

Article

Region Based CNN for Foreign Object Debris Detection on Airfield Pavement

Xiaoguang Cao ^{1,†}, Peng Wang ^{1,†}, Cai Meng ^{1,*}, Xiangzhi Bai ^{1,2,*} , Guoping Gong ¹, Miaoming Liu ¹ and Jun Qi ¹

¹ Image Processing Center, Beijing University of Aeronautics and Astronautics, Beijing 100191, China; xgcao@buaa.edu.cn (X.C.); paulwp@buaa.edu.cn (P.W.); gonggp@buaa.edu.cn (G.G.); liumiaoming@ime.ac.cn (M.L.); qijun1994@126.com (J.Q.)

² State Key Laboratory of Virtual Reality Technology and Systems, Beijing University of Aeronautics and Astronautics, Beijing 100191, China

* Correspondence: tsai@buaa.edu.cn (C.M.); jackybxz@buaa.edu.cn (X.B.); Tel.: +86-010-8233-8578 (C.M.); +86-010-8231-6544 (X.B.)

† These two authors contributed equally to this work.

Received: 10 November 2017; Accepted: 29 January 2018; Published: 1 March 2018

Abstract: In this paper, a novel algorithm based on convolutional neural network (CNN) is proposed to detect foreign object debris (FOD) based on optical imaging sensors. It contains two modules, the improved region proposal network (RPN) and spatial transformer network (STN) based CNN classifier. In the improved RPN, some extra select rules are designed and deployed to generate high quality candidates with fewer numbers. Moreover, the efficiency of CNN detector is significantly improved by introducing STN layer. Compared to faster R-CNN and single shot multiBox detector (SSD), the proposed algorithm achieves better result for FOD detection on airfield pavement in the experiment.

Keywords: foreign object debris; object detection; convolutional neural network; vehicular imaging sensors

1. Introduction

After a plane crashed in Paris de Gaulle, the detection of foreign object debris (FOD) on airfield pavement has become more and more important. Actually, FOD detecting is one of the crucial technologies for intelligent vehicular systems. The safety and convenience of modern transportation could be significantly improved by adopting an efficient FOD detecting system. However, complex airfield environment makes the detection of FOD a challenging task. Due to the variation of background and influence of automotive imaging system, FODs could be hardly detected and recognized on airfield pavement by traditional features, such as scale invariant feature transform (SIFT) [1], histograms of oriented gradients (HOG) [2] and local binary patterns (LBP) [3].

Deep learning is one of the rapidly developing technologies in the area of big data. Since [4] significantly improved the effect of image classification, convolutional neural network (CNN) has been widely introduced into computer vision applications, such as image classification [5–8], face verification [9–11], semantic segmentation [12–15], object detection [16–18] and image annotation [19–21]. Moreover, some CNN based algorithms are proposed to solve transportation problems [22–24]. Based on various public available datasets such as ImageNet [25], Pascal VOC [26] and COCO [27], CNN algorithms have been proved to perform better in detection and recognition than traditional feature methods. Compared with these manually designed features, CNN based features have better resolution and robustness for FOD detection. Actually, the FOD problem consists of two tasks: target location and object classification on pavement. Aimed at these two tasks, a novel

two-stage framework is designed and introduced in this paper. In the first stage, region proposal network (RPN) [28], as a kind of fully convolutional network (FCN) [12], is trained end-to-end to generate FOD location proposals. In the second stage, CNN classifier based on spatial transformer network (STN) [29] is applied to obtain the parameters of scale, rotation and warping. Due to the performance of STN, FODs could be correctly identified by generated features, regardless of image distortion.

A preliminary version of this paper was presented in [30]. The highlights and extensions in this paper could be summarized as follows.

1. A new FOD detection framework based on CNN models for FOD detection is proposed with improved region proposal network and spatial transformer network.
2. RPN is firstly introduced and improved to generate high quality region proposals for FOD detection on airfield pavement. In addition, some candidate select rules are designed to reduce quantity and improve quality of region proposals.
3. The STN based CNN classifier is proved to be effective for FOD classification. Moreover, the proposed framework achieves better result than other detection algorithms, such as faster-RCNN [28] and SSD [17], for FOD detection on airfield pavement.
4. The vehicular imaging system, including DGPS, cameras, alarm, FOD management system and remote query system, is presented and discussed in detail.

The rest of this paper contains four main sections. In Section 2, some recent works are analyzed, including traditional feature based algorithms and some CNN algorithms. In Section 3, the overall detection framework is given in details, including location based on RPN and classification based on STN. In Section 4, comparison with other algorithms is conducted and discussed. Conclusions are given and discussed in Section 5.

2. Related Work

To solve FOD detection problem, some effective algorithms are proposed recently [31–36]. The algorithms based on different sensors, such as actively scanning LiDAR system [31], mm-wave FMCW radar [34] and wideband 96 GHz Millimeter-Wave Radar [35], could achieve good results in different environments. A cosecant squared beam pattern in elevation and a pencil-beam pattern in azimuth, generated through folded reflectarray antenna (FRA) by phase only control, is analyzed to detect objects on ground [36]. A multi-sensor system, based on inherent feature of FOD, is introduced to detect and recognize FOD [32]. Based on a large amount of prior knowledge and manually designed feature extractor, these methods could transform pixel values of an image into a suitable internal representation. Thus, these methods are effective for detection of FODs with less noise, but not working for FODs with complex background and noise.

