Oil Tank Detection With Improved EfficientDet Model

Su Xu, Haowei Zhang, Xiping He, Xiaoli Cao, and Jian Hu

***Abstract*— Rapid detection of oil tank targets has become a topic of significant and increasing interest because the quick acquisition of the distribution and volume of oil storage tanks has high economic and military value. However, research in the literature still faces many challenges in different scenarios. This letter presents a new oil tank detection approach for improved EfficientDet model. The proposed approach has three primary operations: 1) first, adding 3-D deformable convolution to the EfficientDet model as predetection to limit detection in the smaller area; 2) second, loading an attention mechanism appropriate for oil tank detection on the myNet model highlights tiny target information; and 3) lastly, training the myNet model repeatedly by focal loss function to seek better results. The result shows that the overall accuracy of mean average precision in the proposed approach increased by at least 8.25% compared with that of the tested conventional approaches. Therefore, it provides advantageous capabilities for monitoring oil tanks in high-resolution remote sensing imagery.**

***Index Terms*— Deep learning, EfficientDet, high-resolution remote sensing imagery, oil tank detection.**

I. Introduction

With the development of satellite and radar technology, automatic detection of tiny targets such as oil storage tanks is essential to emergency disaster assessment and military target reconnaissance. Therefore, it has been a trending topic and an area of interest in high-resolution imagery. Unfortunately, oil storage tanks are in different dimensions and shapes with different roofs depending on the type of the material to be stored have. Furthermore, the distribution and overlapping characteristic of circles in oil storage tanks have related to many factors, such as illumination, cloud cover, satellite observation and imaging conditions, background environment, and side wall occlusion. Several approaches [1]–[11], [31]–[34] have been studied in the past decade, generally divided into three categories: 1) traditional methods with circular structure information; 2) saliency methods; and 3) machine learning methods. In the first class of methods, Ok [1] proposed a traditional method to detect oil tanks. The method stands on the shadow evidence of the circular structures, thus negatively affecting the precision ratios computed if the structure is not visible or complete. Zhang and Liu [2] proposed a linear clustering saliency analysis-based detection model. Unfortunately, the method needs to consider all possible parameters, such as proportion and rotation, requiring significant computational power and time. Zerman *et al.* [3] proposed a slightly modified version of the fast radial symmetry transform. Ok and Başeski [4] presented a detection method based on circular radial symmetry from a single panchromatic satellite image. However, these methods have great difficulties setting transform parameters, locating the center position, and circling radius information. An unsupervised saliency model with Hough transform [5], Markov Chain [6], and shape-guide saliency [7] for saliency methods have been proposed in the literature. Unfortunately, the circular structures cannot be detected accurately when various contrast areas, scale differences, and shadows are caused by view angle and illumination. In addition, the main causes of the irregular boundary circle phenomenon lie in shadows, 3-D structures, and low contrast. A coarse-to-fine framework [8], hierarchical oil tank detector [9] with deep surrounding features, convolutional neural network (CNN) [10], and learning rotation-invariant CNN [11] machine learning models can efficiently learn features for oil tank detection. However, the attitude and height of satellites may lead to oil tanks suffering certain geometric distortions. Therefore, oil tank detection in high-resolution imagery poses a big challenge, especially in large-scale margins and abundant tiny targets. The traditional methods of detecting tiny targets are mainly feature fusion, cascading networks, model training, and receptive field. He *et al.* [12] proposed a network that applies a spatial pyramid pooling layer to extract and compute features over an entire image regardless of image sizes. Feature pyramid network (FPN) [13] improves the detection effect of tiny targets by introducing a top-down structure. However, for small objects, some gaps still exist in different features

