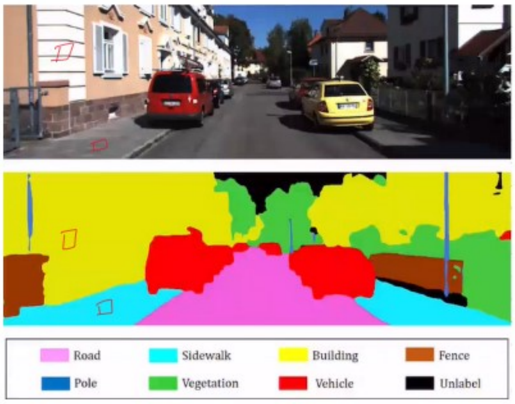
CNN Part - 6

21 September 2023 06:46 PM

U-Net Architecture



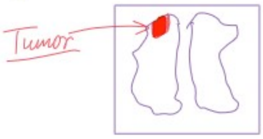
→ It came in 2015, and is one of the most popular papers with 32,000 citations

→ It was specifically designed for bio-medical imaging.

→ Task Performed: Image Segmentation

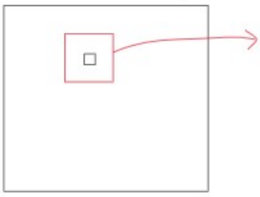
→ The objective is to assign a class label to each and every pixel in the image

It is best used in medical domain, for eg. while detecting the brain tumor, we not only want to detect it but also precisely localize the tumor.



So basically we are segmenting the image into 2 i.e. tumor region and non-tumor region.

* Earlier Methods trained a sliding window setup to predict the class label of every pixel.