Actually, CNN based detection has become more and more popular. There are two basic research methods: region proposal based and non-region proposal based algorithms. Region proposal based algorithm, such as faster R-CNN [28], R-FCN [37] and Mask R-CNN [38], consists of a region of interest (ROI) generator and an object classifier. Region proposal network (RPN) is designed in faster R-CNN [28], by introducing the classification features into ROI extraction module. In this way, better region proposals could be generated in less time. Moreover, this leads to high recognition accuracy and good real-time performance. Semantic segmentation method is introduced in R-FCN [37], which could calculate classification possibility of all region proposals in one image at the same time. This enlightening strategy greatly reduces the running time of ROI classification. Finally, object detection, semantic segmentation and key point detection are solved by one CNN algorithm [38]. Non-region proposal based algorithm, such as YOLO [16] and SSD [17], usually has better real-time performance with lower detection accuracy. Object detection is described as a regression problem in YOLO [16], with an end-to-end network. YOLOv2 and YOLO9000 [39] are two improved versions with faster speed and higher accuracy. However, YOLO could hardly detect small objects such as

FODs on pavement. Similar to YOLO, SSD [17] is also based on regression of target bounding box. To achieve higher accuracy, feature maps from multi-layers are introduced to the locating and scoring module. This strategy has been proved to achieve better performance.

To detect FODs on airport pavement, improved RPN and STN based CNN classifier are designed as given in Section 3.

3. Algorithm

The FOD detection framework is shown in Figure 1, which contains two stages. Firstly, region proposal network (RPN) [28] is adopted to generate a set of original object proposals, which are taken as FOD candidates. Moreover, some select rules are designed to reduce computational cost and fix image size. Secondly, spatial transformer network (STN) [29] is introduced to adjust the targets in region proposals from the influence of scale, rotation and warping. Then, these adjusted proposal images are fed to CNN classifier to extract features and identify FODs.

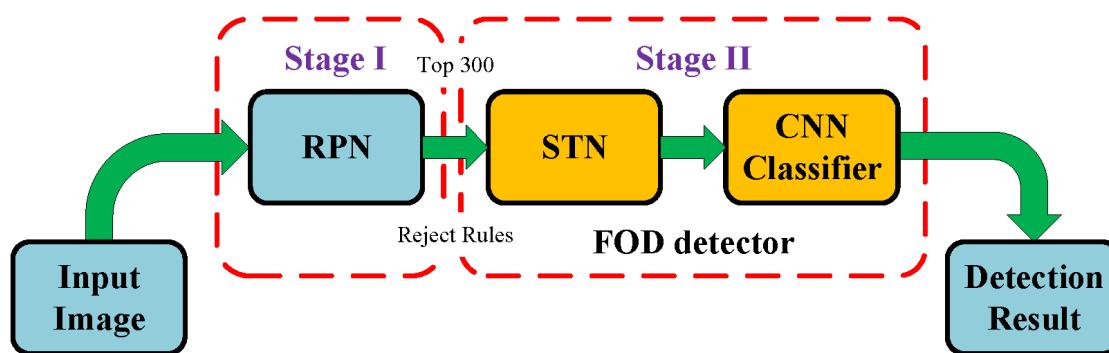


Figure 1. The framework of the FOD detection in this paper.

3.1. Locate FOD Candidates with Improved RPN

Generally, sliding window on images is a widely used strategy for target detection. With well-designed features, such as SIFT [1] and HoG [2], the window with target inside could be identified from background. In this way, the target detection problem could be described as a classification problem in sliding windows. However, sliding window is a low efficiency method, which costs much computing time and memory. Recently, a series of region-based convolutional neural networks (R-CNN) [28,37,38,40,41] is proposed to detect targets using deep learning method. One of the most meaningful ideas from R-CNN algorithms is the region proposal strategy to locate target candidates from background. On the one hand, region proposal algorithm is much faster than sliding window on whole image because region proposal algorithm reduces the number of candidates from millions to hundreds or thousands. On the other hand, region proposal algorithm has higher recall rate for finding all objects in an image, which could improve object location efficiency significantly. In this paper, some prior knowledge is introduced in region proposals to reduce the number of FOD candidates for better location.

The algorithm to generate region proposals is based on region proposal network (RPN) [28]. Actually, RPN takes an image as input and generates a series of rectangular object candidates with corresponding objectness scores. To generate region proposals as shown in Figure 2, a small network is slid over the convolution feature map. In each sliding window, k reference boxes (anchors) [28] is defined in different scales and aspect ratios to predict various region proposals. In other words, RPN is designed with the help of convolution features from detection module. The intersection over union (IoU) between region proposals and ground-truth bounding box, is chosen as evaluate index of the proposed FOD detection framework. For FOD detection in our dataset, targets in average shape

80×80 locate in 2048×2048 input image. To match the scale of FODs, the anchor boxes are set with area of 100^2 and aspect ratios of 1:1, 1:2 and 2:1.

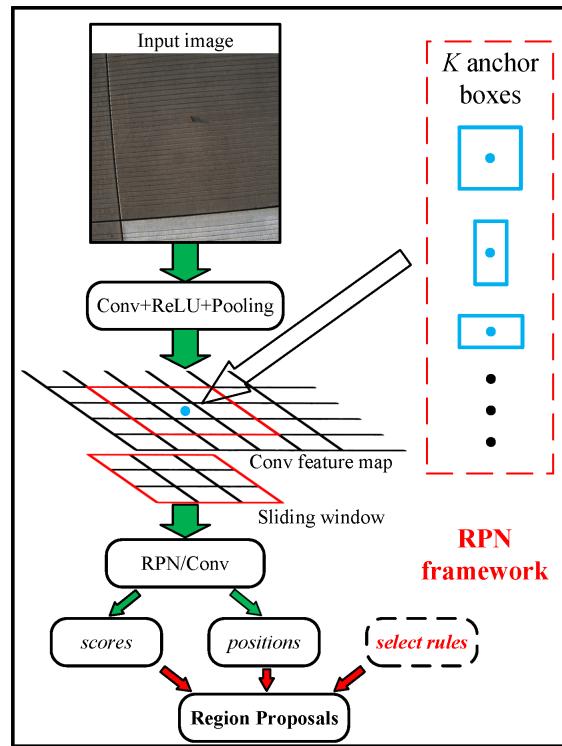


Figure 2. The RPN framework for FOD location.

