

CSE4261: Neural Network and Deep Learning

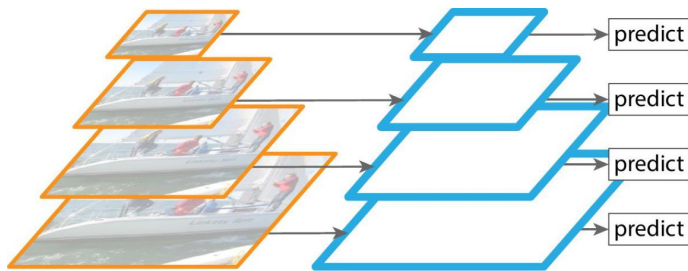
Lecture: 02.07.2025



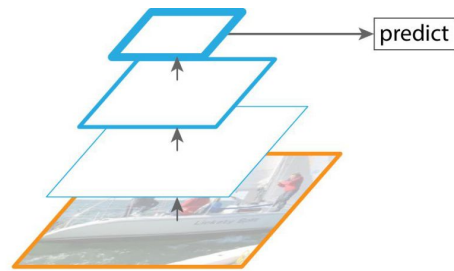
Sangeeta Biswas, Ph.D.
Associate Professor,
University of Rajshahi, Rajshahi-6205, Bangladesh

Feature Pyramid Network (FPN) [2017]

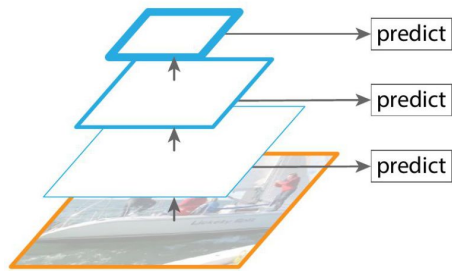
- It creates a feature pyramid from a single-scale input image.
- It enables object detection across various sizes.
- It combines high-resolution, low-level features with low-resolution, high-level semantic features.



