remote sensing

MDPI

Article

Cascade Object Detection and Remote Sensing Object Detection Method Based on Trainable Activation Function

S. N. Shivappriya [1], M. Jasmine Pemeena Priyadarsini [2], Andrzej Stateczny [3,*], C. Puttamadappa [4] and B. D. Parameshachari [5]

1 Kumaraguru College of Technology, Coimbatore 641049, India; shivappriya.sn.ece@kct.ac.in
2 School of Electronic Engineering, Vellore Institute of Technology, Vellore 632014, India; jasmin@vit.ac.in
3 Chair of Geodesy, Gdansk University of Technology, 80232 Gdańsk, Poland
4 Department of Electronics and Communication Engineering, Dayananda Sagar University, Bangalore 560078, India; puttamadappa-ece@dsu.edu.in
5 Department of Telecommunication Engineering, GSSS Institute of Engineering and Technology for Women, Mysuru 570016, India; paramesh@gsss.edu.in
* Correspondence: andrzej.stateczny@pg.edu.pl; Tel.: +48-609-568-961

Abstract: Object detection is an important process in surveillance system to locate objects and it is considered as major application in computer vision. The Convolution Neural Network (CNN) based models have been developed by many researchers for object detection to achieve higher performance. However, existing models have some limitations such as overfitting problem and lower efficiency in small object detection. Object detection in remote sensing hasthe limitations of low efficiency in detecting small object and the existing methods have poor localization. Cascade Object Detection methods have been applied to increase the learning process of the detection model. In this research, the Additive Activation Function (AAF) is applied in a Faster Region based CNN (RCNN) for object detection. The proposed AAF-Faster RCNN method has the advantage of better convergence and clear bounding variance. The Fourier Series and Linear Combination of activation function are used to update the loss function. The Microsoft (MS) COCO datasets and Pascal VOC 2007/2012 are used to evaluate the performance of the AAF-Faster RCNN model. The proposed AAF-Faster RCNN is also analyzed for small object detection in the benchmark dataset. The analysis shows that the proposed AAF-Faster RCNN model has higher efficiency than state-of-art Pay Attention to Them (PAT) model in object detection. To evaluate the performance of AAF-Faster RCNN method of object detection in remote sensing, the NWPU VHR-10 remote sensing data set is used to test the proposed method. The AAF-Faster RCNN model has mean Average Precision (mAP) of 83.1% and existing PAT-SSD512 method has the 81.7%mAP in Pascal VOC 2007 dataset.

Keywords: Additive Activation Function; cascade object detection; Faster Region based Convolution Neural Network; Fourier series and linear combination of activation function; remote sensing

check for updates

Citation: Shivappriya, S.N.; Priyadarsini, M.J.P.; Stateczny, A.; Puttamadappa, C.; Parameshachari, B.D. Cascade Object Detection and Remote Sensing Object Detection Method Based on Trainable Activation Function. *Remote Sens.* **2021**, *13*, 200. https://doi.org/10.3390/rs13020200

Received: 9 November 2020
Accepted: 5 January 2021
Published: 8 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In computer vision, object detection is a significant step to develop various systems such as intelligent surveillance, autonomous driving vehicles, and motion capture. The object detection is required in various applications security, home automation, retail, safety, control applications, traffic monitoring, etc. In intelligent surveillance systems, object detection is an important task to detect the useful insights from Remote Sensing images, image data, such as sidelined objects, intrusion detection, and traffic data collection [1,2]. The object detection in remote sensing images is a challenging task due to the presence of object at different scale, viewpoint variation, and shadows. The motion, color, and shape features are mostly used in the various segmentation and tracking methods in cascade object detection. However, this is difficult to accurately detect the foreground object in the image data [3]. There are many challenges in object detection due to complexities

in an image data such as unpredictable objects motion, noise, partial or full occlusion, variation in background, illumination variation, etc. If the images arecaptured by moving camera, the foreground and background features of each frames change their position [4–6]. Recently, the deep learning method provides the efficient performance in cascade object detection. However, still some challenges in the existing cascade object detection methods is detection of small objects in the video, especially when resources are limited [7,8].

The Convolutional Neural Networks (CNNs) have a significant performance in various computer vision tasks such as object detection, image classification, human pose estimation, semantic segmentation, etc. The CNNs methods have the capacity to effectively learn rich representations compared tohand-craftedconventional representations [9,10]. Among CNN models, the R-CNN framework is one of the most influential methods that performs a classification based on CNN using various methods. Two enhanced R-CNN models such as Fast R-CNN model and Faster R-CNN model are applied for the classification. The Fast R-CNN method learns convolutional feature maps before extracting features from input images for classification. The Faster R-CNN model has shared convolutional layers for a combination of Region Proposal Network (RPN) and Fast R-CNN [11–13]. The cascade learning process determines the cascade learning parameter to increase the classifier efficiency. The cascade parameter involves in a number of stages, and each stage has a number of weak classifier and thresholds [14,15]. In this paper, the AAF-Faster RCNN method is proposed for object detection and it has the advantage of better convergence and clear bounding values. In remote sensing object detection, objects such as airplane, ships, storage tanks, baseball diamonds, tennis courts, basketball courts, ground track fields, harbors, bridges, and vehicles are detected by the proposed AAF-Faster RCNN method. The contributions of the research are discussed as follows:

1. The AAF activation is analyzed in the input data and update the loss function for a Faster RCNN method to improve the detection efficiency. The Fourier series and Linear activation function are used to update the loss function.
2. The proposed AAF-Faster RCNN method is applied in the object detection to increases the efficiency of detection and cascade object detection method is applied to detect the small object in datasets. The analysis shows that AAF-Faster RCNN model has an efficient performance in small object detection.
3. The MS COCO datasets and Pascal VOC 2007/2012 were used to evaluate the robustness of the proposed AAF-Faster RCNN model. The proposed AAF-Faster RCNN model has a robust performance on MS COCO datasets and Pascal VOC 2007/2012 dataset for object detection.
4. In Remote Sensing, object detection is achallenging process due to the presence of objects at different scales, and existing methods have poor localization in small object detection. The proposed AAF-Faster RCNN model has the advantage of better convergence and clear bounding values. The proposed AAF-Faster RCNN model provides the effective detection of object in the Remote Sensing images.
5. The NWPU VHR-10 data set is applied to test the efficiency of proposed AAF-Faster RCNN model for object detection in remote sensing images. The proposed AAF-Faster RCNN model uses the cascade method to effectively detect the small object in the image.

This paper is organizedas follows. The literature review is presented in Section 2 and the proposed AAF-Faster RCNN method is explained in Section 3. The experimental design is provided in Section 4 and the experimental result is provided in Section 5. The conclusion of this research work is provided in Section 6.

2. Literature Review

Convolution Neural Networks (CNN) have been highly applied in object detection and have achieved considerable improvement in the performance. Recent methods involved in applying the cascade object detection method were reviewed in this section.

MOST WIEDZY Downloaded from mostwiedzy.pl

Liu et al. [16] proposed the Pay Attention to Them (PAT) method to combine the bottom-up and top-down operating strategy in CNN for general object detection. The PAT method applies the CNN regression method on the entire input images. The intelligent agent was applied in attention mechanism refine the sub-regions that contain the relevant object in the image. The refining process was carried out until bounding box were scaled and removed the overlapping parts to provide final output. Two benchmark datasets such as Pascal VOC and MS COCO to estimate the PAT method efficiency. The analysis shows that the PAT method increases baseline detector performance. The PAT method uses the discrete action set to realize the attention mechanism and tends to provide the irrelevant information for cascade refinement.

Cai and Vasconcelos [17] proposed cascade R-CNN for the object detection method to reduces the overfitting problem and computational time. The cascade R-CNN method consists of sequence of detection trained with Intersection over Union (IoU). The detectors were trained sequentially, and the output of detector wasused for next detection training. The hypotheses quality was progressive improved by the resampling and reduces the over fitting problem. The benchmark datasets such as COCO, VOC, KITTI, CityPerson, and WiderFace were used to estimate the performance. The experimental analysis shows that the cascade R-CNN has ahigher performance compared with Mask R-CNN. The small object detection was not addressed properly in this method and need to be enhanced.