Although there are much fewer FOD candidates generated by RPN than by common sliding window paradigm, the candidate number is still very large for classification. To further reduce FOD candidates, a novel series of select rules is designed and introduced in RPN framework. Actually, candidate FOD bounding boxes by original RPN vary in different sizes. For all region proposals, FODs such as screws and stones have different shape scales from false alarms. Therefore, three select rules based on prior knowledge are proposed and listed as below.

Rule 1: Region proposals by RPN are filtered by high aspect ratio $Ratio$, which is expressed as

$$Ratio = \frac{\max(w, h)}{\min(w, h)}, \quad (1)$$

$$Ratio < T_{ratio}.$$

w and h are the width and height of the image, respectively, and T_{ratio} is a threshold with a constant value.

Rule 2: The areas of proposals containing FODs should be in range $[T_{min}, T_{max}]$, which is expressed as

$$Area = w \times h, \quad (2)$$

$$T_{min} \leq Area \leq T_{max}.$$

$Area$ is the area of region proposal.

Rule 3: The proposals with higher objectness scores O_l generated by RPN are picked out, which is expressed as

$$O_l > T_{objectness}. \quad (3)$$

$T_{objectness}$ is a constant threshold for judging objectness.

The thresholds for these three rules are set as $T_{ratio} = 1.5$, $T_{min} = 60^2$, $T_{max} = 100^2$ and $T_{objectness} = 0.8$. By deploying these rules, the number of region proposals by improved RPN could be greatly reduced by about 60%. In this way, more accurate candidate proposals with fewer number could be generated. Moreover, the efficiency of FOD classification could be greatly improved.

3.2. Spatial Transformer Network

STN [29] is an effective CNN framework to learn scale, rotation and warping of images. With a predicted affine transformation by STN, an input image $I(x, y)$ could be adjusted to a rectified image $I'(x, y)$. As shown in Figure 3, affine matrix is firstly regressed via a localization network. Then, predicted transformation parameters are used to generate a sampling grid. The sampling grid is a set of points from input map for affine transformation. Finally, feature map and the sampling grid are utilized as inputs to be sampled. The architecture of STN is shown in Figure 4. Details of STN algorithms are given below.

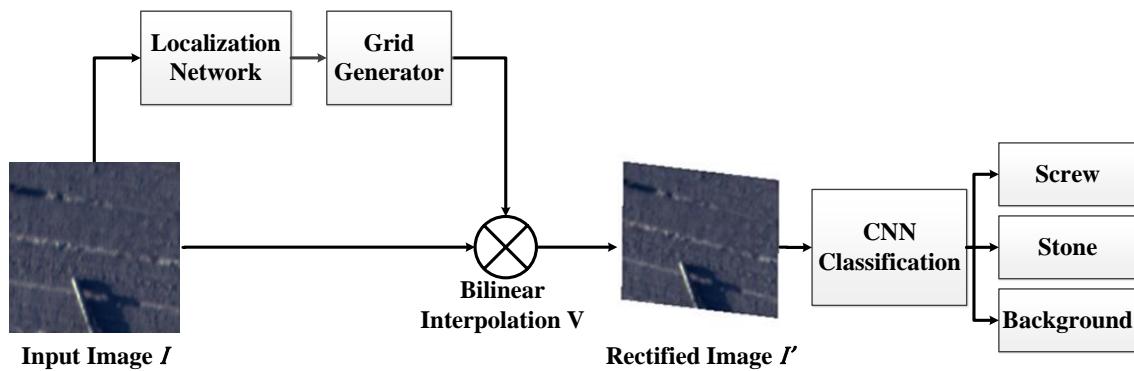


Figure 3. The FOD classification framework.

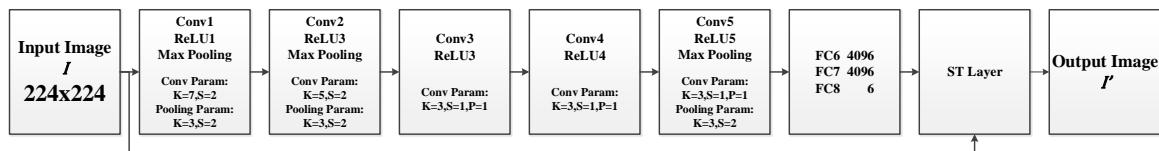


Figure 4. The architecture of spatial transformer network.

3.2.1. Localization Network

The localization network takes feature map with width w , height h and channel c as input U . Then, the transformation parameters could be generated by $O = f_{loc}(U)$. In this paper, the effect of vehicular imaging system is assumed as affine transformation. Therefore, output of localization network O is a six-dimensional vector from fully-connect layer. Actually, there are five convolution layers and three fully-connect layers in the localization network.

3.2.2. Grid Generator

To perform a warping of input feature map, each output pixel is computed by locating a sampling kernel in a particular area of the input feature map. In general, the output pixels are defined on a regular grid $G_i = (x_i^t, y_i^t)$, which form an output feature map P . The relationship between input and output feature maps is expressed as below.

$$\begin{bmatrix} x_i^s \\ y_i^s \\ 1 \end{bmatrix} = \Gamma_\theta \begin{bmatrix} x_i^t \\ y_i^t \\ 1 \end{bmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{bmatrix} x_i^t \\ y_i^t \\ 1 \end{bmatrix} \quad (4)$$

In Equation (4), Γ_θ signifies affine transformation matrix for image distortion.

