2.1.4 R-CNN Training

The details of R-CNN training will be discussed at the following sub-sections.

2.1.4.1 Supervised Pre-training

CNN was previously trained on a large auxiliary data set (ILSVRC2012 classification [9,13]) using only image-level additional tags. CNN was previously trained on data set (ImageNET ILSVRC2012 [9,13]) using only additional tags. This training was carried out using Caffe Deep Learning framework.

2.1.4.2 Domain-Specific Fine-Tuning

In order to adjust Convolutional Neural Network to new task and domain name , SGD training was performed to function parameters using only warped region proposals. Convolutional Neural Networks ImageNetspecific 1000 way classification layer has been changed over with the N+1 way classification layer. Convolutional Neural Network framework has not been changed here. (N = 20 for VOC and N = 200 for ILSVRC2013).

All region proposals, which are equal to or greater than 0.5 iou overlap value, were accepted as positive for the box class and others were accepted as negative. In each SGD iteration, 32 positive windows and 96 background windows are properly sampled to create a mini stack of 128 sizes.

2.1.4.3 Object Category Classifiers

Here, binary classifier training was used to perceive cars. It is a positive example of an image area in which a car is tightly enclosed. In a similar way, a background region that is not interested in cars is a negative example. It is unclear how a partially overlapping region of the car should be labeled. the unclear state is solved by specifying an IoU overlap threshold value. Areas below this threshold value are identified as negative and those above the threshold value as positive.The overlap threshold “0.3” was chosen by conducting a grid search on the verification set. Once the features are removed and the training tags are applied, SVM is applied optimally to all classes.

2.2 Faster R-CNN

The Faster R-CNN composed of two component. The first component is a conventional network used to propose zones called RPN, and The second component is the Faster R-CNN detector which utilize the region proposals. The whole system comes from a single composite network created for object detection [10].

The first component is a conventional network used to propose regions called RPN (Region Proposal Network)

2.2.1 Region Proposal Networks (RPN)

In this study, RPN receives as input image and produces a set of rectangle object tender which all have objectivity score. The RPN is designed with a fully convoluted network. Since calculations are shared with a Faster R-CNN object detection network, it is assumed that both networks share a common set of layers of convolution.

A mini network is moving on the exit of the convolution property map by the last shared convolution layer to produce region proposals. As an input, it takes the space window of the convolutional property map n × n(used as n=3). All sliding windows in work are matched to low dimensional property. This feature composed of two sister fully bound box-regression and box-classification layers. In this mini network, all the fully connected layers are shared in all spatial locations.This framewoek is carried out by the convolution Layer and following the two brothers 1 x 1 convolution layer.

2.2.1.1 Training RPNs (Region Proposal Network)

In this study, RPNs are trained end-to-end with backpropagation and SGD. In order to train this network, "image-centric sampling" strategy is applied.In the study, all the batchs come from the images involving negative and positive sample anchors.

When the missing functions of all the anchors are optimized here, the orientation of the negative examples is realized. For this reason, a random sample of 256 anchors is shown in an image instead. According to this, if there are more than 128 positive samples, it is filled with stacked samples. Otherwise, it is filled with negative examples. In