Cevikalp and Triggs [18] applied Support Vector Machine (SVM) that use a short cascade of asymmetric one-class classifier to reject the negative class within the sliding window framework. The asymmetric representation was highly focused on the coherent positive class and tightly modelling rare extent that can lead to simpler classification and faster rejection. The developed method uses the simple convex model to progressively improve the bound based on positive class. The dataset such as FDDB face detection, INRIA Person and ESOGU face detection were used to evaluate the efficiency of the model and also analyzed in VOC dataset. The model is involved in significantly reducing the computation complexity and computational time due to the elimination of large negative classes. The model has lower efficiency in detecting the small objects and CNN method was needed to analyze the features. The developed method failed to analyze the inter-class interaction information for the overlapping of the object detection.

Zhong et al. [19] proposed a lightweight cascade structure to increase the performance of the Region Proposal Network (RPN) for object detection. The pre-trained RPN, cascade RPN, and constrained ratio of negative over positive class were applied for object detection. The extracted proposal was used for the object detection and datasets such as VOC, COCO, and ILSVRC were used to estimate the method's efficiency. The evaluation shows that the RPN based method achieveda considerable performance in the dataset with less computational cost. The performance of the model degrades for the large threshold of IoU due to the elimination of small object in refinements. The developed method failed to reduce the overheating problem which affects the performance.

Zhu et al. [20] proposed two stage method for object detection technique: (1) a locally sliding line-based point regression (LocSLPR) and (2) a rotated cascade R-CNN method. The LocSLPR method estimates the object outlier that denotes the sliding line intersection and bounding box of the object. The rotated cascade R-CNN methodgradually regresses the target object that increases efficiency of object detection. The developed method has a considerable performance in the aerial image dataset namely DOTA and the developed method has lower efficiency compared to the one-stage detection due to the refinement of the relevant information.

Dai and Wei [21] proposed two-stage regression-based cascade object detection, namely HybridNet, for fast and precise object detection. The regression modes are used in the first and second stage. In second stage, a transitional stage was added to extract the features of desired refinement on high resolution feature map. The datasets such as KITTI and PASCAL VOC were used to analyzethe method's efficiency. The experimental results showed that HybridNet method has higher performance and less computational time in

MOST WIEDZY Downloaded from mostwiedzy.pl

object detection. The small object detection is not well addressed in this method due to the refinement of negative classes.

Zouet al. [22] established multi-task cascade CNN for hierarchical image classification and object detection, in recognition large scale commodity. The object detection method was used to locate the object and the hierarchical clustering method was used to develop a category and an image classification model in a tree shape. The developed method identified the group of classes to provide insight intothe data. The experimental analysis shows that the multi-task cascade CNN method has higher efficiency in object detection and the object detection efficiency needs to be improved.

Xu et al. [23] presented a Deep Regionlets model which combinesadeep neural network and convolution detection for accurate object detection. The Regionlets applies an end-to-end trainable deep learning framework for modelling object deformation and multiple aspect rations. The region selection method provides guidance to select feature for bounding box region. The Regionlet learning modules focus on selecting local features and transform to alleviate the effect of appearance variation. Datasets such as PASCAL VOC and Microsoft COCO were used to evaluate the efficiency of the model. The results show that the Deep Regionlets model has abetterperformance than RetinaNet and Mask R-CNN method. The overfitting problem needs to be solved to increase the efficiency of the model.

Denget al. [24] proposed Concatenated ReLU and Inception module in a Faster RCNN method for the detection of objectsofdifferent sizes in remote sensing images. The developed method increases the reception field size variety and is suitable for multi-class object detection. Two sub-networks such as Multi-Scale Object Proposal Network (MS-OPN) and Accurate Object Detection Network (AODN) are used for object detection. The image blocks are cropped and augment the image with rotation and re-sampling for training the network to detect the large-scale remote sensing images. The Google Earth remote sensing dataset of NWPU VHR-10 was used to evaluate the efficiency of MS-OPN method for object detection in remote sensing. The analysis shows that the MS-OPN method has ahigher efficiency in detecting the object with various scale variation in remote sensing images. The overfitting in the model needs to be solved and deep features of rotation invariant need to be added for object detection.

Ding, et al. [25] proposed a VGG16-Net framework of CNN to reduce the computational time of object detection in remote sensing. The fully convolutional neural network is applied in the Faster RCNN and this reduces the memory requirement of the method as well as the computational time. The dilated convolutional layer is applied to detect the dense object in remote sensing and bootstrapping strategy is applied in Faster RCNN to detect the smaller object. The computational time of the developed method is reduced, and the precision of the detection is also increased. The detection ability of the model needs to be improved and the overfitting of the model needs to be reduced.

Longet al. [26] applied the regional proposal method to detect object in the input high resolution remote sensing images. The CNN model is applied to extract the generic image features from local image of regions. Bounding box regression-based score in a non-maximum suppression is used to optimize the bounding box region and improve the detection performance. The developed method has a higher performance in object detection in remote sensing. The developed method has lower efficiency in detecting small objects in remote sensing.

Li et al. [27] applied RPN with a local-contextual feature fusion method for object detection in remote sensing. Multi-angle anchors based on conventional multi-scale wereapplied to analyzethe characteristics of the geospatial object. A double channel feature fusion method wasapplied to learn the contextual and local region of the image to overcome the ambiguity problem. In the final layer, two kinds of features were combined to provide apowerful joint representation. The publicly available dataset was used to evaluate the performance of the developed model and shows the considerable performance. The model efficiency in detecting the smaller object in the image is low and the false positive rate needs to be reduced.

MOST WIEDZY Downloaded from mostwiedzy.pl

Linet al. [28] applied a faster R-CNN method using the squeeze and excitation mechanism to improve the performance detection in Synthetic Aperture Radar (SAR) image. A multi-scale feature map based on ImageNet pre-trained VGG network was used to provide multi-scale feature map. The scale vector is applied to recalibrate the sub feature maps to suppress the redundant feature map. The analysis on Sentinel-1 images shows ahigher performance in the detection compared to existing method. The overfitting problem in the second-stage classification needs to be reduced for efficient detection.

Cui et al. [29] applied Dense Attention Pyramid Network (DAPN) for the ship detection method in SAR images. The abundant features containing the resolution and semantic information are extracted from multi-scale ship detection. The salient features are integrated with global unblurred features to improve the accuracy intheSAR images. The analysis shows that the DAPN method for multi-scale ships in various scales in various SAR images has high performance compared to existing method. Cascade object detection needs to be applied to improve the learning of feature maps.

Problem Definition and Solution

Various CNN models have been developed for object detection to increase the efficiency. Many existing methods have the limitations of an overfitting problem and low efficiency in detecting small objects. Some methods involve enlarging the image for small object detection and can be localized more easily, while few focus on adding up-sampled high-level features into low-level features to enhance the small object representation. This enlarging method requires more memory and computation time and some of the refinement methods do not fully account for the samples' diversity, thus leaving more precision location and small objects. The object detection in remote sensing images is a challenging task due to the presence of smaller objects and because existing methods have lower efficiency in localizing small objects. The standard IoU threshold value of CNN based methods is 0.5, which leads to noisy detection and a degradation of the performance of some existing methods for a larger threshold. A major problem of the high-quality detector is overfitting due to a vanishing sample for large threshold.

Differ from Existing Faster R-CNN model: The faster R-CNN model using thesqueeze and excitation method [28] has the limitation of the second-stage classification. The DAPN with R-CNN [29] learning performance is affected by the small object and high loss in the activation function. The proposed AAF-Faster R-CNN method applies a linear combination of three activation function to improve the learning. The proposed AAF-Faster R-CNN uses the positive and negative reward update in the feature map to solve the problem of overfitting.

Solutions: This research proposes Fourier series and the linear activation method for the loss value analysis; the loss value is used in the cascade step to the improve thelearning process of the object. The Fourier series and linear activation method havethe advantage of better convergence that helps to solve the overfitting problem by considering positive classes in the analysis. The proposed method also has the advantage of a clear bounding value that helps to detect the small object without enhancement of image and sampling. The proposed AAF method has the advantages of better convergence and clear bounding value, which provides the relevant features that increase the performance for a higher IoU threshold.