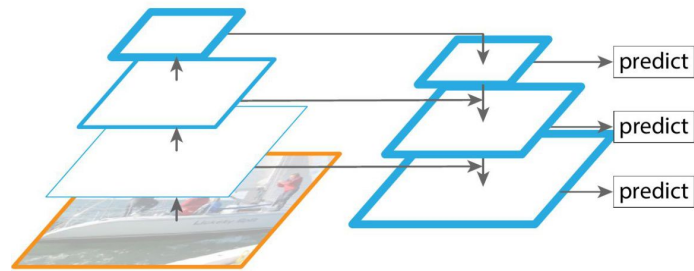
(a) Featurized image pyramid



(b) Single feature map



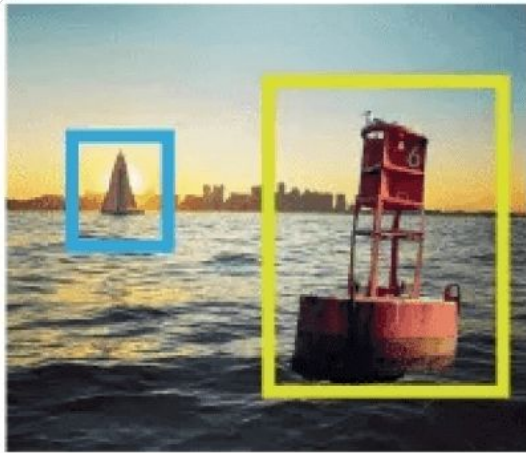
(c) Pyramidal feature hierarchy



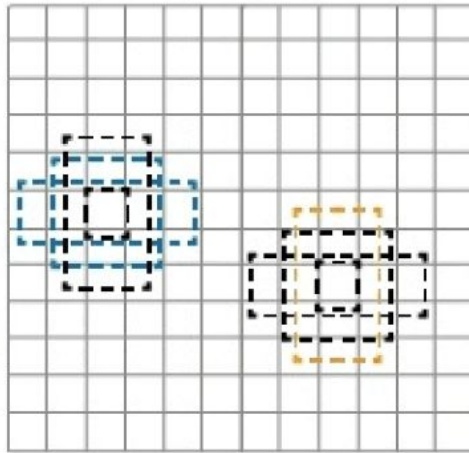
(d) Feature Pyramid Network

Anchors / Anchor Boxes

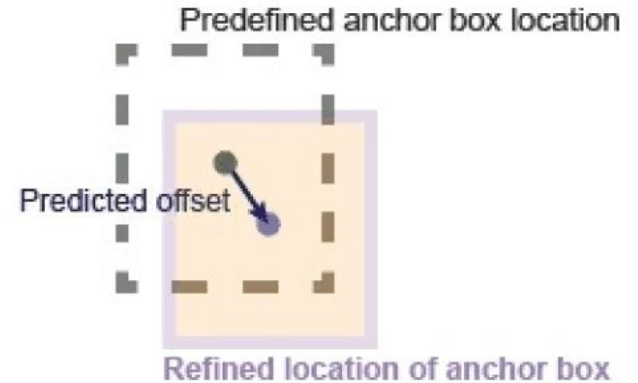
- Anchors are the pre-defined boxes which help to locate objects in an image.



Ground truth image and bounding boxes

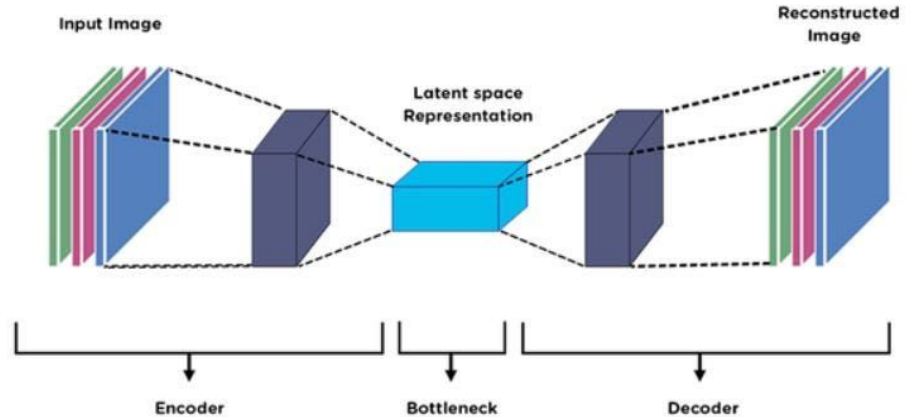


Anchor boxes at each predefined location in each feature map



Encoder-Decoder Architecture

1. It is a neural network framework widely used for sequence-to-sequence tasks, such as image translation, and image captioning.
2. It comprises two main components:
 - a. an encoder: it processes the input sequence and transforms it into a fixed-size context vector (or latent space)
 - b. a decoder: it uses this context vector to generate the output sequence.



Segmentation

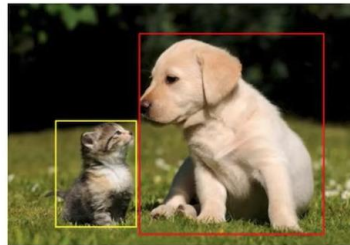
- It is a computer vision technique that partitions a digital image into discrete groups of pixels—image segments—to inform object detection and related tasks.
- It assigns a class label to each pixel in the image, indicating the type of object it belongs to (e.g., road, car, building).

Is this a dog?



Image Classification

What is there in image and where?



Object Detection

Which pixels belong to which object?