Manuscript received May 2, 2022; revised May 29, 2022; accepted June 12, 2022. Date of publication June 15, 2022; date of current version June 24, 2022. This work was supported in part by the Chongqing Engineering Laboratory for Detection, Control and Integrated System, in part by the Chongqing Talents Intelligent Ecological Innovation and Entrepreneurship Demonstration Team under Grant CQYC201903246, in part by the Creative Research Groups of Chongqing Municipal Education Commission under Grant CXQT21034, in part by the Research and Demonstration of Key Technologies for UV Spectrum Monitoring of Surface Fire under Grant CQLK2022-2, in part by Chongqing Education Commission Foundation under Grant KJ1600605, in part by the Project of Chongqing Technology and Business University under Grant 990521004, in part by the Monitoring and Prediction of Pine Silk Worm Disaster Based on Hyperspectral Image Fellowship under Grant 950219067, and in part by the Multimode Human–Computer Interaction Technology to the Ecological Environment of Forestry Fellowship under Grant CSTC2017zdcy-zdyfX0067. *(Corresponding author: Haowei Zhang.)*

Su Xu is with the College of Computer Science and Information Engineering, Chongqing Technology and Business University, Chongqing 400067, P. R. China (e-mail: xusu@ctbu.edu.cn).

Haowei Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 473079, P. R. China (e-mail: haoweizhang@whu.edu.cn).

Xiping He, Xiaoli Cao, and Jian Hu are with the Chongqing Engineering Laboratory for Detection, Control and Integrated System, Chongqing Technology and Business University, Chongqing 400067, P. R. China (e-mail: hxp@ctbu.edu.cn).

Digital Object Identifier 10.1109/LGRS.2022.3183350

1558-0571 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Oestfold University College. Downloaded on March 26,2023 at 11:20:14 UTC from IEEE Xplore. Restrictions apply.

in FPN, for example, the gap between the high-resolution features with low semantic information and low-resolution features with high semantic information. Therefore, an unsupervised score is proposed based on bounding-box regression for precise positioning by image pyramid [28]. It mainly focuses on small region segments but cannot scale up to handle arbitrary massive input. R^2CNN algorithm [29] adopts backbone Tiny-Net, intermediate global attention block, and customized detector to locate tiny target accurately. However, the adverse impact of noise and lightweight architecture of Tiny-Net might lead to a drastic performance drop. Therefore, SF-Net and MDA-Net are adopted as substitutes for Tiny-Net in R^2CNN++ [30]. Mask-RCNN algorithm [14] identifies specific targets in the underlying distribution by adding a third branch to the faster R-CNN architecture. Mask R-CNN is only implemented with a region of interest align when considering the lack of mask annotations. Tan *et al.* [15] proposed EfficientDet and SpineNet [16] with faster training speed and better parameter efficiency than previous models. Unfortunately, they also obtain inferior results because the imbalanced and complex background deceives the affinity learning on small objects. Detecting tiny targets in high-resolution remote sensing images remains challenging because high background complexity leads to tedious work. Moreover, just a few pixel features carried by the tiny target can cause a loss of spectral information after downsampling. Finally, model degeneration can be attributed to an extreme foreground–background unbalanced distribution. Therefore, this letter proposes an improved EfficientDet network module to handle the aforementioned problems. The specific contributions are as follows.

1) Weakening the influence of noise and highlighting tiny target information by adding an attention mechanism (AM) on the EfficientDet model.
2) Developing a new model called myNet to fit the structure of tiny pixels, deforming sampling, and pooling patterns to increase the ability to capture tiny targets on adaptive regions effectively. myNet easily captures anomaly distributing in high-resolution remote sensing images by residual deformable 3-D convolution (RD3C).
3) Multiclass focal loss (MFL) is used to decrease the blindness and increase the foresight, until it finally improves the effect of model training. Experimental results show that this approach is an effective way for oil tank detection.

II. Methodology

This section presents the proposed improved EfficientDet network module named myNet, accomplished by adding AM, deformable convolution, and MFL. Fig. 1 shows the flowchart of the proposed approach. The details of each procedure are provided in Sections II-A–II-C.