3. Proposed Method

Object detection plays animportant role in computer vision applications such as surveillance system and vehicle identification. The existing methods have the limitations of overfitting and low efficiency in small object detection. This research applies the FLS-Faster RCNN model to increases the efficiency of small object detection. The MS COCO datasets and Pascal VOC 2007/2012 were used to evaluate the performance of the proposed FLS-Faster RCNN model. The proposed FLS-Faster RCNN model is based on ResNet-101

MOST WIEDZY Downloaded from mostwiedzy.pl

architecture. The block diagram of the FLS-Faster RCNN model in object detection is shown in Figure 1.

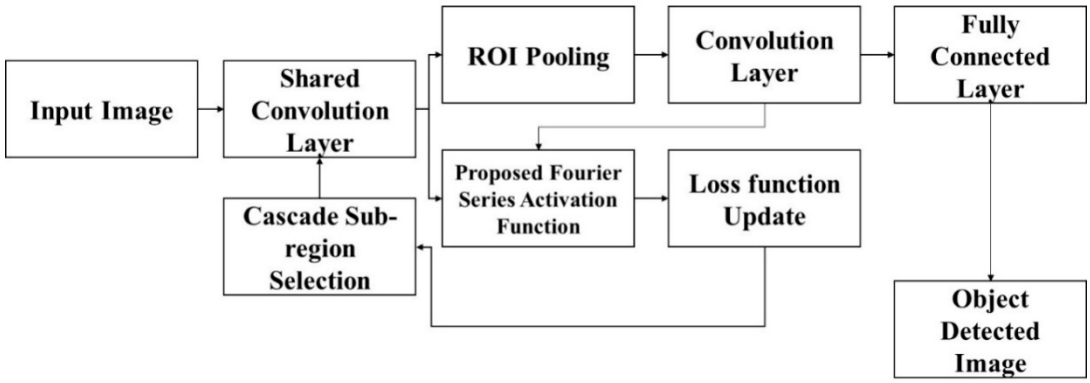

Figure 1. The proposed FLS-Faster region-based Convolution Neural Network (RCNN) method forobject detection.

From the input images, the feature maps are created and stored in the shared convolution layer based on the weight values. ROI pooling reshape the input image with arbitrary size for a size constrained fully connected layer. The convolutional layer applies the set of filters for learning the feature maps based on the activation function. The proposed Fourier series activation function updates the loss function based on the feature maps and provide rewards for cascade sub-region selection. The cascade sub-region selection updates the shared convolution layer based on the loss function from the activation function. The final detected object is provided with a boundary box based on updated feature maps. The input image is used to analyze the convolutional feature map and RPN in the detector. The regions are extracted based on the feature map and ROI pooling is performed to provide fixed size of the image. The Faster R-CNN method stores the RPN and convolutional feature maps in shared convolutional layer. The cascade object detection method provides the reward function to the Faster R-CNN method to learn the features. The AAF is used to measure the loss function to update the Faster R-CNN method.

3.1. Cascade Object Detection

The current object detection datasets rarely contain cascade attention labels and this is not easy to directly train the classifier. To overcome this problem, the Markov Decision Process (MDP) is used to develop attentional region generation. An agent is developed in MDP to make decision sequentially and ground truths of cascade are not required.

Generally, the MDP contains an actions set (A), a states set (S) and a reward function (R). The agent analyzesthe state of an environment and based on its policy function, selects an action. A states to actions probability distribution are mapped in policy function. The state present in the environment changes based on the selected action and current state. A reward signal of real value is applied to the agent to punish or award the choice [16]. The analysis process continues for a step of a finite number or until the stopping signal is received from the environment.

The parameters and details in MDP are as follows.

From the observed regions of image, the CNN feature maps are extracted are called as states. Through an ROI pooling layer, the feature maps are down sampled to a fixed size to handle multi-scale inputs.

A pre-defined hierarchy are used to cascade the attended region for five movement actions. The five candidates related to five actions in the observed region, i.e., four quarters plus a central one. Overlapped regions and non-overlapped regions are explored as two versions of the elemental design.

MOST WIEDZY Downloaded from mostwiedzy.pl

The reward function is important to build a classical reward [30] and the agent is shown in Equation (1):

$$R_\alpha(s, s') = sign(IoU(b', g) - IoU(b, g) \quad (1)$$

where b and s are the predicted box and current state, when agent selects action a, then the b and s are the predicted box and next state. The ground truth boxes are represented as g and IoU denotes the Intersection over Union. If IoU between the ground truth and predicted box is improved, a positive rewardisgiven, otherwise a negative reward is given to the actions, provided in Equation (2):

$$R_\alpha(s, s') = \begin{cases} sign(AvgIoU(B', G'_s) - AvgIoU(B, G_s)), & if\ G'_s > 0 \\ -\eta, & otherwise \end{cases} \quad (2)$$

where G_s and B are denotes the ground truth and predicted boxes on observed region and when agent select action a, G_s and B are the ground truth and predicted boxes on the next region. The $AvgIoU$ denotes the average IoU between the ground truths and predicted boxes in an image region. A positive value of reward function is return when $AvgIoU$ is increased and if $AvgIoU$ is not optimized, a negative value is returns. If there is no target box is located in the attended area, then G_s is zero and the situation is penalized by a negative reward $-\eta$.

3.2. Faster R-CNN

The Faster R-CNN key aspects are briefly described in this section. The Faster R-CNN original paper [31] is referred to for a detailed description.

In the RPN, a 3×3 convolutional layer isfollowed to pre-train the convolution layer. In the input image, the convolution layer performs mapping large spatial window or a receptive field at a center stride to reduce the dimensional feature vector. For the regression and classification of all spatial windows, two 1×1 convolutional layers are added.

The anchors are introduced in RPN to analyzevarious objects aspect ratios and scales. Each convolutional map's sliding location contains an anchor and at each spatial window center. Each anchor is related with an aspect ratio and a scale and the default settings of research [31], 3 aspect rations (1: 1, 1: 2, and 2: 1), 3 scales (1282, 2562, and 5122 pixels), leads to k = 9 at each location. The parameters of each proposal arerelated to an anchor. The most possible proposals $W \times H_k$ are present for convolutional feature map size $W \times H$. Instead of training a single regressor and k sets features extraction, the same features are present for all the sliding location to regress $k = 9$ proposals. The Stochastic Gradient Descent (SGD) is applied to train RPN in an end-to-end manner for classification and regression branches. Both RPN and Fast R-CNN modules are considered for the entire system to share convolutional layers. In this research, the approximate joint learning method is adopted for training [32]. The Fast R-CNN and RPN are trained in an end-to-end manner, independently. The Fast R-CNN input is dependent on the RPN output and this is not a trivial optimization problem [32].

An image patch x of the four coordinates ispresent in a bounding box $b = (b_x, b_y, b_w, b_h)$. The bounding box regressor $f(x,\ b)$ regress a candidate bounding box b into a target bounding box g. Minimizing the risk in a training set $(g_i,\ b_i)$, as shown in Equations (3) and (4):

$$R_{loc}[f] = \sum_i \mathcal{L}(f(x_i, b_i), g_i) \quad (3)$$

$$\mathcal{L}(a, b) = \sum_{i \in \{x,y,w,h\}} smooth_{\mathcal{L}}(a_i - b_i) \quad (4)$$

Assign an image patch x based on Function $h(x)$ to one of $M + 1$ classes, where the background is present in class 0 and detect the remaining classes of the object. The posterior distribution over classes of a $M + 1$ dimensional estimate is based on Function

MOST WIEDZY Downloaded from mostwiedzy.pl

$h(x)$, i.e.,$h_k(x) = p(y = k|x)$, where y is the class label. Consider that thetraining set (x_i, y_i) is learned by minimizing the classification risk, as shown in Equations (5) and (6):

$$R_{cls}[h] = \sum_i \mathcal{L}_{cls}(h(x_i) \times R_\alpha, y_i) \tag{5}$$

$$\mathcal{L}_{cls}(h(x), y) = -log h_y(x) \tag{6}$$

where $\mathcal{L}_{cls}(h(x), y)$ is cross-entropy loss.