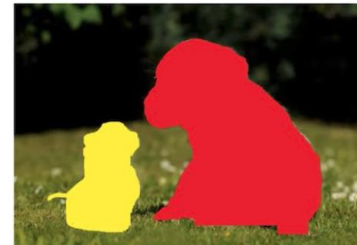
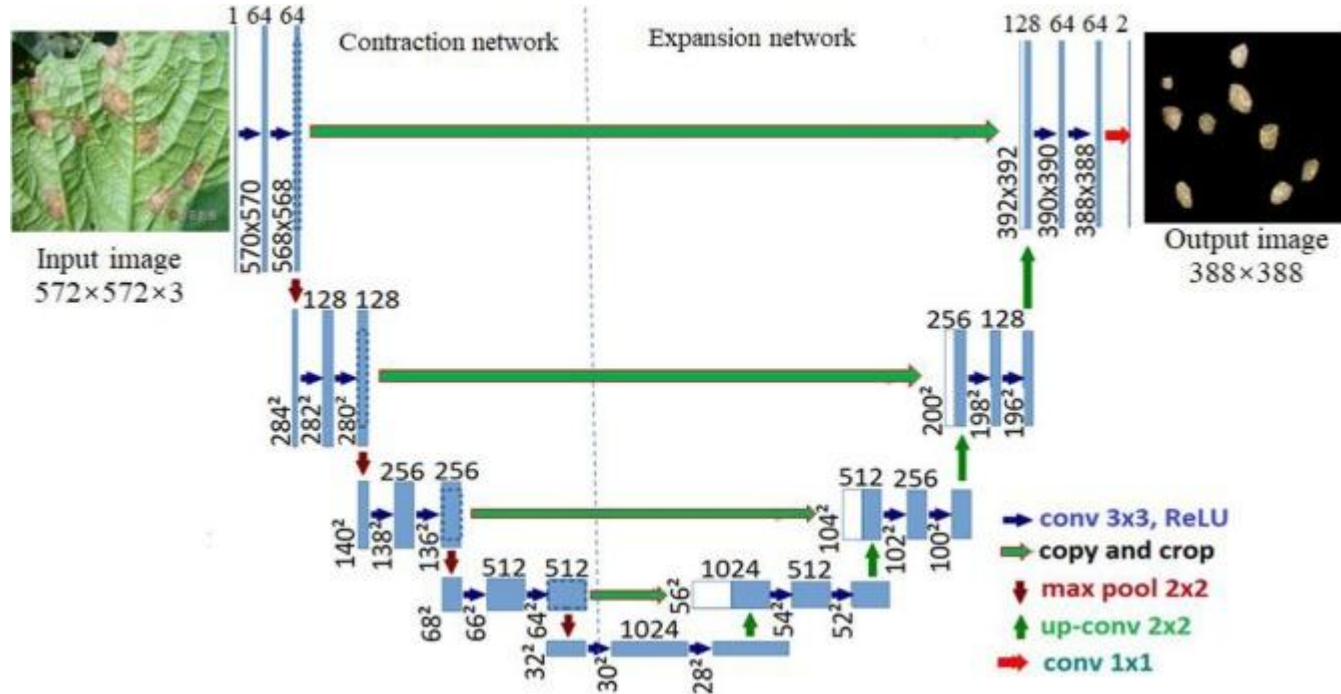


