- The architectural decisions were based on Hebbian theory (**Hebbian theory** is a neuroscientific theory claiming that an increase in synaptic efficacy arises from a presynaptic cell's repeated and persistent stimulation of a postsynaptic cell) and multi-scale approach.
- RCNN uses 2 step approach:
  - Finds object location proposals based on color distribution and texture.
  - Uses CNN to detect object in those locations.
- Work of Arora et al. [2] states that if the probability distribution of the dataset is representable by a large, very sparse(non-uniform) deep neural network, then the optimal network topology can be constructed layer after layer by analyzing the correlation statistics of the preceding layer activations and clustering neurons with highly correlated outputs.
- Inception architecture started out as a case study for assessing the hypothetical output of a sophisticated network topology construction algorithm that tries to approximate a sparse structure implied by [2] for vision networks and covering the hypothesized outcome by dense, readily available components.
- In the lower layers (the ones close to the input) correlated units would concentrate in local regions.
- It suggests that adding an alternative parallel pooling path in each such stage should have additional beneficial effect too.
- Features of higher abstraction are captured by higher layers, their spatial concentration is expected to decrease. => more no. of filters needed to retain more information => ratio of 3×3 and 5×5 convolutions should increase as we move to higher layers.
- It is also based on the practical intuition that visual information should be processed at various scales and then aggregated so that the next stage can abstract features from the different scales simultaneously.
- Auxiliary layers are added to prevent gradient vanishing at initial layers and also to regularise by adding to the final loss function and restricting the model.
- They trained different models with same architecture only by changing random order of images and type of data augmentation and followed Andrew Howard cropping.