3.3. Detection Network

For cascade object detection, the modified version of Faster R-CNN method is used. Based on Faster R-CNN, the coarse to fine bounding boxes are used to detect the object.

Coarse to Fine Forward: The region proposal network $B_1 = \{B_i, 1\} i \in \{1, ..., K\}$ is developed based on a first set of K object proposal from an input image. A feature map is used to extract the regions and the ROI Pooling [14] is used to pool to a fixed size. The extracted regions are applied in a network and reduced using offset transformations. A second set of K objects $B_2 = \{B_i, 2\} i \in \{1, ..., K\}$ is applied and the final set of bounding boxes B3 is developed by repeating the process. This bounding box process differs from Faster R-CNN, refinement sets overcome variationconstraints in large object scale and provide more accurate detection. In this method, first convolution feature maps are used to extract ROI pooled regions for keeping high resolution to detect small objects.

3.4. Faster R-CNN Network Training

The section denotes the tasks of network and associated loss functions. Three refinement levels $l \in \{1, 2, 3\}$ and five functions is used to minimize the loss function: $\mathcal{L}_{rpn}$, $\mathcal{L}_{ReLU}$, $\mathcal{L}_{sigmoid}$, $\mathcal{L}_{linear}$ and $\mathcal{L}_{tanh}$. $\mathcal{L}_{rpn}$ is the RPN loss function [32]. The Faster R-CNN framework [33] based on RPN learn end-to-end model. The network joint optimization based on an input image minimizes the global function, as shown in Equations (7)–(10):

$$\mathcal{L} = \mathcal{L}^1 + \mathcal{L}^2 + \mathcal{L}^3 \tag{7}$$

With

$$\mathcal{L}^1 = \mathcal{L}_{rpn} \tag{8}$$

$$\mathcal{L}^2 = \sum_i act(x)_{Fourier} \tag{9}$$

$$\mathcal{L}^3 = \sum_i act(x)_{linear} \tag{10}$$

3.4.1. Fourier Series Activated Function

Fourier series is selected for better convergence and activation function naturally satisfy the Dirichlet Fourier series conditions. The Fourier series represent any suitable activation function [34]. Dirichlet Fourier series conditions are explained as follows.

1. Integrable over a period
2. Bounded interval has bounded variation
3. Discontinuities are in finite number in any given bounded interval

At continuous points, the Fourier series converge to function, and discontinuity points havethe mean of the negative and positive limits. The activation function is continuous or non-continuous points that will make the output neuron meaningless, integrable, and havezero discontinuities. The activation function should have finite extreme points and as for the neuron, it should be stable for similar input and any given bounded interval has bounded variation for this reason. The Dirichlet Fourier series conditions satisfy the activation functions and any functions suitable for activation function can be represented in Fourier series including ReLU, Sigmoid, and the tanh activation function. The best performance can be seen in thenetwork with Fourier series as the activation function has

MOST WIEDZY Downloaded from mostwiedzy.pl

higher performance than the Sigmoid, ReLU, and tanh activation functions. Fourier series activation function satisfy the Dirichlet Fourier series conditions and Fourier series can represent any function that is suitable for activation function (e.g., ReLU, Sigmoid, and tanh). Hence, the performance of Fourier series activation function is higher than ReLU, Sigmoid, and tanh function. The positive and negative reward function is used in the cascade learning to handle the loss in the activation function.

The Fourier series can be written as Equation (11).

$$act(x) = A + \sum_{n=1}^{\infty} (a_n \cos(nwx) + b_n \sin(nwx)) \tag{11}$$

where parameters A, ω, a_n, and b_n are trainable parameters. The Fourier series rank does not need to be high. In this method, this is fixed to 5 (i.e., n = 1, 2, ..., 5).

The gradient descent algorithms is used to train the activation function. The gradient descent algorithm normalizes the range of the activation function and helps to compute the activation function easily.The gradient descent method normalizes the value of three activation function into the range of 0 to 1 to update the loss function. The gradient of such activation function is given in Equations (12)–(16):

$$\frac{\partial act(x)}{\partial A} = 1 \tag{12}$$

$$\frac{\partial act(x)}{\partial a_n} = -nwsin(nwx) \tag{13}$$

$$\frac{\partial act(x)}{\partial b_n} = nwcos(nwx) \tag{14}$$

$$\frac{\partial act(x)}{\partial w} = \sum_{n=1}^{\infty} (a_n nxcos(nwx) - b_n nxsin(nwx)) \tag{15}$$

$$= \sum_{n=1}^{\infty} nx(a_n \cos(nwx) - b_n \sin(nwx)) \tag{16}$$

If $\sin(n\omega x)$, $\cos(n\omega x)$ and x is stored in the local, then the gradient computing task is greatly simplified and perform several multiplies. In training, memory and time complexity are both $O(n)$ and n is a very small integer which represents Fourier series rank.

3.4.2. Linear Combination of Activated Function

The candidate functions arethe process of linearly combining multiple functions and this is another way to train the activation function [34]; such activation functions in CNNs are called LC-CNN. The dot product of a hyperspace unit vector and vector of various activation functions are used for such activation function, as shown in Equations (17)–(19):

$$act(x) = \frac{\sum_{i=1}^{n} w_i act_i(x)}{\sum_{i=1}^{n} w_i} \tag{17}$$

$$= [w_1,\ w_2, \ldots, w_{n-1}, w_n] \times [act_1(x), act_2(x), \ldots, act_{n-1}(x), act_n(x)]^T \tag{18}$$

$$= \frac{W \times Acts}{||W||} \tag{19}$$

where $||W|| = \sum_{i=1}^{n} w_i$.

A linear combination is used for two reasons. One reason is thatsuch a combination is easily converted into simple activation functions based on letting weight vector W to be one-hot. This means that such method will not have lower perform than single activation

MOST WIEDZY Downloaded from mostwiedzy.pl

function. Another advantage is the gradient of these functions are easy to compute. The gradients are given in Equations (20)–(22).

$$\frac{\partial act(x)}{\partial(w_k)} = \frac{act_k(x)\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i act_i(x)}{\left(\sum_{i=1}^{n} w_i\right)^2} \tag{20}$$

$$= \frac{act_k(x)\sum_{i=1}^{n} w_i - act(x)\sum_{i=1}^{n} w_i}{\left(\sum_{i=1}^{n} w_i\right)^2} \tag{21}$$

$$= \frac{act_k(x) - act(x)}{||W||} \tag{22}$$

The sum of weights and each activation function output is stored to easily compute the gradient. In training, the efficiency of gradient decent training is considerable. In this method, the activation function is the combination of ReLU, tanh, Sigmoid, and linear.

4. Experimental Design

Various CNN based models have been developed for the object detection and shows considerable performance. The proposed AAF-Faster RCNN method have been applied for object detection. The experimental design of the AAF-Faster RCNN model is explained in this section.

Datasets: The Pascal 2007/2012 [35] and Microsoft (MS) COCO dataset [36] contains 20 and 80 objects. The Pascal 2007/2012 and MS COCO datasets were used to estimate the proposed model efficiency. From the Pascal VOC 2007/2012 dataset, 16,551 images were applied for training and 4952 images were applied for testing. From MS COCO dataset, 118 k images were applied for training and 20 k images were applied for testing. The NWPU VHR-10 dataset consists of 565 color images collected from Google Earth and has objects such as Airplanes, ships, baseball diamonds, storage tanks, ground track fields, basketball courts, tennis courts, bridges, and vehicles. In NWPU VHR-10 dataset evaluation, 60% is used for training and 40% is used for testing. Generally, training-testing of 80–20 was applied to estimate the performance of the object detection model. The training-testing of 60–40 is used to evaluate the efficiency of the proposed AAF-Faster RCNN mode.