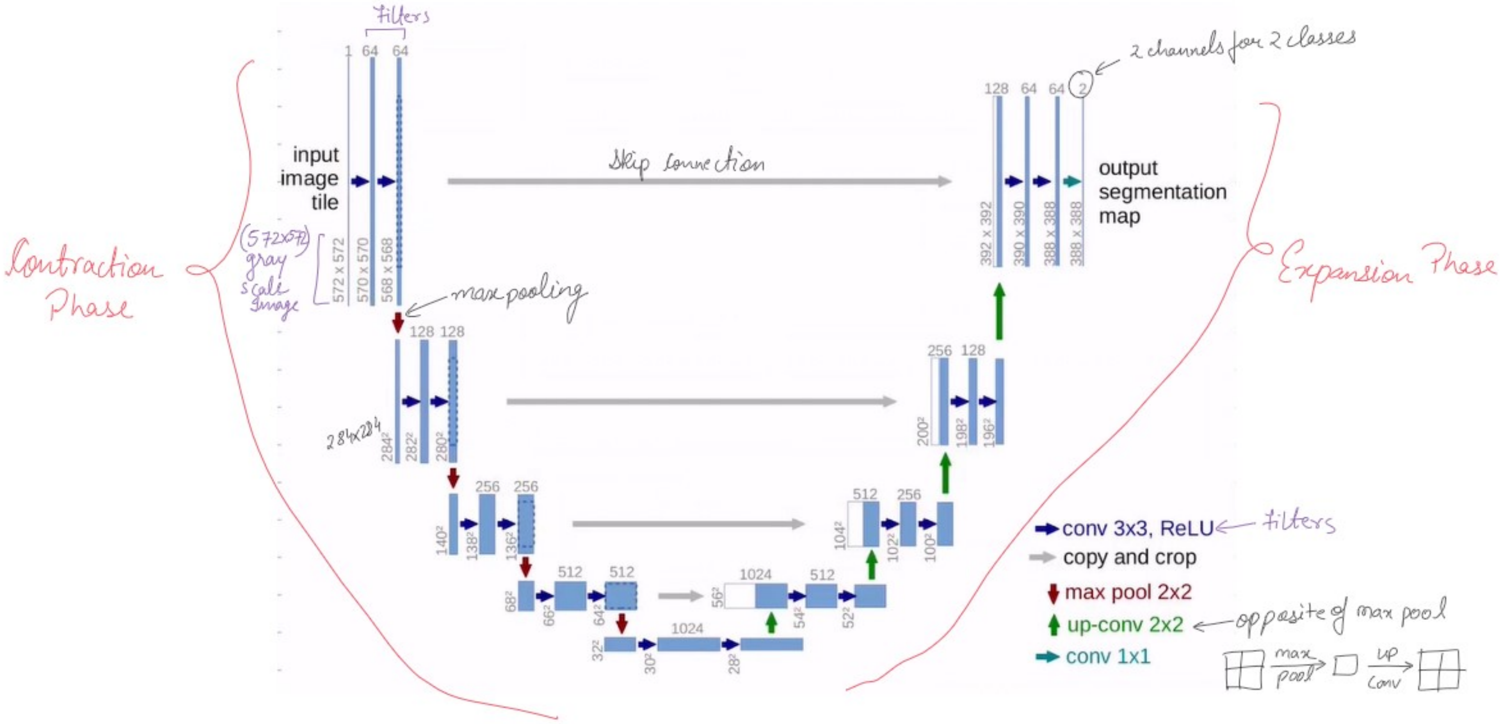


Sliding Window → To find the class label of the black pixel, we train a CNN model to identify the class.

↳ CNN required the entire window as image pixels are correlated to nearby pixels in some context, so that is why the entire window image is given.

Drawback → It is a very expensive process to be performed for each and every pixel.

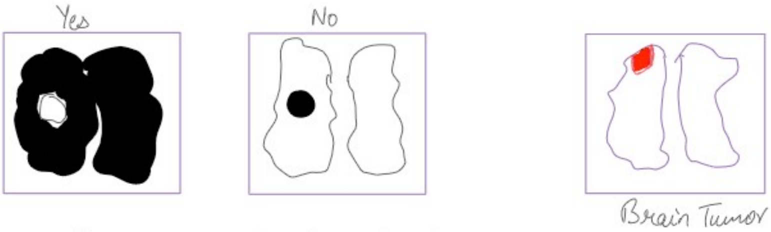
Architecture of U-Net:



→ One of the reasons for gray scale images is that most of the images in medical domain are in gray scale.

→ In the contraction phase, the architecture tries to learn "What features are good in terms of discriminating b/w the classes?"

→ In the expansion phase, the architecture tries to learn "Where in the image is that feature present?"



☆ In the expansion phase we are trying to map the learned features in previous layers through the skip connections to the location in the image.

☆ The size of input image is larger than the output image because of padding; padding is done so that each pixel in the input image gets equal attention/importance.

Main Idea: Supplement the usual contraction phase by an expansion phase where instead of max-pooling we use up-conv, so as to learn what features are good for what region in the output image.

☆ Siamese Network (Facial Recognition)

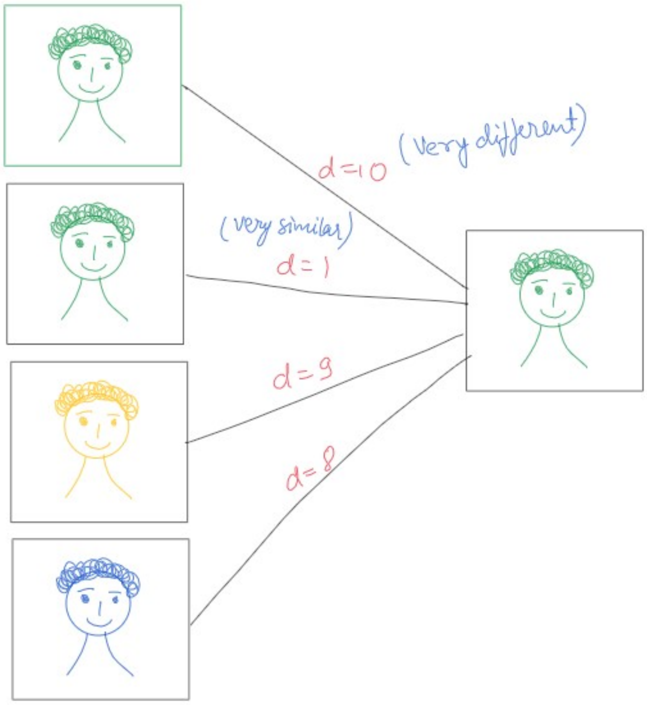
Setup for facial recognition → There are 'K' persons
→ you get an input image
→ Recognize if the input is one of the 'K' people