3.2.3. Sampler

In the sampler module, pixels in P' are calculated through regional bilinear interpolation on the input image P . With all pixel values obtained, the rectified image I could be generated as:

$$I' = V(P, I) \quad (5)$$

V represents the bilinear interpolation. Specifically, V is also a differentiable module for back propagation.

3.3. FOD Classification with Convolutional Neural Network

In FOD classification framework, each extracted region proposal is identified as background, stone or screw. A convolutional neural network (CNN) classifier is adopted for this application, as presented in Figure 5. The network is based on the type C of visual geometry group (VGG) model [5], which is one of the most famous CNN algorithm. There are thirteen convolution layers followed by three fully-connect layers in VGG model. The FOD classification network is fine-tuned from pre-trained models of VGG on ILSVRC2012 dataset [25], with the Caffe toolbox [42] in our experiments.

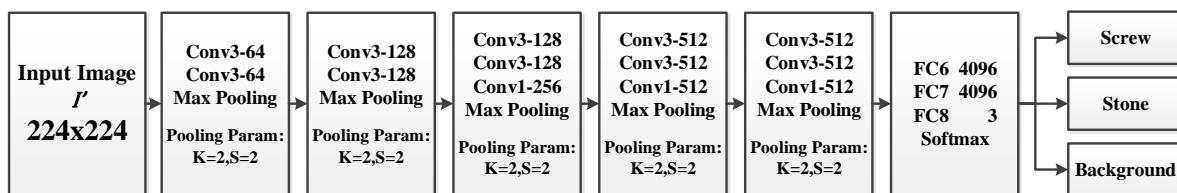


Figure 5. The FOD classification architecture.

Actually, the last layer of VGG model is a fully-connect layer with 1000 output neurons. To match our application with the pre-trained VGG model, the first fifteen layers of VGG model are taken as feature extractor. The last layer of VGG model is replaced by a classifier to predict the candidate rectangular regions as background, screw or stone. Therefore, only three output neurons are necessary in the last layer of classification network.

4. Experiments

The airfield pavement images for experiment are sampled by a vehicular imaging system in Tianjin Binhai International Airport (ZBTJ). Actually, there are already some automatic vehicular imaging and processing systems, such as PCES, PAVUE and ARAN. The framework of vehicular FOD detection system on airfield pavement in this paper is shown in Figure 6. There are mainly two parts: FOD detecting car and offline database in this framework. In detail, some important modules are listed as below.

DGPS: The differential GPS (DGPS) base station and mobile station could provide the real-time position of the FOD detecting car and the detected FODs.

Cameras: There are four GT2050C cameras with 2048×2048 resolution scanning for 5m in width at the same time.

Alarm: The information of predicted FODs, including FOD classes and accurate locations from the image and DGPS, is sent as alarm and saved in offline database.

FOD Management System: The detection and cleaning of FODs are managed and upgraded by the management system in offline database.

Remote Query System: The real-time status of FODs in database could be remotely inquired by others, such as airfield pavement cleaning system and plane operating system.

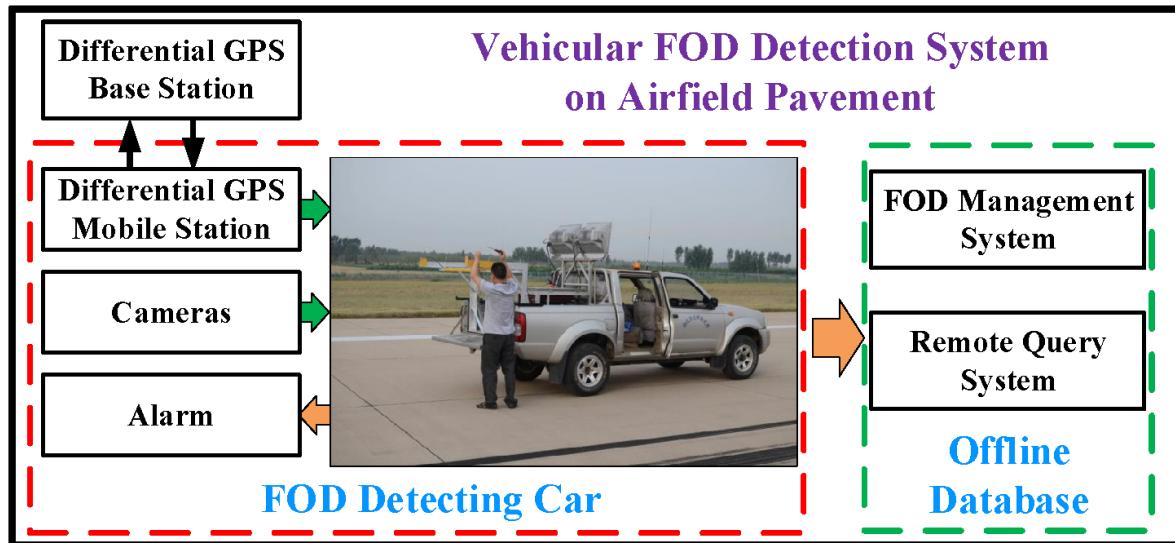


Figure 6. The framework of vehicular FOD detection system on airfield pavement.

To meet the 25 fps real-time sampling frequency, the speed of FOD detection car should be less than 31.25 m/s and the processing time for each frame should be less than 40 ms. With GTX 1080ti gpu, the proposed detection algorithm could achieve 26 fps with high accuracy, which is faster than the original faster R-CNN [28] with 14 fps and slower than SSD [17] with 31 fps.

For FODs on pavement, lamp covers, marker lines, dilapidations and tire marks are the main noise and false alarms. This makes FOD detection a complex and challenging task. Actually, screws and stones are major targets to be recognized for airfield pavement FOD detection. These two kinds of FODs and background make up the final three predict classes.