Metrics: The mean Average Precision (AP) and Average Precision (AP) were used to estimate the efficiency of the proposed AAF-Faster RCNN. The sensitivity is measured using analytic tool [37] and computational time were also analyzed in the model. The AP is measured for standard VOC IoU (0.5) andanalyzed for IoU (0.75). The proposed AAF-Faster RCNN method is analyzed to detect the small, medium, and large objects.

Parameter Settings: The proposed AAF-Faster RCNN method is based on ResNet-101 architecture. The learning rate is set as 1×10^{-4} and η is set as 0.3. The analysis shows there is no performance gain after 20 epoch and training is applied as 20 epochs. The image size of 512×512 is used to train the proposed method. The NWPU VHR-10 dataset consists of 60% training and 40% testing in the evaluation.

System Requirement: The proposed AAF-Faster RCNN method is implemented inthe system and consists of an Intel i7 processer with 16 GB RAM and 4 GB graphic card. The proposed AAF-Faster RCNN method is developed and tested on Python 3.7. The proposed and existing method is tested on the same dataset and in the same environment.

5. Experimental Results

Various CNN based models were developed for object detection and achievedconsiderable performance in detection. The existing methods of object detection have the limitation of overfitting and low efficiency in small object detection. This research applies the AAF-Faster RCNN model to increase the object detection efficiency. The AAF has abetter convergence and has clear bounding variance for the analysis. The developed model is based on the ResNet-101 structure. The Pascal VOC 2007, Pascal VOC 2012 and

MOST WIEDZY Downloaded from mostwiedzy.pl

Microsoft COCO dataset were used to analysis the performance. The detailed description of the AAF-Faster RCNN model performance is provided in this section.

The output samples of AAF-Faster RCNN samples on PASCAL VOC 2007 dataset is shown in Figure 2: (a) plane, (b) swarm and (c) Two dogs. The proposed AAF-Faster RCNN method has higher performance in object detection and provides ahigher performance in multiple object detection.

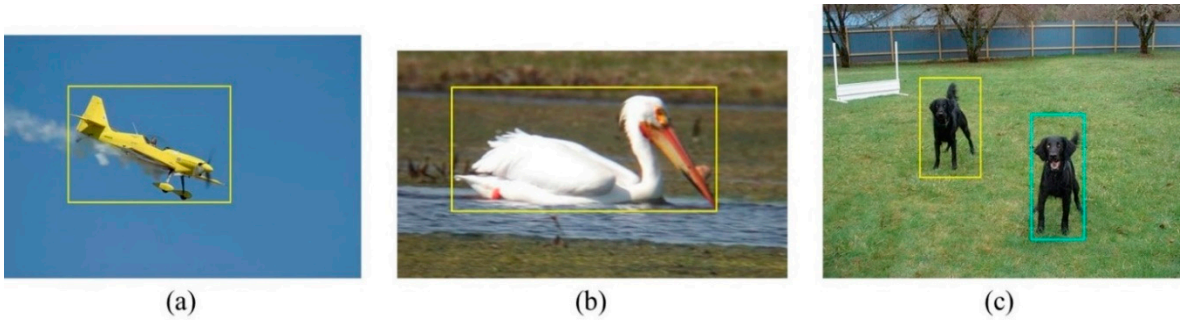

Figure 2. (**a**) plane, (**b**) swarm and (**c**) Two dogs: The output samples of proposed AAF-Faster RCNN method on PASCAL VOC 2007 dataset.

The output samples of AAF-Faster RCNN method on PASCAL VOC 2012 dataset are shown in Figure 3a–c. The proposed AAF-Faster RCNN method has higher efficiency in the object detection of the overlap region due to the fact that the AAF-Faster RCNN method has clear bounding values for the object and better convergence, as shown in Figure 3a. The proposed AAF-Faster RCNN model has ahigher efficiency in detecting the small object due to cascade method is applied in the proposed AAF-Faster RCNN method to analyze the sub-region in the image, as shown in Figure 3b. The existing methods has a lower performance in detecting the small object. The existing method has lower efficiency in detecting the overlap region in the image. To overcome this problem, the proposed AAF-Faster RCNN method has clear bounding values based on the loss function.

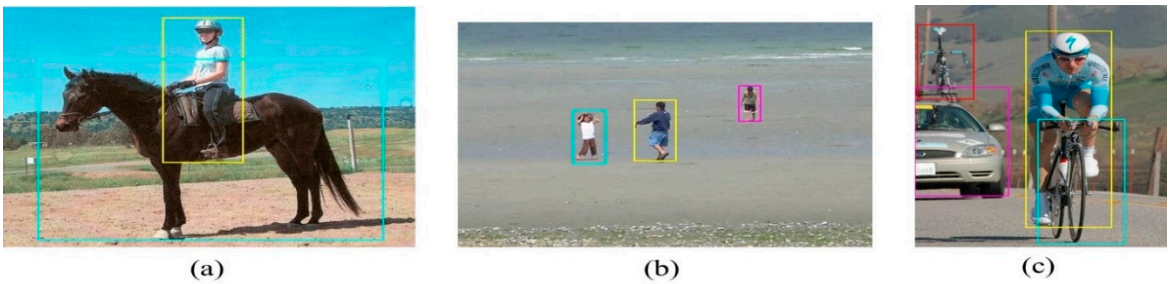

Figure 3. (**a**) Boy on the horse, (**b**) three children in the beach, and (**c**) Cyclist: The output samples of proposed AAF-Faster RCNN method on PASCAL VOC 2012 dataset.

The output samples of AAF-Faster RCNN method on MS COCO dataset, as shown in Figure 4a–c. The proposed AAF-Faster RCNN method has the advantage of better convergence and clear bounding values based on loss function. The proposed AAF-Faster RCNN method has a higher performance in detecting the overlap object and small object, as shown in the Figure 4c.

MOST WIEDZY Downloaded from mostwiedzy.pl

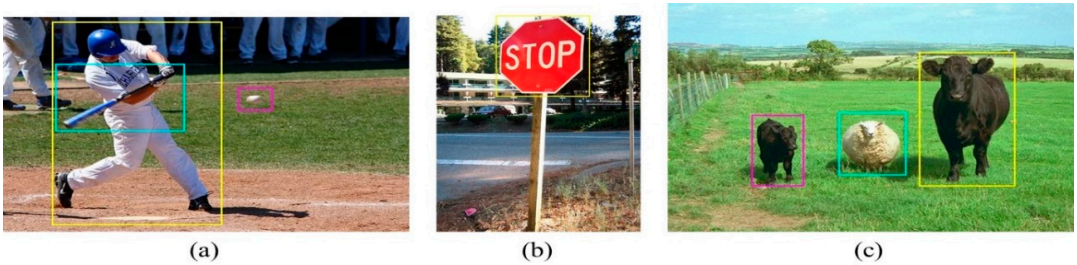

Figure 4. (**a**) Baseball, (**b**) Stop sign, and (**c**) Domestic animals: The output samples of proposed AAF-Faster RCNN method on MS COCO dataset.

The output samples of proposed AAF-Faster RCNN method on PASCAL VOC 2007, PASCAL VOC 2012, and MS COCO dataset are shown in Figures 2–4, respectively. The proposed AAF-Faster RCNN method has the ability to detect the small object, as shown in Figure 3b. The AAF-Faster RCNN method has the abilityto effectively detect the overlap region of the image, as shown in Figures 3a–c and 4a. The proposed AAF-Faster RCNN method has the advantage of better convergence and clear bounding value and this helps to detect the small object in the image and solves the overfitting problem.

The sample images of proposed AAF-Faster RCNN method for object detection in Remote sensing are shown in Figure 5a,b. The efficiency of the proposed AAF-Faster RCNN method is high in the object detection in remote sensing images. Figure 5a shows that the proposed AAF-Faster RCNN method has a higher efficiency in detecting the airplane and Figure 5b shows that the proposed AAF-Faster RCNN method has ahigher efficiency in detecting the ship. The proposed AAF-Faster RCNN method has the advantage of better convergence and clear bounding.

Figure 5. (**a**) Airplanes, and (**b**) Ships: The proposed AAF-Faster RCNN method sample object detection in Remote sensing images.