A. Attention Mechanism

myNet utilizes an update gate to model the global context features from multiscale information for addressing the tiny target detection problem. It sequentially consists of three stages: spatial, pooling, and channel attention. First, a global

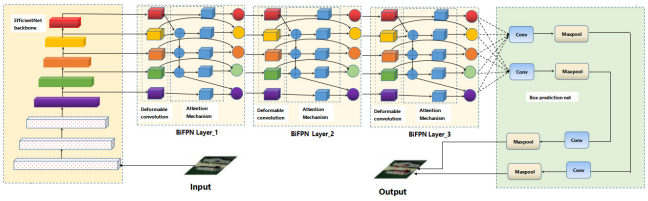

Fig. 1. Improved scheme by different methods.

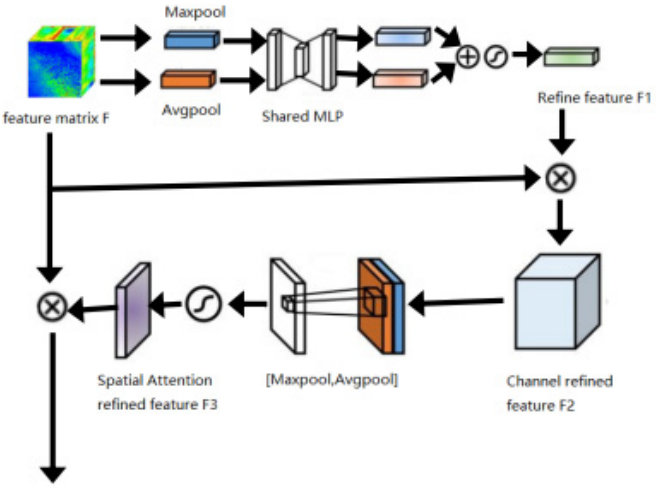

Fig. 2. Convolutional block attention module.

context modeling is performed to obtain a spatial attention map by calculating spatialwise weights of each position. Second, a softmax function tank normalizes the final spatial attention map. The reason is that ReLU inevitably destroys feature representational power while the tank preserves information more smoothly.

Conversely, the tank is more prone to cause gradient to vanish as the increasing depth increases. Lastly, the convolutional block attention module generates the global context feature map to shrink channel and spatial information. Fig. 2 shows the detailed procedure of the convolutional block attention module.

B. Residual Deformable 3-D Convolution

Deformable 3-D convolution (D3C) [20] is introduced to utilize the spatio-spectral information in high-resolution remote sensing image to enhance the capability of modeling geometric transformations. The reason is that a high-resolution remote sensing image has a 3-D structure, with each point consisting of two spatial coordinates (pixel position) and a spectral coordinate (wavelength). As a result, the 3-D structure is adopted as input and carries out corresponding D3C: 3-D convolution, 3-D max pooling, and 3-D upsampling layer, as shown in Fig. 3.

C. Multiclass Focal Loss

MFL [17] expressed in the following equation is used to reduce the error caused by the imbalance of positive and

Authorized licensed use limited to: Oestfold University College. Downloaded on March 26,2023 at 11:20:14 UTC from IEEE Xplore. Restrictions apply.

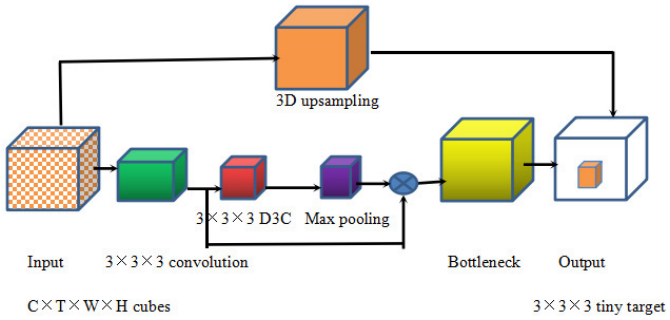

Fig. 3. Example diagram of RD3C.