4.1. The Dataset and Training

The airfield pavement image dataset contains 12,231 images with shape 2048×2048 sampled by the vehicular imaging system in our dataset. In these images, there are 3562 screws and 4202 stones. The number of foreign object debris in one image varies from zero to four. The shape of boundary box for screws and stones is 80×80 on average, in the dataset ground truth. Region proposals by RPN are employed to fine-tune the FOD detector with STN. The region proposals with $\text{IoU} \geq 0.7$ over ground-truth boxes are treated as positive, while the region proposals with $\text{IoU} \leq 0.3$ are labeled as negative, as shown in Figure 7. The top 100 score region proposals are chosen as training samples. This makes the number of negative samples much larger than that of positive samples in our application. To obtain a reasonable number of candidates, 16 stone windows, 16 screw windows and 32 background windows are used in each training batch.

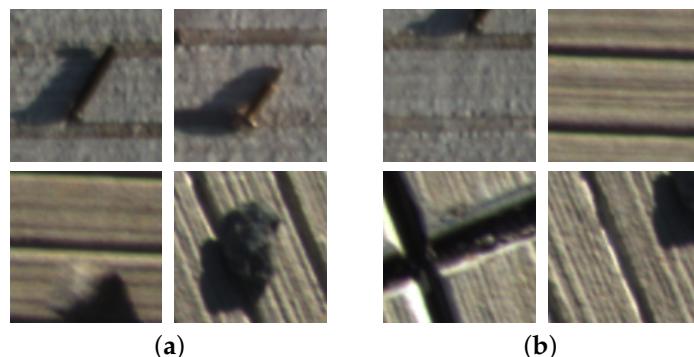


Figure 7. (a) Positive FOD samples with $\text{IoU} \geq 0.7$; and (b) negative FOD samples with $\text{IoU} \leq 0.3$.

To train the FOD Detector, 60% of images are taken as training dataset, another 20% are taken as validate dataset and the rest for test. The Caffe Toolbox [42] are chosen to train our networks, with Stochastic Gradient Descent (SGD) [43] for optimization. As one of the most widely used CNN models, VGG model for fine-tuning is trained on 1.2 million images from ILSVRC2012 dataset with 1000 classes. Before fine-tuning, parameters of all layers except FC8 layer are initialized with the pre-trained type C VGG model. In this way, the training time could be apparently reduced and a better fine-tuned model could be generated. The number of training iterations is set as 50,000. The learning rate is initially set as 0.001 and decaying by 0.1 after each 10,000 iterations.

4.2. The Experiments of Location

Due to the sequential structure of the most target detection algorithms, location efficiency significantly affects classification results. Candidates with more accurate locations could be more easily classified as targets or background, especially for CNN-based classifier. Actually, RPN is improved by adding prior knowledge based select rules, to obtain FOD candidates with high quality. To evaluate generated proposals, Figure 8 shows the candidate rectangles extracted by Selective Search [40,41,44], original RPN in faster R-CNN [28] and improved RPN for images in our dataset.

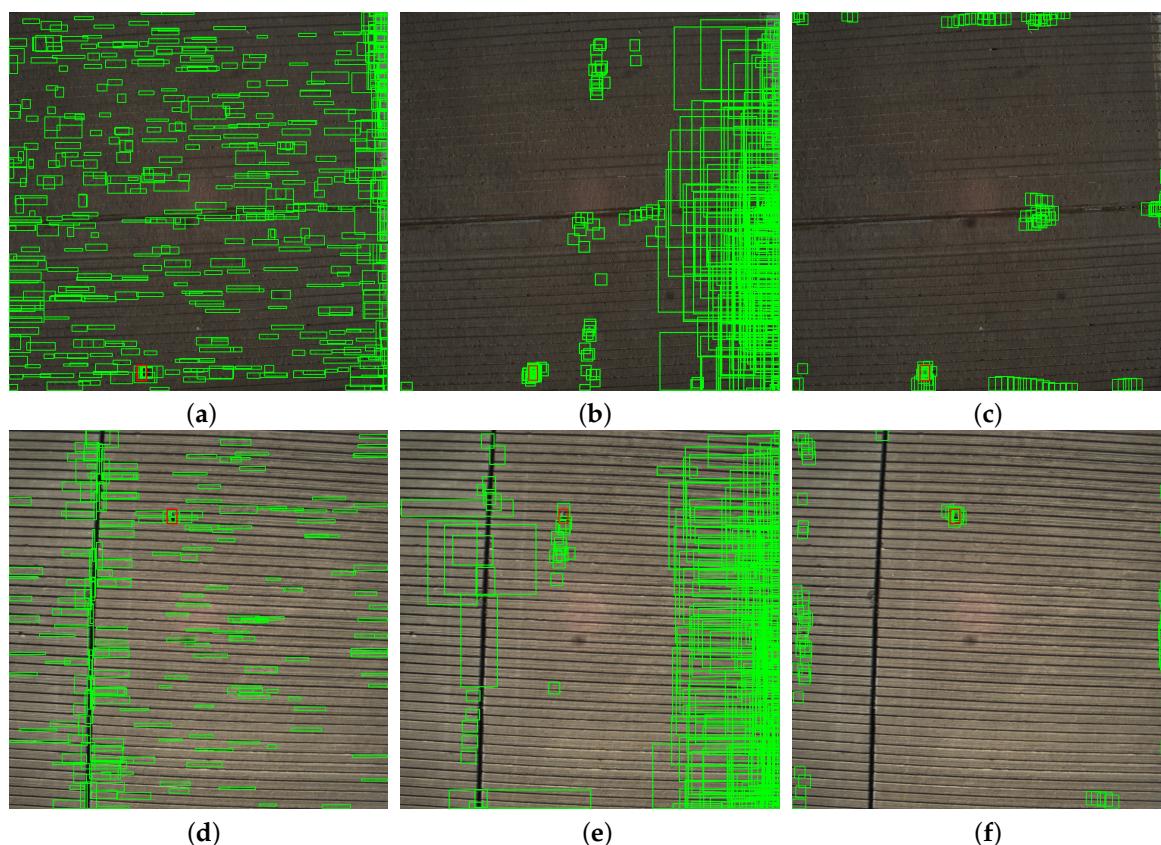


Figure 8. (a,d) FOD location results by Selective Search; (b,e) results of original RPN; and (c,f) results of improved RPN. The green boxes are generated FOD candidates by these two methods respectively, while red boxes are ground truths.