The detection error of proposed AAF-Faster RCNN method in PASCAL 2012 and MS COCO images are shown in Figure 6. In Figure 6a, one boat is missed in the classification due to the presence of object in the narrow pattern and in Figure 6b, instead of two hot dogs one hot dog is detected due to the presence of a similar pattern.

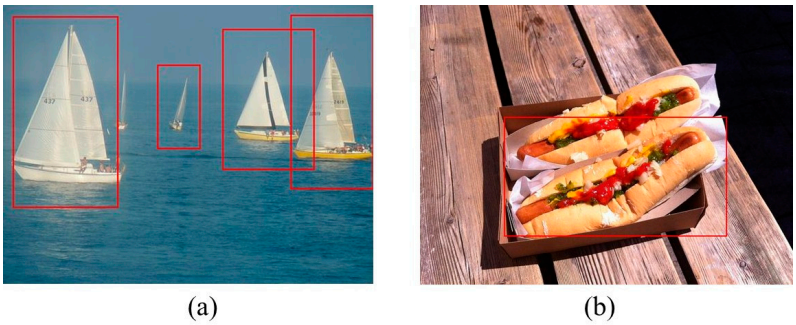

Figure 6. (**a**) Boat, and (**b**) Hot Dogs: Detection error in PASCAL 2012 and MS COCO.

The detection error of the proposed AAF-Faster RCNN in NWPU VHR-10 dataset method is shown in the Figure 7a,b. In the Figure 7a, the airplane is falsely detected due to the color features and small object of the image. In Figure 7b, the small airplane is undetected due to the presence of small object in the image.

(a) (b)

Figure 7. (**a**) Street, and (**b**) Plane runway: Detection error in the NWPU VHR-10 dataset.

5.1. Performance Analysis on Pascal VOC 2007 Dataset

The proposed AAF-Faster RCNN method is tested on the Pascal VOC 2007 dataset and compared with other models, as shown in Table 1. The mAP metric is measured for various models in the Pascal dataset and compared with existing methods.

Table 1. Performance analysis of the Pascal VOC 2007 dataset.

Method	Structure	mAP
Faster w ResNet [38]	ResNet-101	76.4
R-FCN [39]	ResNet-101	80.5
R-FCN w Deformable CNN	ResNet-101	82.6
AOL Net	AlexNet	46.1
Attention Net [40]	VGG	70.7
Faster [31]	VGG	73.2
PAT-Faster [16]	VGG	75.9
YOLOv2 352 [41]	Darknet	73.7
PAT-YOLOv2 352 [16]	Darknet	77.1
YOLOv2 544 [41]	Darknet	78.6
PAT-YOLOv2 544 [16]	Darknet	80.1
SSD300 [42]	VGG	77.2
DSSD321	ResNet-101	78.6
PAT-SSD300 [16]	VGG	80.5
SSD512 [42]	VGG	79.8
DSSD513	ResNet-101	81.5
PAT-SSD512 [16]	VGG	81.7
AAF-Faster RCNN	ResNet-101	83.1

The Fourier series linear combination of activation function is used in the proposed AAF-Faster RCNN model to improve the efficiency of the detection. The proposed AAF-Faster RCNN model uses the cascade detection to increase the learning rate and effectively

detect the small object. The existing methods [16,41,42] have overfitting problem in cascaded object detection. The proposed model overcame the overfitting problem using better convergence and clear bounded variation in the analysis. The models with various structure are analyzed in the Pascal VOC 2007 dataset and compared with the proposed AAF-Faster RCNN model. The evaluation shows that the proposed AAF-Faster RCNN model has the higher mAP value compared to other existing models. The proposed AAF-Faster RCNN method also has abetter convergence and clear bounded variation in the analysis. The proposed AAF-Faster RCNN method achieves the mAP of 83.1% and existing PAT-SSD512 method achieves 81.7% mAP in the Pascal VOC 2007 dataset. The proposed AAF-Faster RCNN model is trained and tested with the image of 512 × 512 size and existing PAT method also trained and tested with the same image size.

5.2. Performance Analysis of PASCAL VOC 2012 Dataset

The proposed AAF-Faster RCNN method is analyzedinthe PASCAL VOC 2012 dataset and compared with other models, as shown in Table 2. The same parameter settings of PASCAL VOC 2007 dataset areused in this analysis.

Table 2. Performance analysis in the PASCAL VOC 2012 dataset.

Method	Faster w ResNet [38]	R-FCN [39]	Attention Net [40]	Faster [31]	PAT-Faster [16]	YOLO-v2 544 [41]	PAT-YOLO-v2 544 [16]	PAT-SSD300 [16]	PAT-SSD512 [42]	AAF-Faster RCNN
mAP	73.8	77.6	65.6	70.4	74	73.4	74.5	78.2	80.6	81.11
aero	86.5	86.9	79.1	84.9	84.2	86.3	86.9	89.9	91.3	92.1
bike	81.6	83.4	68.9	79.8	80.6	82	82.8	86.7	88.2	89.2
bird	77.2	81.5	65.5	74.3	75.5	74.8	75.6	77	81	82.2
boat	58	63.8	52.3	53.9	59.6	59.2	61	64.4	68.6	69.1
bottle	51	62.4	55.9	49.8	56.5	51.8	55.7	54.1	62.7	63.5
bus	78.6	81.6	73.5	77.5	79.6	79.8	79.6	84.2	85.2	85.4
car	76.6	81.1	76.5	75.9	81.5	76.5	78.9	82.5	86.7	87.2
cat	93.2	93.1	79.1	88.5	89.4	90.6	89.9	91.6	92.5	92.7
chair	48.6	58	42.3	45.6	54.3	52.1	52.1	61.5	64	64.2
cow	80.4	83.8	70.9	77.1	78.7	78.2	82.9	84.3	84.9	85.1
table	59	60.8	42.7	55.3	57.9	58.5	57.8	65.1	68.1	68.7
dog	92.1	92.7	76.8	86.9	86.7	89.3	87.1	89.9	90.4	90.5
horse	85.3	86	73.9	81.7	83.1	82.5	86.7	86.9	88.5	88.7
mbike	84.8	84.6	73.3	80.9	84	83.4	84.9	87.7	89.1	89.3
person	80.7	84.4	79.2	79.6	83.6	81.3	83.5	85.3	87.6	88.2
plant	48.1	59	48.1	40.1	52.4	49.1	49.8	57	58.6	58.7
sheep	77.3	80.8	74.1	72.6	78	77.2	78.5	82.7	86.2	87.1
sofa	66.5	68.6	44.9	60.9	64.4	62.4	62.9	72.1	73	74.5
train	84.7	86.1	73.7	81.9	81.9	83.8	83.8	87.8	86.7	87.8
tv	65.6	72.9	62.2	68.8	68.8	68.7	70.4	74.2	77.4	78.1

The proposed AAF-Faster RCNN model uses the reward function to improve the learning based on the Fourier series and linear combination of activation function. The cascade object detection method is used in proposed AAF-Faster RCNN model to effectively detect the small object. The existing models [16,42] have the limitation of notfully accounting for the samples' diversity and they have lower efficiency. The proposed AAF-Faster RCNN model considers the samples' diversity based on cascade object detection. The performance analysis of proposed AAF-Faster RCNN model is analyzed with mAP on the PASCAL VOC 2012 dataset, as shown in Table 2. The analysis shows that the AAF-Faster RCNN model has ahigher mAP value compared to the existing methods. The PAT-SSD 512 method has the second highest performance in object detection. The AAF-Faster CNN method has the advantage of better convergence and clear bounded variation in the analysis. The proposed

MOST WIEDZY Downloaded from mostwiedzy.pl

AAF-Faster RCNN model is based on the ResNet-101 structure for the object detection. The proposed AAF-Faster RCNN model has amAP of 81.11% and state of art method of PAT-SSD512 method has 80.6% mAP value in PASCAL VOC 2012 dataset. The state of art PAT method uses discrete set of value for realization and proposed AAF method integrate over the period of time for clear bounding variance. The analysis clearly shows that the proposed AAF-Faster RCNN model has ahigher performance compared to other models.