TABLE I
EXPERIMENT ENVIRONMENT

Hardware	Configuration
CPU	Intel (R) Core (TM) i7-4790 K 4.OO Hz
GPU	NVIDIA GEFORCE GTX 1080
RAM	16GB
Hard disk	Toshiba SSD 5126

negative samples:

$$\mathrm{MFL}_{\mathrm{softmax}} = -\alpha_c(1-p_c)^{\gamma}\log(p_c). \tag{1}$$

The hyperparameters such as tunable focusing parameter, weighting factor field $\alpha \in [0, 1]$, model outputs category probabilities $p \in [0, 1]$, and category label c in loss function need to be determined quickly. A modified linear-weighted sum method with an additional and conditional constraint is adopted to solve the multiobjective problem. Ignoring its favorable separable structure emerging in the objective function and the constraint in (1) can be expressed as follows:

$$\mathrm{MFL}_{\mathrm{softmax}} = \begin{cases} p^{k+1} \in \underset{p\in\Re^{m\times n}}{\arg\min}\{L(p, \gamma^k, \alpha^k)\} \\ \gamma^{k+1} \in \underset{\gamma\in\Re^{m\times n}}{\arg\min}\{L(p^{k+1}, \gamma, \alpha^k)\} \\ \alpha^{k+1} = \alpha^k - \beta(p^{k+1} + \gamma^{k+1} - D). \end{cases} \tag{2}$$

Mathematically, ignoring the direct application for (2) can be made up by the well-known alternating direction method, which optimizes the variables p, γ, and α serially.

III. RESULTS

To demonstrate the validity of our model, all experimental configurations are shown in Table I. More details, 512 × 512 as input image size, 4 as batch size, and 8 as freeze size are adopted. The mean average precision (mAP) as a key performance indicator and NWPU VHR-10 [10] as the training data are manually annotated with horizontal bounding boxes. Only 655 storage tanks are available and data augmentation is essential. Thus, many transformations such as rotating 90°, 180°, and 270°, scaling 15%–25% along the y-axis, blurred, light adjustment operation, and add Gaussian and impulse noise are used to avoid overfitting. The split ratios of the training and test datasets were 90% and 10%. During training frozen or unfrozen backbone network, the tunable parameters in myNet contain *LR*, batch size, and epoch, which are set to 1e-3 or 5e-5, 16 or 8, and 100 or 1000, respectively.

TABLE II
ABLATION EXPERIMENT

EfficientDet(D0)	AM	RD3C	MFL	mAP
√	×	×	×	73.198.
√	×	×	√	74.637.

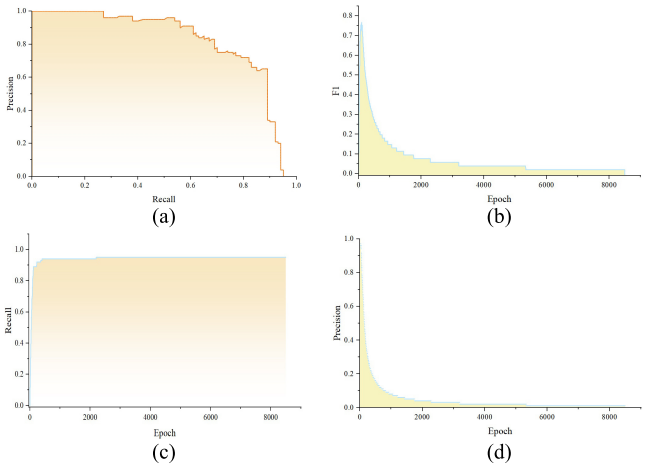

Fig. 4. (a) AUC, (b) F1, (c) precision, and (d) recall from storage tanks.

A. Ablation Studies

The results from the experiments were analyzed using the NWPU VHR-10 dataset to understand which stages are critical for detection performance in our proposed approach.

Table II describes the results from the ablation experiment with FL having little effect on detection performance. There is a significant improvement on AM + MFL with about 10% increase because AM pays attention to small targets. When AM + RD3C is adopted, detection performance decreased from 84.4530 to 83.6211 because MFL lies in speed and accuracy tradeoff on benchmarks. The results show that it can especially help improve the performance by adding AM, RD3C, and MFL.