$d(x^i, x^j)$ = degree of difference between the two images x^i and x^j

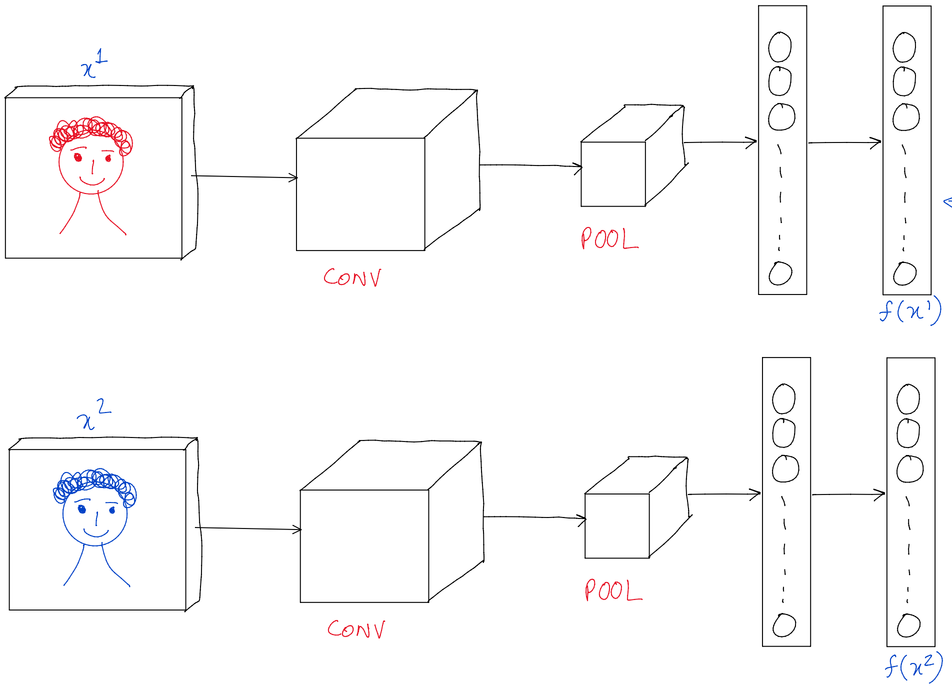
if $d(x^i, x^j) \leq T$: recognize

$> T$: do not recognize.

T → Hyper parameter



Siamese Network



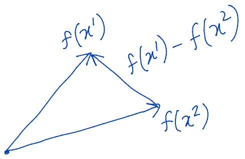
Stopping before the o/p layer

No output layer

← learns very general features

$$d(x^1, x^2) = \underbrace{\left| f(x^1) - f(x^2) \right|}_{\downarrow}{}_2^2$$ ← we take 2 norm and square it.

This is vector
So we convert to scalar by taking its norm.



Goal of learning: learn the parameters of the model such that:
if x^i and x^j are images of the same person, then
$$\left| f(x^i) - f(x^j) \right|_2^2 \text{ should be small.}$$

Q. Why can't we use pixel for pixel comparison?

Ans. It is because for the same person if the lighting conditions are different, the difference between the pixel values will be huge and the result will be "different image", where as in reality it is the image of same person.
Similar Scenario will occur if the person tilts or the new image is mirror image.