5.3. Performance Analysis on Microsoft COCO Dataset

The proposed AAF-Faster RCNN model is analyzed in the MS COCO dataset and compared with existing models, as shown in Table 3. The MS COCO dataset consists of small objects and this is challenging for the object detection model. The proposed AAF-Faster RCNN model and existing PAT with YOLOv2 are trained in same parameter settings. The dataset is divided in three kinds, namely: Small (S) i.e., area less than 322, Medium (M) i.e., area between 32^2 and $96^{2,}$ and Large (L) i.e., area greater than 96^2 pixels.

Table 3. Performance analysis in theMS COCO dataset.

Method	Structure	AP IoU				AP Area		
		Time	0.5:0.95	0.5	0.75	S	M	L
Faster+++ [38]	ResNet-101	336	3.49	55.7	37.4	15.6	38.7	50.9
R-FCN [39]	ResNet-101	110	29.9	51.9	-	10.8	32.8	45
Mask R-CNN	ResNext-101-FPN	210	37.1	60	39.4	16.9	39.9	53.5
Faster [31]	VGG	147	21.9	42.7	-	-	-	-
PAT-Faster [16]	VGG	160	26.9	47	27.9	11.9	31	37.8
YOLOv2 544 [41]	Darknet	25	21.6	44	19.2	5	22.4	35.5
PAT-YOLOv2 544 [16]	Darknet	50	27.4	50.7	27	11.8	30.9	36.4
SSD300 [42]	VGG	22	25.1	43.1	25.8	6.6	25.9	42.4
PAT-SSD 300 [16]	VGG	37	28.9	48.7	30	11.2	31.8	42.3
SSD 512 [42]	VGG	53	28.8	48.5	30.3	10.9	31.8	43.5
PAT-SSD 300 [16]	VGG	85	31.9	53.1	33.5	16.2	36.1	43
PAT-SSD 800 [16]	VGG	110	37.1	57.7	39.8	22.5	41.5	50.7
AAF-Faster RCNN	ResNet-101	82	37.2	58.1	40.2	23.1	43.4	51.2

The proposed AAF-Faster RCNN model uses the cascade learning method to improve the learning model. The Fourier series and linear combination of activation function areused in the proposed AAF-Faster RCNN model is used to improve the efficiency of the object detection. The existing detection models [16,42] have the limitation of overfitting due to the samples vanishing for a large threshold. The proposed AAF-Faster RCNN model uses reward function to improve the learning and solve the overfitting problem. The performance analysis of Proposed AAF-Faster RCNN method is analyzed in the MS COCO dataset, as shown in Table 3. The analysis shows that the AAF-Faster RCNN method has higher mAP value compared to existing method in cascade object detection. The AAF-Faster RCNN method has the advantage of better convergence and clear bounding variance. The proposed AAF-Faster RCNN method has lower computation time compared to the state of art PAT SSD 300 method. The proposed AAF-Faster RCNN method uses a 512 $\times$ 512 image size and PAT SSD method has a 300 $\times$ 300 image size. The proposed AAF-Faster RCNN method has lower computation time due to integral of value and higher mAP value due to clear bounding variance. The proposed AAF-Faster RCNN method has a higher mAP value for various IoU and various area sizes. The proposed AAF-Faster

MOST WIEDZY Downloaded from mostwiedzy.pl

RCNN method has the mAP of 43.4% in a medium-sized area and existing PAT-SSD 800 method has 41.5% mAP in the MS COCO dataset.

5.4. Performance Analysis on Small Object Detection

The proposed AAF-Faster RCNN method and state of art PAT-SSD 300 is measured with sensitivity using an analysis tool [37]. The Pascal VOC challenging classes such as boat, bird, chair, bottle, plant and table were used to analyzethe performance of the proposed AAF-Faster RCNN method, as shown in Table 4.

Table 4. Performance analysis forsmall object detection.

	PAT-SSD300 [16]			**AAF-Faster RCNN**		
	Small	**Medium**	**Large**	**Small**	**Medium**	**Large**
Bird	0.83	0.87	0.91	0.85	0.89	0.92
Boat	0.79	0.8	0.92	0.8	0.83	0.93
Bottle	0.57	0.64	0.75	0.59	0.66	0.76
Chair	0.63	0.77	0.73	0.65	0.78	0.74
Table	0.78	0.94	0.96	0.79	0.95	0.97
Potted plant	0.56	0.62	0.72	0.56	0.63	0.73
tv monitor	0.66	0.83	0.9	0.67	0.84	0.91

The proposed AAF-Faster RCNN model applies the Fourier series and linear combination of activation function to effectively detect small objects in the dataset. The cascade object detection improves the learning performance of the proposed AAF-Faster RCNN model. The existing models [16] do not fully account for thesamples' diversity, thus leaving more precision location and small objects. The proposed AAF-Faster RCNN model uses loss function to consider samples diversity that improves the efficiency. The proposed AAF-Faster RCNN is based on ResNet-101 structure helps to detect the object with considerable computation time and it is clear that AAF activation method boosts the performance of the Faster RCNN model. The proposed AAF-Faster RCNN method is analyzedforsmall object detection in Pascal VOC dataset, as shown in Table 4. The analysis shows that the proposed AAF-Faster RCNN method is more stable compared to the PAT-SSD 300 method. The proposed AAF-Faster RCNN method has the advantage of better convergence and clear bounding variance. The sensitivity of the proposed AAF-Faster RCNN method is 84% for Medium area and existing PAT-SSD300 method has 83% sensitivity forsmall object detection.

Performance Analysis of Object Detection in Remote Sensing Images

The efficiency of the proposed AAF-Faster RCNN model is evaluated in the remote sensing images to detect objects. The NWPU VHR-10 dataset was used to estimate the performance of the proposed AAF-Faster RCNN model. The proposed AAF-Faster RCNN model and existing model mAP were measured for object detection in the NWPU VHR-10 dataset, as shown in Table 5. The proposed AAF-Faster RCNN model has ahigher mAP compared to existing methods in object detection. The proposed AAF-Faster RCNN model has the advantage of better convergence and clear bounding value. The MS-OPN with fuse feature map has the second highestmAP value due to its capacity to detect the small object.

In remote sensing images, the objects are present in various scale and the existing models have lower efficiency due to poor localization. The proposed AAF-Faster RCNN model uses Fourier series and the linear activation function to improve the localization. The proposed AAF-Faster RCNN model uses the cascade learning method to improve the efficiency of small object detection based on reward function. The proposed AAF-Faster RCNN model has mAP value of 95.21% and existing MS-OPN with fuse method has value of 94.87%mAP in NWPU VHR-10 dataset.

MOST WIEDZY Downloaded from mostwiedzy.pl

Table 5. Performance analysis of object detection in remote sensing.

Methods	mAP (%)
Fast—RCNN	72.32
Faster—RCNN	76.42
YOLO1	66.72
YOLO2	78.73
R-FCN	92.8
SSD	89.39
MS-OPN	91.52
MS-OPN with Fuse	94.87
AAF-Faster RCNN	95.21

6. Conclusions

Object detection is animportant task in the computer vision application such as surveillance system, vehicle identification, etc. Various CNN based models have been developed for the object detection and have achieved considerable performance. However, the existing methods suffers from the overfitting problem and lower efficiency in small object detection. This research proposed the AAF-Faster RCNN model to increases the efficiency in object detection. The proposed AAF-Faster RCNN model has the advantage of better convergence and clear bounding variance. The datasets such as Pascal VOC 2007/2012 and MS COCO were used to analyzethe efficiency of the proposed AAF-Faster RCNN method. The proposed AAF-Faster RCNN method is evaluated in the NWPU VHR-10 dataset to estimate itsperformance forremote sensing images. The analysis shows that the proposed AAF-Faster RCNN model has higher performance compared to the existing modelsin terms of AP. The AAF-Faster RCNN model has ahigher mAP of 83.1% and existing PAT-SSD512 method has 81.7%mAP in the Pascal VOC 2007 dataset.

