORRECT ANSWER in first attempt. All later attempts will get you ZERO score.

- effective, size bias
- **effective, position bias**
- ineffective, position bias
- ineffective, angle bias