B. Training Studies

In Fig. 4, the result from area under the curve (AUC), F1, precision, and recall reveals that the training process has a good convergence in our model. Fig. 4(a) shows that AUC as an important indicator is quite robust and reaches 97.72%. F1-score as another important indicator for precision evaluation of object recognition reaches 76.68%, as shown in Fig. 4(b). Fig. 4(c) and 4(d) shows that the changes of the precision rate and the recall rate under different epochs, respectively. The experimental results show that our model has a good performance of both precision and recall in the training process.

Authorized licensed use limited to: Oestfold University College. Downloaded on March 26,2023 at 11:20:14 UTC from IEEE Xplore. Restrictions apply.

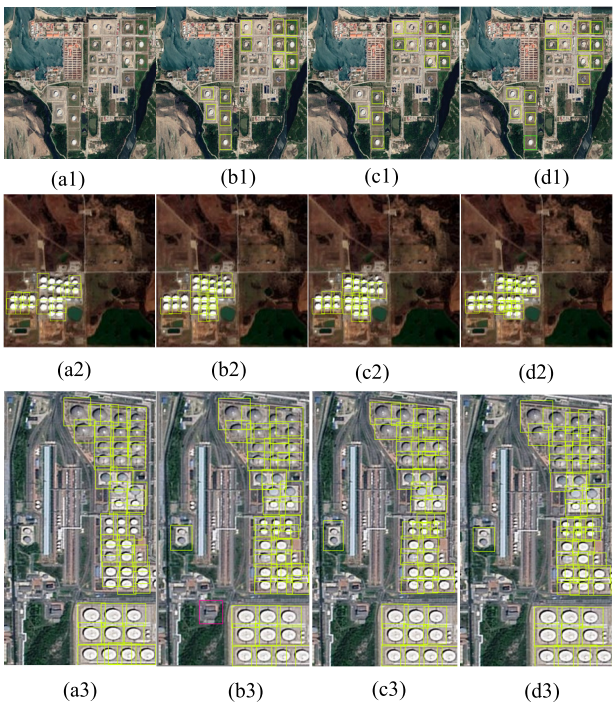

Fig. 5. Comparative results from different scenes on Google Maps: (a1)–(a3) RICNN; (b1)–(b3) SSD; (c1)–(c3) EfficientDetD0; and (d1)–(d3) proposed approach.

C. Testing Studies

To illustrate the effectiveness of the proposed approach, transferred CNN [10], RICNN [10], Sketch-R2CNN [18], ASSD [19], G-CNN [20], MT-DSSD [21], multiscale CNN [22], R-P-Faster R-CNN [23], single-shot detection (SSD) [24], cascade faster R-CNN [25], DSOD [26], and deformable R-FCN [27] were adopted.

Different results can be obtained by different target detection approaches from three scenes, oil reservoirs along the Gulf of Mexico, in Hinku, USA and at Dalian Port, PRC on Google Maps. In order to simplify the discussion here, we will only discuss Dalian Port as an example. As described from Fig. 5(a3) to (d3), RICNN [10] has worst-case performance compared with other results, and only a few storage tanks were detected, as shown in Fig. 5(a3). Conversely, SSD [24] detects 44 storage tanks but one error in the bottom right corner in Fig. 5(b3) which is marked in red. More than 40 storage tanks could be detected by EfficientDet (D0) [15] in Fig. 5(c3). Only proposed approach detects 49 storage tanks correctly, which reflect detailed information of ground objects, as shown in Fig. 5(d3).