Author Contributions: The paper investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by S.N.S. and M.J.P.P. The paper conceptualization, methodology, software, validation, and formal analysis have been done by B.D.P. Supervision, project administration, and final approval of the version to be published wereconductedby A.S. and C.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Pascal VOC 2007 dataset is available at http://host.robots.ox.ac.uk/pascal/VOC/voc2007/, PASCAL VOC 2012 dataset is available at http://host.robots.ox.ac.uk/pascal/VOC/voc2012/, MS COCO dataset is available at https://cocodataset.org/, and NWPU VHR-10 dataset is available at https://github.com/chaozhong2010/VHR-10_dataset_coco.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shijila, B.; Tom, A.J.; George, S.N. Simultaneous denoising and moving object detection using low rank approximation. *Future Gener. Comput. Syst.* **2019**, *90*, 198–210.
2. Cuevas, C.; Yáñez, E.M.; García, N. Labeled dataset for integral evaluation of moving object detection algorithms: LASIESTA. *Comput. Vis. Image Underst.* **2016**, *152*, 103–117. [CrossRef]
3. Sengar, S.S.; Mukhopadhyay, S. Moving object detection using statistical background subtraction in wavelet compressed domain. *Multimed. Tools Appl.* **2020**, *79*, 5919–5940. [CrossRef]

MOST WIEDZY Downloaded from mostwiedzy.pl

4. Ray, K.S.; Chakraborty, S. Object detection by spatio-temporal analysis and tracking of the detected objects in a video with variable background. *J. Vis. Commun. Image Represent.* **2019**, *58*, 662–674. [CrossRef]
5. Chen, Z.; Wang, R.; Zhang, Z.; Wang, H.; Xu, L. Background–foreground interaction for moving object detection in dynamic scenes. *Inf. Sci.* **2019**, *483*, 65–81. [CrossRef]
6. Boukhriss, R.R.; Fendri, E.; Hammami, M. Moving object detection under different weather conditions using full-spectrum light sources. *Pattern Recognit. Lett.* **2020**, *129*, 205–212. [CrossRef]
7. Lu, R.; Ma, H.; Wang, Y. Semantic head enhanced pedestrian detection in a crowd. *Neurocomputing* **2020**, *400*, 343–351. [CrossRef]
8. Lin, H.; Zhou, J.; Gan, Y.; Vong, C.M.; Liu, Q. Novel up-scale feature aggregation for object detection in aerial images. *Neurocomputing* **2020**, *411*, 364–374. [CrossRef]
9. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; et al. Deep high-resolution representation learning for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**. [CrossRef]
10. Shuai, H.; Liu, Q.; Zhang, K.; Yang, J.; Deng, J. Cascaded regional spatio-temporal feature-routing networks for video object detection. *IEEE Access* **2017**, *6*, 3096–3106. [CrossRef]
11. Li, J.; Liang, X.; Li, J.; Wei, Y.; Xu, T.; Feng, J.; Yan, S. Multistage object detection with group recursive learning. *IEEE Trans. Multimed.* **2017**, *20*, 1645–1655. [CrossRef]
12. Yang, D.; Zou, Y.; Zhang, J.; Li, G. C-RPNs: Promoting object detection in real world via a cascade structure of Region Proposal Networks. *Neurocomputing* **2019**, *367*, 20–30. [CrossRef]
13. Sun, X.; Wu, P.; Hoi, S.C. Face detection using deep learning: An improved faster RCNN approach. *Neurocomputing* **2018**, *299*, 42–50. [CrossRef]
14. Pang, Y.; Cao, J.; Li, X. Cascade learning by optimally partitioning. *IEEE Trans. Cybern.* **2016**, *47*, 4148–4161. [CrossRef] [PubMed]
15. Bria, A.; Marrocco, C.; Molinara, M.; Tortorella, F. An effective learning strategy for cascaded object detection. *Inf. Sci.* **2016**, *340*, 17–26. [CrossRef]
16. Liu, S.; Huang, D.; Wang, Y. Pay Attention to Them: Deep Reinforcement Learning-Based Cascade Object Detection. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *31*, 2544–2556. [CrossRef] [PubMed]
17. Cai, Z.; Vasconcelos, N. Cascade R-CNN: High quality object detection and instance segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**. [CrossRef]
18. Cevikalp, H.; Triggs, B. Visual object detection using cascades of binary and one-class classifiers. *Int. J. Comput. Vis.* **2017**, *123*, 334–349. [CrossRef]
19. Zhong, Q.; Li, C.; Zhang, Y.; Xie, D.; Yang, S.; Pu, S. Cascade region proposal and global context for deep object detection. *Neurocomputing* **2020**, *395*, 170–177. [CrossRef]
20. Zhu, Y.; Ma, C.; Du, J. Rotated cascade R-CNN: A shape robust detector with coordinate regression. *Pattern Recognit.* **2019**, *96*, 106964. [CrossRef]
21. Dai, X. HybridNet: A fast vehicle detection system for autonomous driving. *Signal Process. Image Commun.* **2019**, *70*, 79–88. [CrossRef]
22. Zou, X.; Zhou, L.; Li, K.; Ouyang, A.; Chen, C. Multi-task cascade deep convolutional neural networks for large-scale commodity recognition. *Neural Comput. Appl.* **2020**, *32*, 5633–5647. [CrossRef]
23. Xu, H.; Lv, X.; Wang, X.; Ren, Z.; Bodla, N.; Chellappa, R. Deep regionlets: Blended representation and deep learning for generic object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**. [CrossRef] [PubMed]
24. Deng, Z.; Sun, H.; Zhou, S.; Zhao, J.; Lei, L.; Zou, H. Multi-scale object detection in remote sensing imagery with convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 3–22. [CrossRef]
25. Ding, P.; Zhang, Y.; Deng, W.J.; Jia, P.; Kuijper, A. A light and faster regional convolutional neural network for object detection in optical remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2018**, *141*, 208–218. [CrossRef]
26. Long, Y.; Gong, Y.; Xiao, Z.; Liu, Q. Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2486–2498. [CrossRef]
27. Li, K.; Cheng, G.; Bu, S.; You, X. Rotation-insensitive and context-augmented object detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 2337–2348. [CrossRef]
28. Lin, Z.; Ji, K.; Leng, X.; Kuang, G. Squeeze and excitation rank faster R-CNN for ship detection in SAR images. *IEEE Trans. Geosci. Remote Sens. Lett.* **2018**, *16*, 751–755. [CrossRef]
29. Cui, Z.; Li, Q.; Cao, Z.; Liu, N. Dense attention pyramid networks for multi-scale ship detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8983–8997. [CrossRef]
30. Caicedo, J.C.; Lazebnik, S. Active object localization with deep reinforcement learning. In Proceedings of the IEEE International Conference on Computer Vision, Las Condes, Chile, 11–18 December 2015; pp. 2488–2496.
31. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2015; pp. 91–99.
32. Jiang, H.; Learned-Miller, E. Face detection with the faster R-CNN. In Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition, Washington, DC, USA, 30 May–3 June 2017; pp. 650–657.
33. Chabot, F.; Chaouch, M.; Rabarisoa, J.; Teuliere, C.; Chateau, T. Deep manta: A coarse-to-fine many-task network for joint 2d and 3d vehicle analysis from monocular image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2040–2049.

MOST WIEDZY Downloaded from mostwiedzy.pl

34. Liao, Z. Trainable Activation Function Supported CNN in Image Classification. *arXiv* **2004**, arXiv:2004.13271. Available online: https://arxiv.org/abs/2004.13271 (accessed on 9 November 2020).
35. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
36. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
37. Hoiem, D.; Chodpathumwan, Y.; Dai, Q. Diagnosing error in object detectors. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; pp. 340–353.
38. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
39. Dai, J.; Li, Y.; He, K.; Sun, J. R-fcn: Object detection via region-based fully convolutional networks. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 379–387.
40. Yoo, D.; Park, S.; Paeng, K.; Lee, J.Y.; Kweon, I.S. Action-driven object detection with top-down visual attentions. *arXiv* **2016**, arXiv:1612.06704. Available online: https://arxiv.org/abs/1612.06704 (accessed on 9 November 2020).
41. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
42. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37.

MOST WIEDZY Downloaded from mostwiedzy.pl