**Differences between R-CNN, Fast R-CNN and Faster R-CNN**

R-CNN: At test time, the R-CNN model generates around 2000 category-independent region proposals for the input image using selective search, extracts a fixed-length feature vector from each proposal using a CNN, and then classifies each region with category-specific linear SVMs. However, R-CNN is slow because it performs a ConvNet forward pass for each object proposal, without sharing computation.

Fast R-CNN: In R-CNN, training is a multi-stage pipeline that involves extracting features, fine-tuning a network with log loss, training SVMs, and finally fitting bounding-box regressors. To deal with this problem Fast R-CNN was proposed. It takes the whole image and region proposals as input in its CNN architecture in one forward propagation. Fast R-CNN introduced Region of Interest (RoI) pooling, which allows CNN to directly extract features from the proposed regions, eliminating the need for separate feature extraction for each region. It also uses the softmax layer instead of SVM in its classification of region proposal, which proved to be faster and generate better accuracy than SVM.

Faster R-CNN: After the Fast R-CNN, the bottleneck of the architecture is selective search. As it needs to generate 2000 proposals per image, it constitutes a major part of the training time of the whole architecture. So, to solve this problem, selective search was replaced by a region proposal network (RPN). RPN eliminates the dependency of selective search by integrating region proposals into the model itself. RPN uses convolution layers to analyse the image features and simultaneously predict. RPN and Fast R-CNN were merged into a single network by sharing their convolution features. This scheme converges quickly and produces a unified network with convolutional features that further improve performance and smoother inference flow.