The patches generated by Selective Search and original RPN vary in different sizes and may contain multiple objects. This uncertainty affects the accuracy and robustness of FOD detector and produces more false positive results. The improved RPN generates much less region proposals with more accurate locations. To compare the quality and quantity of generated proposals, proposal numbers and recall rates in different IoU levels are shown in Table 1.

Table 1. The recall of Selective Search and RPN.

Methods	IoU	Recall Num	Total Num	Recall Rate	Average Num
Selective Search	IoU > 0.5	2108	2469	85.37%	800
	IoU > 0.6	1875	2469	75.94%	800
Region Proposal Network (RPN)	IoU > 0.5	2263	2469	91.65%	Top5
	IoU > 0.6	2253	2469	91.25%	Top5
	IoU > 0.5	2399	2469	97.16%	Top10
	IoU > 0.6	2394	2469	96.96%	Top10
	IoU > 0.5	2462	2469	99.72%	Top20
	IoU > 0.6	2461	2469	99.60%	Top20

In general, the number of candidate boxes generated by RPN is less than that generated by Selective Search, but the recall rate is about 10% higher. It indicates that region proposals by RPN are with high quality and easy to be identified by the followed FOD classifier.

4.3. The Experiments of Classification

The STN is proposed to learn and rectify image distortion, such as scale, rotation and warping. Actually, CNN classifier with STN could improve classification accuracy, especially on small dataset. To measure the effect of STN, four comparative experiments are conducted. These four experiments are FOD detector without fine-tune, STN based FOD detector without fine-tune, FOD detector with fine-tune and STN based FOD detector with fine-tune. These results are presented in Table 2.

Table 2. The results of classification.

FOD Detector	Recall Rate
FOD classification (no fine-tune)	94.52%
STN + FOD classification (no fine-tune)	96.31%
FOD classification + fine-tune	96.45%
STN + FOD classification + fine-tune	97.67%

As presented in Table 2, the recall rates of FOD classification are significantly increased by introducing STN. This is because STN decreases the effect of image distortion while sampling. Compared to ImageNet or Pascal VOC, the sample number is much smaller in our dataset, which the various image distortions may not be fully represented by samples. With fine-tuning from VGG model, STN based FOD classifier could achieve high accuracy for classification.

4.4. Comparison with Other Algorithms

To test the performance of proposed method, three other object detection algorithms are introduced for comparison. These algorithms are faster R-CNN [28], Single Shot MultiBox Detector (SSD) [17] and Selective Search [44] with FOD detector. Faster R-CNN is the third generation of the region proposal based CNN algorithms, in which RPN is proposed to enhance the real-time detection performance. RPN is designed to generate region proposals through convolution feature maps instead of original image. With the help of shared features from classification network, RPN could achieve high recall rate with less running time. As region proposals by RPN have good robustness for various target scales, faster R-CNN are one of the most popular detection CNN algorithms. In our experiment, the Zeiler and Fergus model [45] based faster R-CNN is deployed for comparison. SSD, as an end-to-end fully convolutional network, is based on regression of target bounding box. Usually, regression based detection algorithm, such as [16,17], has good real-time performance with lower

accuracy, comparing to region based CNNs. To increase the accuracy of location and classification, features from multi-layers are introduced to the locating and scoring module in SSD framework. As one of the most famous regression based detection CNNs, SSD512 model is deployed for comparison with the proposed algorithm. For compared faster R-CNN and SSD, parameters are set as the same to the original algorithms [17,28], except for object class numbers. Moreover, Selective Search method with the proposed STN based FOD detector is also conducted to evaluate detection efficiency.

For all experiments, if the correctly identified candidate has a greater value than 0.5 intersection over union (IoU) overlap with a ground-truth box, this will be recorded as a successful recall. All wrongly detected candidates, including wrongly sorted targets, are counted as false alarms. In detail, false alarm rate (FAR) and recall rate (RR) are, respectively, defined as:

$$\begin{aligned} FAR &= \frac{TN}{TP + TN'} \\ RR &= \frac{TP}{TP + FP}. \end{aligned} \quad (6)$$

In Equation (6), TP is the number of true positive samples, TN is the number of true negative samples, FP represents the number of false positive samples, and FN signifies the number of false negative samples.

FARs of four algorithms are shown in Table 3, while RRs for screws and stones are shown in Table 4. It indicates that the proposed algorithm could predict FODs with the least false alarms and the highest recall rates. Faster R-CNN uses RPN as candidate detection network, which could achieve high recall rates for all classes regardless of FOD scales. However, more backgrounds are also introduced as region proposals. This is because the sample number of dataset is not large enough to train the model with good candidate selecting capability. Although SSD greatly enhances the real-time performance, detection for small objects is not effective. Actually, the average length ratio between FODs and whole images is about 0.04, which is usually considered as small target detection problem [46–49]. It means that convolution features could possibly be lost through pooling operation with no fully-connect layers. This leads to high FAR and low PR of SSD algorithm for small objects, such as FODs. Selective Search algorithm generates more FOD candidates as shown in Figure 8, which lead to high FAR usually. However, Selective Search with FOD detector achieves low FAR due to the proposed high efficiency classifier. In other words, the proposed STN based classifier correctly identified most of the candidates by Selective Search. Moreover, RRs for both classes are limited by the low recall rate of Selective Search, as shown in Table 1 in Section 4.2.

