RCNN

↳ Convolutional Neural Network

Region Proposal

Image classification → human

→ car

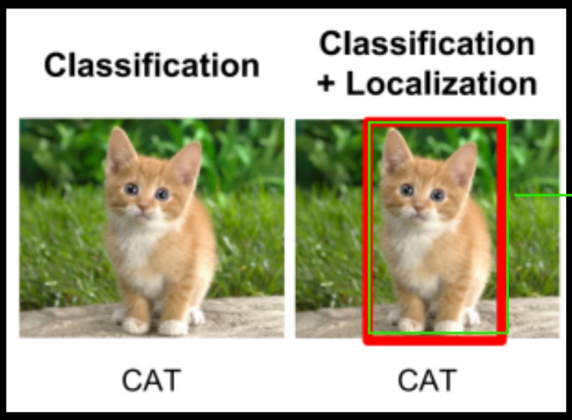
→ dog

CNN

↳ AlexNet

VGGNet

ResNet

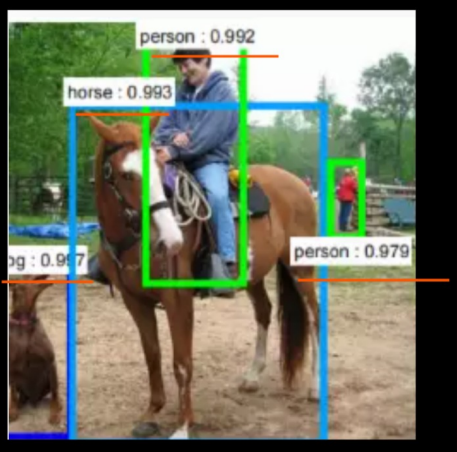


bounding box

cat → bounding box

Classification + Localization

Object detection



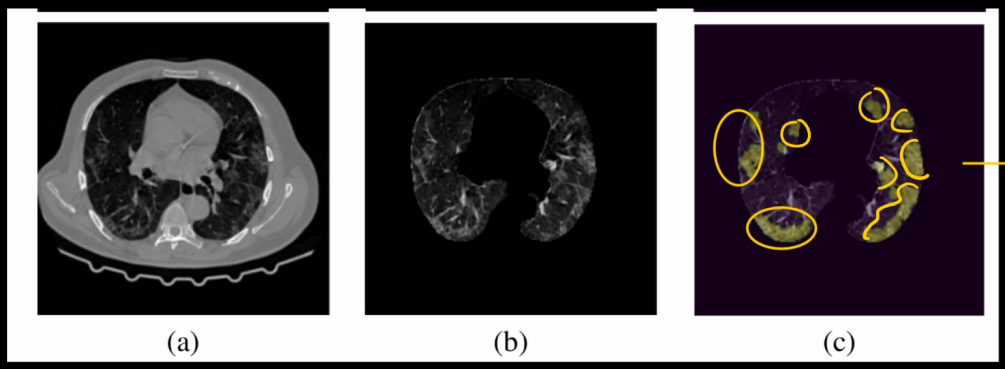
classification

Localisation (single class detection)

object detection (multiple objects available in a single image)