More details of detection results from different approaches are shown in Table III. AUC increases from 66.10% in transferred CNN to 100% in the proposed approach. F1 scored remarkable effects and reached 98.65%. Precision and recall have also come to the same conclusion. Log-average miss rate (lamr) has dramatically increased from 42.90 in transferred CNN to 95.03 in the proposed approach. The total mAP as an important indicator accomplishes optimal performance at 100% in the proposed approach from 59.70% in transferred CNN [10]. In general, CNN architecture and its derivatives like RICNN [10], transferred CNN [10], Sketch-R2CNN [18], G-CNN [20], multiscale CNN [22], R-P-Faster R-CNN [23], and R-CNN [25] adopt fixed-size receptive field which is not suitable for the 3-D structure of high-resolution remote sensing image. Conversely, SSD [26] architecture and its improved schemes, such as ASSD [19] and MT-DSSD [21], have a pyramid-like structure with a generalized semantic network. However, it is too simple to remedy the depth search shortage. Moreover, DSOD [26] has somewhat flexible to learn object detectors from scratch but is not ideal for tiny targets in high-resolution remote sensing images. Deformable R-FCN [27] improves the ability to focus on pertinent image regions by increased modeling power and more robust training but supports resource-bounded devices. While EfficientDet (D0) [15] still has room for improvement, the imbalanced and complex background deceives the affinity learning on small objects. The bold numbers in Table III conclude that the proposed approach has the best overall performance.

TABLE III
COMPARATIVE EXPERIMENT (%)

Method	AUC	F1	Precision	Recall	lamr	mAP
Transferred CNN [10]	66.10	84.30	45.90	80.00	42.90	59.70
RICNN [10]	88.35	88.12	58.45	86.73	71.10	72.63
R-P-Faster R-CNN [23]	90.40	89.90	79.00	87.70	73.20	76.50
SSD [24]	90.40	89.90	82.60	98.30	76.70	78.40
MT-DSSD [21]	86.50	90.30	85.10	78.20	74.20	78.80
EfficientDet D0 [15]	97.66	96.52	90.80	100	58.60	78.99
DSOD [26]	82.70	90.10	87.80	82.10	81.20	79.80
G-CNN [20]	81.70	90.30	80.20	89.80	78.30	76.30
Deformable R-FCN [27]	87.30	90.40	81.60	90.30	75.50	79.10
Cascade Faster R-CNN [25]	90.70	89.50	89.30	97.20	88.80	84.40
ASSD [19]	99.60	95.00	94.80	95.30	86.50	89.50
Multiscale CNN [22]	99.30	97.20	92.60	98.10	85.90	89.60
Sketch-R^2CNN [18]	100	97.22	87.54	99.40	90.10	91.75
Proposed Scheme	**100**	**98.65**	**90.03**	**100**	**95.03**	**100**

IV. CONCLUSION

This letter proposes an oil tank detection approach on the improved EfficientDet model. The proposed approach weakens

Authorized licensed use limited to: Oestfold University College. Downloaded on March 26,2023 at 11:20:14 UTC from IEEE Xplore. Restrictions apply.

the influence of noise and highlights the object information by a multidimensional attention network. Moreover, RD3C as a receptive field matching can easily capture anomaly distributing in high-resolution remote sensing images. Finally, MFL is used to decrease the blindness and increase the foresight and improve the effect of model training. The experimental results support the inference with the stipulation that the proposed approach is an efficient way to improve the ability to detect oil tanks. This approach can be scaled and widely used for oil tank detection in the future.

REFERENCES

[1] A. O. Ok, "A new approach for the extraction of aboveground circular structures from near-nadir VHR satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3125–3140, Jun. 2014, doi: 10.1109/TGRS.2013.2270372.

[2] L. Zhang and C. Liu, "Oil tank detection based on linear clustering saliency analysis for synthetic aperture radar images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 2981–2985, doi: 10.1109/ICIP.2019.8803347.

[3] E. Zerman, E. Batı, G. B. Akar, E. Başeski, and Ş. Düzgün, "Circular target detection algorithm on satellite images based on radial transformation," in *Proc. 22nd Signal Process. Commun. Appl. Conf. (SIU)*, Apr. 2014, pp. 1790–1793, doi: 10.1109/SIU.2014.6830598.