Table 3. The detection evaluations by FAR.

Methods	FAR
faster R-CNN	11.02%
SSD	8.19%
Selective Search + FOD Detector	1.21%
RPN + FOD Detector	0.66%

Table 4. The recall rates of screw and stone.

Methods	Screw RR	Stone RR
faster R-CNN	83.51%	93.84%
SSD	87.72%	88.63%
Selective Search + FOD Detector	80.63%	81.46%
RPN + FOD Detector	96.90%	96.40%

The proposed algorithm achieves low FAR and high RR due to improved RPN and STN based classifier with high accuracy. Actually, the improved RPN has some extra select rules which could remove more false alarms for FOD detection. The STN based classifier could correctly identify most of the rest region proposals as one of FOD classes or background.

To compare the overall performance of four algorithms, mean average precision (mAP) is calculated according to different precision-recall pairs. The mAPs of different algorithms are listed in Table 5. The proposed algorithm achieves the highest mAP, which means that the proposed algorithm is effective and robust for FOD detection.

Table 5. The mean average precisions.

Methods	mAP
faster R-CNN	89.43%
SSD	89.92%
Selective Search + FOD Detector	96.65%
RPN + FOD Detector	98.41%

These experiments generally indicate that FOD detector with RPN outperforms faster R-CNN, SSD and Selective Search with FOD detector in our dataset. This is because the candidate boxes produced by improved RPN are with better quality and quantity than other detection algorithms. STN based classifier could correctly identify most of the generated region proposals as one of FOD classes or background, regardless of image distortion.

5. Conclusions

This paper presents a method based on region proposal network (RPN) and convolutional neural network (CNN) with spatial transformer network (STN) to detect foreign object debris (FOD). Some extra region proposal select rules are introduced to improve the quality and quantity of FOD candidates generated by RPN. Moreover, CNN classifier with STN could learn rotation invariant features from image, which usually are not fully represented by a small number of samples. The experiments indicate that the proposed algorithm is more effective and robust than some popular detection algorithms, such as faster R-CNN and Single Shot MultiBox Detector for FOD detection.

Acknowledgments: This work is supported by the National Natural Science Foundation of China under Grant U1736217, the Program for New Century Excellent Talents in Universities under Grant NCET-13-0020, the Fundamental Research Funds for the Central Universities under Grant YWF-17-BJ-Y-69.

Author Contributions: Xiaoguang Cao and Peng Wang conceived and designed the algorithm and experiments; Peng Wang and Guoping Gong performed the experiments and comparisons; Xiangzhi Bai analyzed the data and result; Xiaoguang Cao and Xiangzhi Bai provided the vehicular imaging sensor system; Peng Wang and Xiangzhi Bai wrote the paper; Cai Meng helped proofread the manuscript; and Miaoming Liu and Jun Qi ran the vehicular system and collected FOD data.

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

References

1. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.
2. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–25 June 2005; pp. 886–893.
3. Ojala, T.; Pietikäinen, M.; Mäenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987.
4. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances In Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.

5. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
6. 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, Seattle, WA, USA, 27–30 June 2016; pp. 770–778.
7. Kim, K.W.; Hong, H.G.; Nam, G.P.; Park, K.R. A Study of Deep CNN-Based Classification of Open and Closed Eyes Using a Visible Light Camera Sensor. *Sensors* **2017**, *17*, 1534, doi:10.3390/s17071534.
8. Satat, G.; Tancik, M.; Gupta, O.; Heshmat, B.; Raskar, R. Object classification through scattering media with deep learning on time resolved measurement. *Opt. Express* **2017**, *25*, 17466–17479.
9. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. Deepface: Closing the gap to human-level performance In face verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1701–1708.
10. Sun, Y.; Liang, D.; Wang, X.; Tang, X. Deepid3: Face recognition with very deep neural networks. *arXiv* **2015**, arXiv:1502.00873.
11. Chen, J.; Patel, V.M.; Chellappa, R. Unconstrained face verification using deep cnn features. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Lake Placid, NY, USA, 7–9 March 2016; pp. 1–9.
12. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
13. Liu, Z.; Li, X.; Luo, P.; Loy, C.; Tang, X. Semantic image segmentation via deep parsing network. In Proceedings of the IEEE International Conference on Computer Vision (2015), Los Alamitos, CA, USA, 7–13 December 2015; pp. 1377–1385.
14. Chen, L.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, doi:10.1109/TPAMI.2017.2699184.
15. Nguyen, T.; Bui, V.; Lam, V.; Raub, C.B.; Chang, L.C.; Nehmetallah, G. Automatic phase aberration compensation for digital holographic microscopy based on deep learning background detection. *Opt. Express* **2017**, *25*, 15043–15057.
16. Redmon, J.; Divvala, S.K.; Girshick, R.B.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 27–30 June 2016; pp. 779–788.
17. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.; 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.
18. Jiao, J.; Li, F.; Deng, Z.; Ma, W. A Smartphone Camera-Based Indoor Positioning Algorithm of Crowded Scenarios with the Assistance of Deep CNN. *Sensors* **2017**, *17*, 704.
19. Wu, F.; Wang, Z.; Zhang, Z.; Yang, Y.; Luo, J.; Zhu, W.; Zhuang, Y. Weakly semi-supervised deep learning for multi-label image annotation. *IEEE Trans. Big Data* **2015**, *1*, 109–122.
20. Lev, G.; Sadeh, G.; Klein, B.; Wolf, L. RNN fisher vectors for action recognition and image annotation. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 833–850.
21. Murthy, V.N.; Maji, S.; Manmatha, R. Automatic image annotation using deep learning representations. In Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, Shanghai, China, 23–26 June 2015; pp. 603–606.
22. Yao, Y.; Tian, B.; Wang, F. Coupled Multivehicle Detection and Classification WithPrior Objectness Measure. *IEEE Trans. Veh. Technol.* **2017**, *66*, 1975–1984.
23. Li, X.; Li, L.; Flohr, F.; Wang, J.; Xiong, H.; Bernhard, M.; Pan, S.; Gavrila, D.M.; Li, K. A unified framework for concurrent pedestrian and cyclist detection. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 269–281.
24. Chen, L.; Hu, X.; Xu, T.; Kuang, H.; Li, Q. Turn Signal Detection During Nighttime by CNN Detector and Perceptual Hashing Tracking. *IEEE Trans. Intell. Transp. Syst.* **2017**, *99*, 1–12.
25. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252.