Image Segmentation

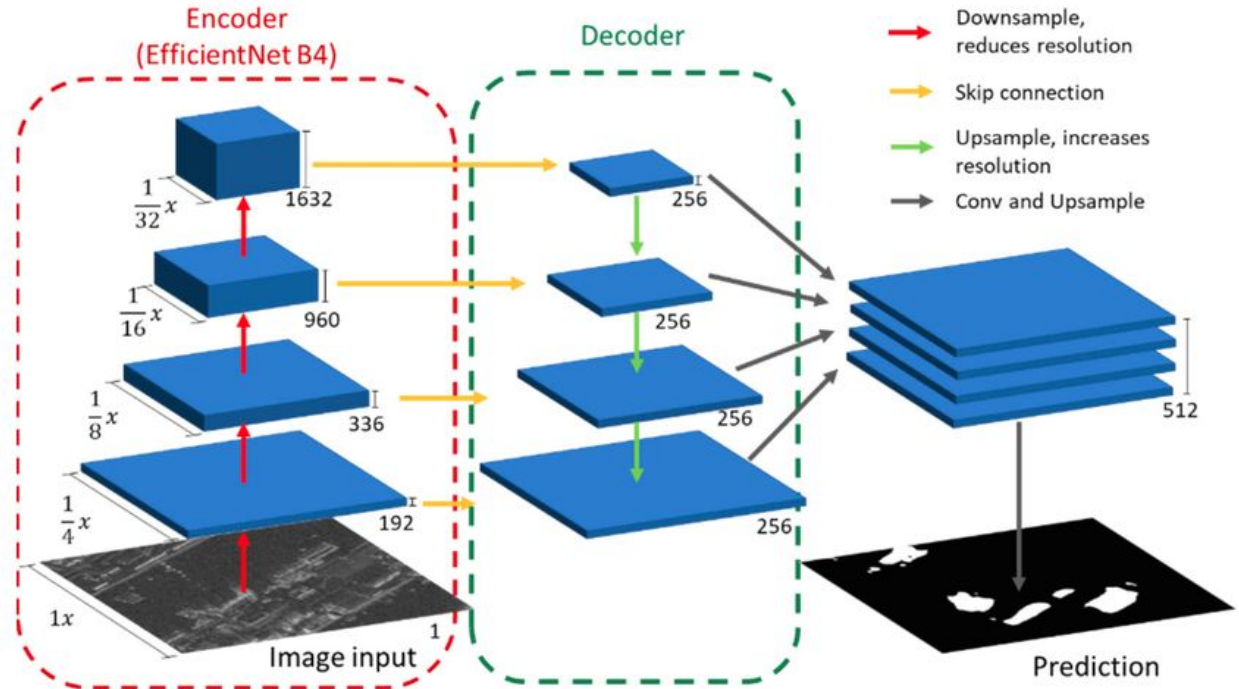
U-Net Segmenter [2015]



Code: https://keras.io/examples/vision/oxford_pets_image_segmentation/

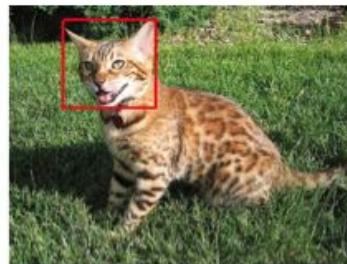
FPN based Image Segmenter

- FPN based segmenter is used model to capture both fine-grained details and global context.
- This technique creates a feature pyramid, which consists of feature maps at multiple scales



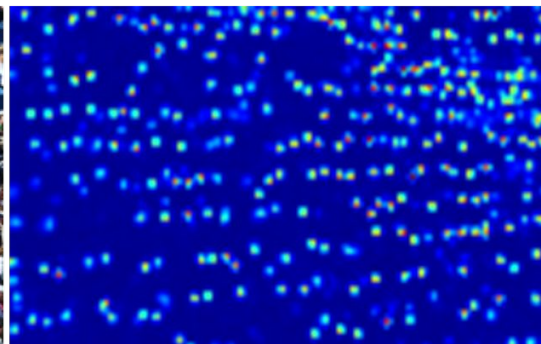
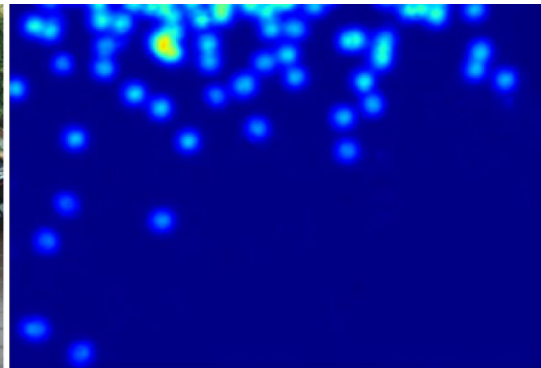
Oxford IIIT Datasets

