Masters Course
Neural Networks

# Summary
# On
# Weeks 2 & 3
# CNNs Course

Submitted to: Dr/Omar Nasr

Name: Noha Ahmad Darwish
Mohamed Ahmad Nassar

- Image Classification refers to determining the class that an object in an image belongs to.
- Classification with localization refers to classifying an object in image as well as determining its position in the image.
- Detection is when there are multiple objects in an image and we need to localize each of them.
- In **Object Localization**: Locate the presence of objects in an image and indicate their location with a bounding box.

❖ Input: An image with one or more objects, such as a photograph.

❖ Output: One or more bounding boxes (e.g. defined by a point, width, and height).

- The loss function in case of an object detection problem can be divided into 3 parts:-
  1- The loss of the probability that there's an object or not
  2- The loss of the boundary box coordinates
  3- The loss of the classes predictions'
- Landmark detection is one of the applications for the object detection algorithms where we may detect a face and the position of the mouth in the face to determine if the person is happy.
- In the Sliding window algorithm, To detect a car in a test input image, we start by picking sliding window of size (x) and then feeding input region (x) to trained convnet by sliding window over every part of input image
- For each input region, convnet outputs whether it has a car or not. We run sliding windows multiple times over the image with different window sizes, from smaller to larger, hoping a window size would fit the car and allow convnet to detect it.
- Computational cost is a huge disadvantage of sliding window algorithm.
- Increasing window and stride size makes it faster but at cost of decreased accuracy
- The convolution implementation for the sliding window saves computational power and time by processing the entire image in a single forward path instead of processing it as grid by grid.

- Intersection over union (IOU) is a measure of the overlap between 2 bounding boxes.
- IOU is the ratio between:-
  1- the intersection of the ground truth and the predicted BB
  2- the union between the ground truth and the BB
- Non Max Suppression is used to eliminate the multiple detection for the same object according to the probability of detection times the probability of a certain class.
- After choosing the BB with the highest probability, we then eliminate the BBs having high IOU with the chosen BB.
- Anchor boxes are used to detect multiple objects in the same grid cell.
- To do multiple object detection in the same grid cell, we can predefine 2 different anchor boxes and then associate 2 predictions with these anchor boxes.
- Upon using the anchor boxes, the output is now the original output multiplied by 2.
- YOLO algorithm uses the previous mentioned algorithms for detection and classification.
- Region proposals use segmentation and blobs detection to minimize the convolution operations performed on an image.
- **Face Verification**. A one-to-one mapping of a given face against a known identity (e.g. is this the person?).
- **Face Identification**. A one-to-many mapping for a given face against a database of known faces (e.g. who is this person?).
- One-shot learning are classification tasks where many predictions are required given one (or a few) examples of each class, and face recognition is an example of one-shot learning.
- Siamese networks are an approach to addressing one-shot learning in which a learned feature vector for the known and candidate example are compared.
- Contrastive loss and later triplet loss functions can be used to learn high-quality face embedding vectors that provide the basis for modern face recognition systems.

- Triplet loss involves an anchor example and one positive or matching example (same class) and one negative or non-matching example (differing class).
- The loss function penalizes the model such that the distance between the matching examples is reduced and the distance between the non-matching examples is increased.
- The result is a feature vector, referred to as a ‘face embedding,’ that has a meaningful Euclidean relationship, such that similar faces produce embeddings that have small distances (e.g. can be clustered) and different examples of the same face produce embeddings that are very small and allow verification and discrimination from other identities.
- Deep layers learn more complex features while earlier layers detect simple features.