bounding box

segmentation → Cancer cells Lungs in CT scan

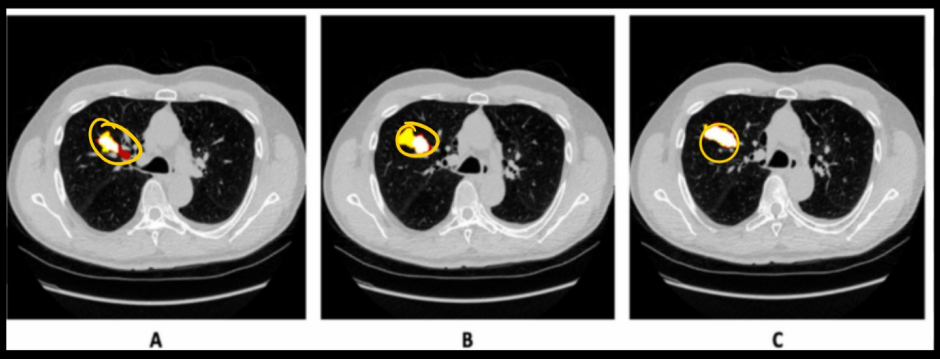


→ Volume of inflamed region

CT scan

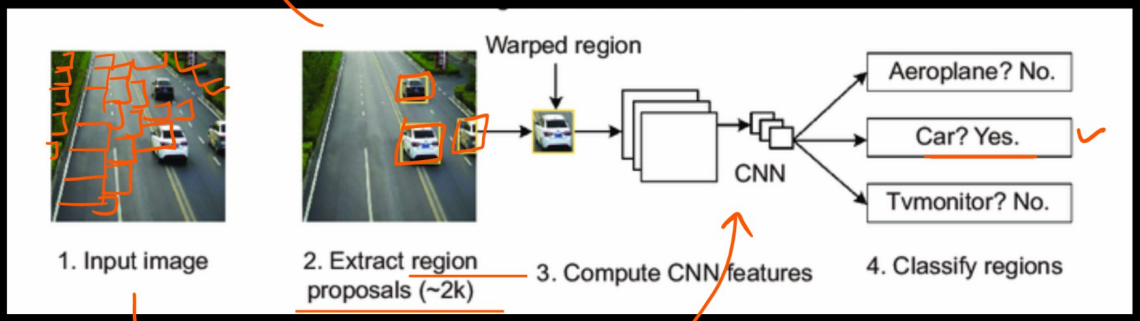
Region of interest

Segmentation



Optimizations

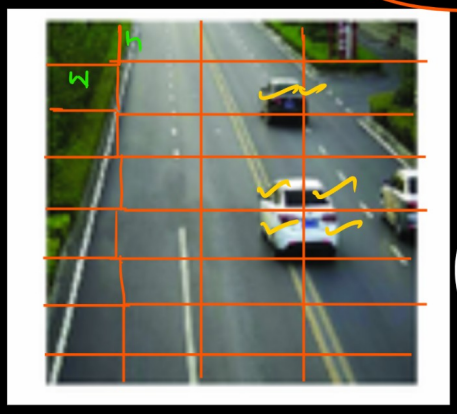
How RCNN Works??



60 M → 2K

Traditional Approach — big failure

↳ CNN



224=H

W = 224

car

Road

CNN (softmax)

car

truck

dog

↓

multiclass Classification

Computationally expensive

for a given input image, how many possible number of sub images can be formed?

$$\sum_{h=1}^{H}\sum_{w=1}^{W}(W-w+1)*(H-h+1)$$

$W = 224$, $H = 224$

$w = h = 1$

600 million boxes

Pipeline

Input Image → Extract Region Proposals (2K) → squere sheped (preprocessing) → Convolutional & maxpooling (AlexNet) (Compute feature Vectors) → 2D Image (4 x 32 x 32)

↓

Flatten → 1D Image (4096,1) → fully connected Layer

↓

bounding box → (x', y', w', h') → Regression → Correct dimension (x, y, w, h) → MSE

car → 0.90

human → 0.01

dog → 0.01

truck → 0.01

cat → 0.01

horse → 0.01

mountain → 0.01

beach → 0.01

cycle → 0.01

scooty → 0.01

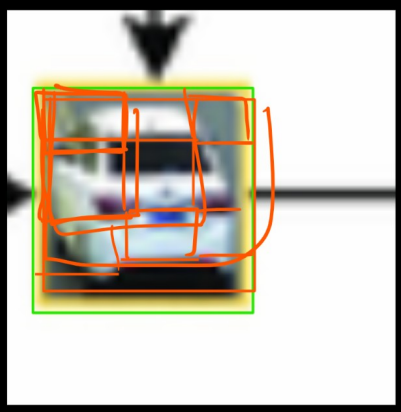
Classification task

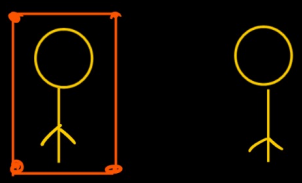
10 classes

output layer

softmax

cross entropy





target box → (x, y, w, h)

dimension

(x', y', w', h')

Regression based problems → Continuous data