- Images are from 37 pet category
- Roughly 200 images for each category.
- Annotated information:
 - species and breed name
 - a tight bounding box (ROI) around the head of the animal
 - a pixel level foreground-background segmentation (Trimap)



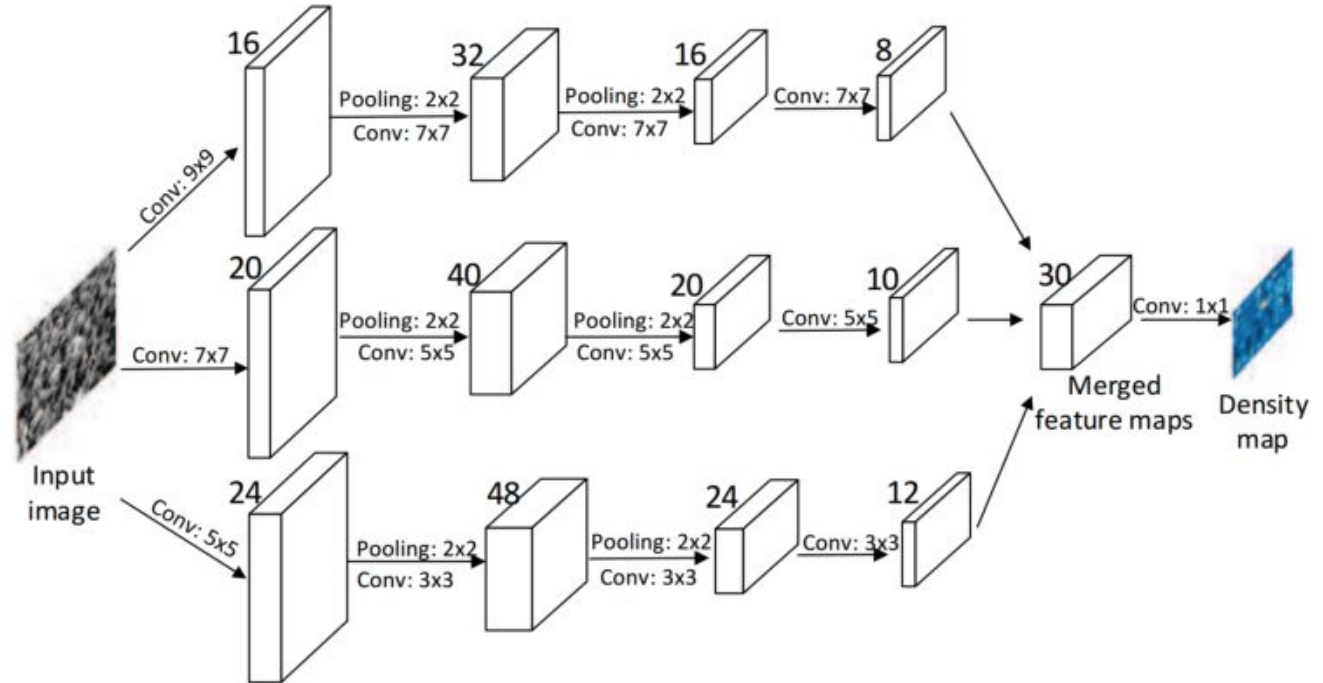
Crowd Counting

- Crowd counting is the act of counting the total number of people present in a certain area.
- It is important for crowd control, public safety
- Output of a DNN based crowd counter can be:
 - Density map
 - Estimated headcount



Multi-Column CNN (MCNN) based Crowd Counting

- MCNN uses the filters of different sizes to model the density maps corresponding to heads of different scale.



UCF-QNRF - A Large Crowd Counting Data Set

- https://www.crcv.ucf.edu/data/ucf-qnrf/UCF-QNRF_ECCV18.zip
- Number of total images: 1535
- Number of Annotations: 1,251,642