[4] A. O. Ok and E. Başeski, "Circular oil tank detection from panchromatic satellite images: A new automated approach," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1347–1351, Jun. 2015, doi: 10.1109/LGRS.2015.2401600.

[5] X. Cai, H. Sui, R. Lv, and Z. Song, "Automatic circular oil tank detection in high-resolution optical image based on visual saliency and Hough transform," in *Proc. IEEE Workshop Electron., Comput. Appl.*, May 2014, pp. 408–411, doi: 10.1109/IWECA.2014.6845643.

[6] Z. Liu, D. Zhao, Z. Shi, and Z. Jiang, "Unsupervised saliency model with color Markov chain for oil tank detection," *Remote Sens.*, vol. 11, no. 9, pp. 1089–1108, 2019, doi: 10.3390/rs11091089.

[7] M. Jing, D. Zhao, M. Zhou, Y. Gao, Z. Jiang, and Z. Shi, "Unsupervised oil tank detection by shape-guide saliency model," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 477–481, Mar. 2019, doi: 10.1109/LGRS.2018.2873024.

[8] C. Zhu, B. Liu, Y. Zhou, Q. Yu, X. Liu, and W. Yu, "Framework design and implementation for oil tank detection in optical satellite imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 6016–6019, doi: 10.1109/IGARSS.2012.6352236.

[9] L. Zhang, Z. Shi, and J. Wu, "A hierarchical oil tank detector with deep surrounding features for high-resolution optical satellite imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 10, pp. 4895–4909, Oct. 2015, doi: 10.1109/JSTARS.2015.2467377.

[10] X. Yao, X. Feng, G. Cheng, J. Han, and L. Guo, "Rotation-invariant latent semantic representation learning for object detection in VHR optical remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul./Aug. 2019, pp. 1382–1385, doi: 10.1109/IGARSS.2019.8899285.

[11] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016, doi: 10.1109/TGRS.2016.2601622.

[12] K. He, X. Zhang, J. Sun, and S. Ren, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015, doi: 10.1109/TPAMI.2015.2389824.

[13] Y. Zhang, J. H. Han, Y. W. Kwon, and Y. S. Moon, "A new architecture of feature pyramid network for object detection," in *Proc. IEEE 6th Int. Conf. Comput. Commun. (ICCC)*, Dec. 2020, pp. 1224–1228, doi: 10.1109/ICCC51575.2020.9345302.

[14] E. Polat, H. M. A. Mohammed, A. N. Omeroglu, N. Kumbasar, I. Y. Ozbek, and E. A. Oral, "Multiple barcode detection with mask R-CNN," in *Proc. 28th Signal Process. Commun. Appl. Conf. (SIU)*, Oct. 2020, pp. 1–4, doi: 10.1109/SIU49456.2020.9302141.

[15] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10778–10787, doi: 10.1109/CVPR42600.2020.01079.

[16] X. Du *et al.*, "SpineNet: Learning scale-permuted backbone for recognition and localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11589–11598, doi: 10.1109/CVPR42600.2020.01161.

[17] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.

[18] L. Li, C. Zou, Y. Zheng, Q. Su, H. Fu, and C.-L. Tai, "Sketch-R2CNN: An RNN-rasterization-CNN architecture for vector sketch recognition," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 9, pp. 3745–3754, Sep. 2021, doi: 10.1109/TVCG.2020.2987626.

[19] T. Xu, X. Sun, W. Diao, L. Zhao, K. Fu, and H. Wang, "ASSD: Feature aligned single-shot detection for multiscale objects in aerial imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022, doi: 10.1109/TGRS.2021.3089170.

[20] Q. Lu, C. Liu, Z. Jiang, A. Men, and B. Yang, "G-CNN: Object detection via grid convolutional neural network," *IEEE Access*, vol. 5, pp. 24023–24031, 2017, doi: 10.1109/ACCESS.2017.2770178.