26. Everingham, M.; Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338.
27. Lin, T.; Maire, M.; Belongie, S.J.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects In Context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
28. Ren, S.; He, K.; Girshick, R.B.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the Advances In Neural Information Processing Systems, Montreal, QC, Canada, 11–12 December 2015; pp. 91–99.
29. Jaderberg, M.; Simonyan, K.; Zisserman, A.; Kavukcuoglu, K. Spatial transformer networks. In Proceedings of the Advances In Neural Information Processing Systems, Montreal, QC, Canada, 11–12 December 2015; pp. 2017–2025.
30. Cao, X.; Gong, G.; Liu, M.; Qi, J. Foreign Object Debris Detection on Airfield Pavement Using Region Based Convolution Neural Network. In Proceedings of the International Conference on Digital Image Computing: Techniques and Applications, Gold Coast, Australia, 30 November–2 December 2016; pp. 1–6.
31. Mund, J.; Zouhar, A.; Meyer, L.; Fricke, H.; Rother, C. Performance evaluation of LiDAR point clouds towards automated FOD detection on airport aprons. In Proceedings of the 5th International Conference on Application and Theory of Automation In Command and Control Systems, Toulouse, France, 30 September–2 October 2015; pp. 85–94.
32. Li, Y.; Xiao, G. A new FOD recognition algorithm based on multi-source information fusion and experiment analysis. *Proc. SPIE* **2011**, doi:10.1117/12.900576.
33. Li, J.; Deng, G.; Luo, C.; Lin, Q.; Yan, Q.; Ming, Z. A Hybrid Path Planning Method In Unmanned Air/Ground Vehicle (UAV/UGV) Cooperative Systems. *IEEE Trans. Veh. Technol.* **2016**, *65*, 9585–9596.
34. Ölzen, B.; Baykut, S.; Tulgar, O.; Belgül, A.U.; Yalçın, İ.K.; Şahinkaya, D.S.A. Foreign object detection on airport runways by mm-wave FMCW radar. In Proceedings of the 25th IEEE Signal Processing and Communications Applications Conference, Antalya, Turkey, 15–18 May 2017; pp. 1–4.
35. Futatsumori, S.; Morioka, K.; Kohmura, A.; Okada, K.; Yonemoto, N. Detection characteristic evaluations of optically-connected wideband 96 GHz millimeter-wave radar for airport surface foreign object debris detection. In Proceedings of the 41st International Conference on Infrared, Millimeter, and Terahertz waves, Copenhagen, Denmark, 25–30 September 2016; pp. 1–2.
36. Zeitler, A.; Lanteri, J.; Pichot, C.; Migliaccio, C.; Feil, P.; Menzel, W. Folded reflectarrays with shaped beam pattern for foreign object debris detection on runways. *IEEE Trans. Antennas Propag.* **2010**, *58*, 3065–3068.
37. Dai, J.; Li, Y.; He, K.; Sun, J. R-FCN: Object Detection via Region-based Fully Convolutional Networks. In Proceedings of the Conference on Neural Information Processing Systems, Barcelona, Spain, 5–6 December 2016, pp. 379–387.
38. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R.B. Mask r-cnn. *arXiv* **2017**, arXiv:1703.06870.
39. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. *arXiv* **2016**, arXiv:1612.08242.
40. Girshick, R.B.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
41. Girshick, R.B. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Los Alamitos, CA, USA, 7–13 December 2015; pp. 1440–1448.
42. Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.B.; Guadarrama, S.; Darrell, T. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; pp. 675–678.
43. Bottou, L. Large-scale machine learning with stochastic gradient descent. In Proceedings of the COMPSTAT'2010, Paris, France, 22–27 August 2010; pp. 177–186.
44. Uijlings, J.R.; Sande, K.E.; Gevers, T.; Smeulders, A.W. Selective search for object recognition. *Int. J. Comput. Vis.* **2013**, *104*, 154–171.
45. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. In Proceedings of the European Conference on Computer Vision, Portland, OR, USA, 23–28 June 2013; pp. 818–833.
46. Bai, X.; Zhou, F. Analysis of new top-hat transformation and the application for infrared dim small target detection. *Pattern Recognit.* **2010**, *43*, 2145–2156.

47. Bi, Y.; Bai, X.; Jin, T.; Guo, S. Multiple Feature Analysis for Infrared Small Target Detection. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1333–1337.
48. Bai, X.; Bi, Y. Derivative Entropy-Based Contrast Measure for Infrared Small-Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2018**, doi:10.1109/TGRS.2017.2781143.
49. Li, Z.; Xie, J.; Tu, D.; Choi, Y.-J. Sparse Signal Recovery by Stepwise Subspace Pursuit In Compressed Sensing. *Int. J. Distrib. Sens. Netw.* **2013**, *1*, 945–948.



© 2018 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 (<http://creativecommons.org/licenses/by/4.0/>).