- Problem with CNN : They require a fixed input image size (e.g., 224×224), which limits both the aspect ratio and the scale of the input image.
- The cropped region may not contain the entire object, while the warped content may result in unwanted geometric distortion. Recognition accuracy can be compromised due to the content loss or distortion.
- Only FCs need fixed input size => output of convnet(with arbitrary input size) can be changed to suit FCs.
- SPP uses sliding windows of sizes proportional to input image size => output size is constant.
- It pools the variable size feature map with different sizes and then concatenates them to form fixed length vector.
- With a pyramid level of n×n bins, we implement this pooling level as a sliding window pooling, where the window size win = ceil(a/n) and stride str = floor(a/n)
- SPP was trained on images of different sizes by 2 fixed size networks that take 180x180 and 224x224 images. Both networks share all parameters.Images that are big are not cropped but resized to fit 180x180 from 224x224 which is cropped from resized image.
- When multi-level pooling was applied with 50 bins and 30 bins, both showed high accuracies compared to no SPP network => improvement due to multi-level pooling; not due to more parameters.
- Random multi-size training between [180,224] yields better results than single size but worse compared to two-size as 224 size is less likely.
- Full image testing improves accuracy in which aspect ratio of image is maintained and min dim set to 256, compared to 224x224 centre crop.
- Multiple views are tested on an image by resizing it to different scales and cropping 224x224 windows from them.
- SPP can be used for object detection in RCNN to get multi-scale feature vector for each region from the same image feature map.