[21] R. Araki, T. Onishi, T. Hirakawa, T. Yamashita, and H. Fujiyoshi, "MT-DSSD: Deconvolutional single shot detector using multi task learning for object detection, segmentation, and grasping detection," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 10487–10493, doi: 10.1109/ICRA40945.2020.9197251.

[22] R. Jing, Z. Gong, and H. Guan, "Land cover change detection with VHR satellite imagery based on multi-scale SLIC-CNN and SCAE features," *IEEE Access*, vol. 8, pp. 228070–228087, 2020, doi: 10.1109/ACCESS.2020.3045740.

[23] B. Pradhan, M. N. Jebur, H. Z. M. Shafri, and M. S. Tehrany, "Data fusion technique using wavelet transform and Taguchi methods for automatic landslide detection from airborne laser scanning data and quickbird satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1610–1622, Mar. 2016, doi: 10.1109/TGRS.2015.2484325.

[24] L. Cai, F. Dong, K. Chen, K. Yu, W. Qu, and J. Jiang, "An FPGA based heterogeneous accelerator for single shot MultiBox detector (SSD)," in *Proc. IEEE 15th Int. Conf. Solid-State Integr. Circuit Technol. (ICSICT)*, Nov. 2020, pp. 1–3, doi: 10.1109/ICSICT49897.2020.9278177.

[25] C.-Y. Sun, X.-J. Hong, S. Shi, Z.-Y. Shen, H.-D. Zhang, and L.-X. Zhou, "Cascade faster R-CNN detection for vulnerable plaques in OCT images," *IEEE Access*, vol. 9, pp. 24697–24704, 2021, doi: 10.1109/ACCESS.2021.3056448.

[26] Z. Shen, Z. Liu, J. Li, Y.-G. Jiang, Y. Chen, and X. Xue, "DSOD: Learning deeply supervised object detectors from scratch," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1937–1945, doi: 10.1109/ICCV.2017.212.

[27] X. Zhu, H. Hu, S. Lin, and J. Dai, "Deformable ConvNets v2: More deformable, better results," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9300–9308, doi: 10.1109/CVPR.2019.00953.

[28] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate object localization in remote sensing images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2486–2498, May 2017, doi: 10.1109/TGRS.2016.2645610.

[29] J. Pang, C. Li, J. Shi, Z. Xu, and H. Feng, "$\mathcal{R}^2$-CNN: Fast tiny object detection in large-scale remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5512–5524, Aug. 2019, doi: 10.1109/TGRS.2019.2899955.

[30] X. Chen, H. Li, Q. Wu, F. Meng, and H. Qiu, "Bal-R^{2}CNN: High quality recurrent object detection with balance optimization," *IEEE Trans. Multimedia*, vol. 24, pp. 1558–1569, 2022, doi: 10.1109/TMM.2021.3067439.

[31] G. Akbarizadeh, "A new statistical-based kurtosis wavelet energy feature for texture recognition of SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4358–4368, Nov. 2012, doi: 10.1109/TGRS.2012.2194787.

[32] C. Wang, H. Zhang, F. Wu, B. Zhang, and S. Tian, "Ship classification with deep learning using COSMO-SkyMed SAR data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 558–561, doi: 10.1109/IGARSS.2017.8127014.

[33] J. Geng, H. Wang, J. Fan, and X. Ma, "Change detection of SAR images based on supervised contractive autoencoders and fuzzy clustering," in *Proc. Int. Workshop Remote Sens. with Intell. Process. (RSIP)*, May 2017, pp. 1–3, doi: 10.1109/RSIP.2017.7958819.

[34] X. Xu, B. Zou, and L. Zhang, "Polsar image classification based on polarimetric object-based morphological profiles," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 3270–3273, doi: 10.1109/IGARSS.2017.8127695.

Authorized licensed use limited to: Oestfold University College. Downloaded on March 26,2023 at 11:20:14 UTC from IEEE Xplore. Restrictions apply.