

Article

Detecting Crop Circles in Google Earth Images with Mask R-CNN and YOLOv3

Mohamed Lamine Mekhalfi ¹, Carlo Nicolò ¹, Yakoub Bazi ², Mohamad Mahmoud Al Rahhal ^{3,*} and Eslam Al Maghayreh ^{3,4}

¹ Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy; mohamed.mekhalfi@alumni.unitn.it (M.L.M.); carlo.nicolo@alumni.unitn.it (C.N.)

² Computer Engineering Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia; ybazi@ksu.edu.sa

³ Applied Computer Science Department, College of Applied Computer Science, King Saud University, Riyadh 11543, Saudi Arabia; ealmaghayreh@ksu.edu.sa

⁴ Computer Science Department, Yarmouk University, Irbid 21163, Jordan

* Correspondence: mmalrahhal@ksu.edu.sa



Citation: Mekhalfi, M.L.; Nicolò, C.; Bazi, Y.; Al Rahhal, M.M.; Al Maghayreh, E. Detecting Crop Circles in Google Earth Images with Mask R-CNN and YOLOv3. *Appl. Sci.* **2021**, *11*, 2238. <https://doi.org/10.3390/app11052238>

Academic Editors: Simone Priori and Anna Brook

Received: 23 January 2021

Accepted: 1 March 2021

Published: 3 March 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/>).

Abstract: Automatic detection and counting of crop circles in the desert can be of great use for large-scale farming as it enables easy and timely management of the farming land. However, so far, the literature remains short of relevant contributions in this regard. This letter frames the crop circles detection problem within a deep learning framework. In particular, accounting for their outstanding performance in object detection, we investigate the use of Mask R-CNN (Region Based Convolutional Neural Networks) as well as YOLOv3 (You Only Look Once) models for crop circle detection in the desert. In order to quantify the performance, we build a crop circles dataset from images extracted via Google Earth over a desert area in the East Oweinat in the South-Western Desert of Egypt. The dataset totals 2511 crop circle samples. With a small training set and a relatively large test set, plausible detection rates were obtained, scoring a precision of 1 and a recall of about 0.82 for Mask R-CNN and a precision of 0.88 and a recall of 0.94 regarding YOLOv3.

Keywords: crop circles; target detection; instance segmentation; Google Earth; deep learning; precision farming; Mask R-CNN; YOLOv3

1. Introduction

Land use and land cover are two areas that have attracted ongoing interest in the remote sensing community. In particular, it is evident that remote sensing image classification and object detection remain the most active topics so far. Further, characterizing, representing, and pinpointing targets of interest in remote sensing data are common bottlenecks.

The last decade has witnessed a swift shift from traditional shallow spectral/spatial feature manipulations to higher deep representations, which have demonstrated cutting-edge performance in spite of the relatively elevated hardware demands.

Deep learning has been tailored to many scopes in remote sensing thus far [1–5]. For instance, in [6], the well-known Convolutional Neural Network (CNN) was altered by introducing a metric learning regularization term that is meant to mitigate intra-class variations and map inter-class samples apart. In [7], a feature selection scheme based on Deep Belief Networks (DBN) is presented. Basically, the DBN is utilized as a feature reconstructor, where the most reconstructible features are selected for remote sensing scene classification. In [8], a Deep Stacking Network is applied for spatio-temporal prediction in satellite remote sensing imagery. Cloud detection in remote sensing images was also addressed in [9], where Simple Linear Iterative Clustering is used to infer superpixels from the input image. Afterward, a two-branch CNN was involved to extract multiscale features and categorize the pixel of interest into thin cloud, thick cloud, or noncloud. A

patch-to-patch mapping was implemented within a deep learning architecture for remote sensing image registration in [10]. In [11], ternary change detection in Synthetic Aperture Radar data is addressed, where an autoencoder was used to learn meaningful features, followed by a three-group clustering. The resulting features are fed into a CNN for change classification.

Owing to viewpoint, rotation and scale changes on the one hand and the lack of abundant data (i.e., often manual annotation may be somewhat laborious) on the other, object detection in remote sensing images remains harder to approach than the typical scene classification. However, as per object detection, the literature suggests fewer contributions. On this point, the work in [12] combines a multiscale object proposal network with an accurate object detection network by fusing feature maps, and satisfactory results were yielded. In order to tackle the rotation invariance issue that may be encountered with CNNs, in [13], a rotation-invariant layer was trained prior to fine-tuning the CNN model. This enables drawing the training samples close to each other before and after rotation. To cope with scale change, a scale-adaptive architecture was presented in [14], where multilayer region proposal networks are envisioned, topped with a fusion network for object detection. Another rotation-invariant network was devised in [15], where multiangle anchors are incorporated into the region proposal network. Further, a fusion network was suggested to learn contextual cues. In [16], a saliency mechanism was implemented within a deep belief network to locate potential objects coarsely within a probe image. This coarse object proposal has demonstrated its ability to lessen exhaustive search for objects.

In this respect, deep learning has also found its way into other applications such as precision farming. In [17], transfer learning (based on a VGG-16 pretrained network) was applied to identify mildew disease in a pearl millet. In [18], summer crop classification was addressed based on Landsat Enhanced Vegetation Index by means of a long short-term memory network and uni-dimensional CNN, among other state-of-the-art classifiers. It turned out that 1D-CNN performs the best. A comparative study of three pre-trained deep learning models was carried out in [19] for water stress assessment in maize (*Zea mays*), okra (*Abelmoschus esculentus*) and soybean (*Glycine max*). Specifically, AlexNet, GoogLeNet and Inception V3 were confronted on a 1200 optical image dataset. Thus, GoogLeNet stands out among the three models. In [20], AlexNet and GoogLeNet were evaluated on a dataset amounting to 54,306 images pertaining to 14 crop species and 26 plant diseases. Although both of the networks perform fairly equally, GoogLeNet yields slightly higher scores under a transfer learning scenario. However, the score gap increases when the networks are trained from scratch. In [21], a digital surface model was combined with a radiometric index image and fed as input to a CNN for soil and crop segmentation in remote sensing data.

Crop yield estimation/counting via artificial intelligence holds a great interest in precision agriculture. For instance, it is useful for the farmers in order to allocate essential logistics such as transportation means and labor force. This implies that overestimating the yield may raise management costs, whereas in the case of underestimation, crops may be subject to waste if not transferred to storage facilities shortly after the harvest. In [22], a computer vision system was devised for kiwifruit yield estimation. The system incorporates a lightweight optical sensor mounted on a tractor that drives along the kiwi plants for image acquisition. Regarding the fruit detection part, a cascaded Haar feature-based classifier is adopted. In [23], cotton yield estimation is approached by considering several features fed into an Artificial Neural Network (ANN) as input, namely multi-temporal features (i.e., including canopy cover, canopy height, canopy volume, normalized difference vegetation index, excessive greenness index) and non-temporal features (i.e., cotton boll count, boll size and boll volume), supplemented with irrigation status. Another ANN that takes as input multi-temporal features (e.g., canopy cover) and weather information for tomato yield estimation was presented in [24]. A review on machine learning in agriculture is given in [25–31].

Crop circles have been evolving as a large-scale farming scheme, especially in countries consisting mostly of desert. Crop circles are normally planned around areas in the desert with a large body of underground water reserves. Their circular shape is owed to the irrigation practice, commonly referred to as pivot irrigation, mainly adopted for optimizing water use. Automatic crop circle detection is necessary for a precise, low-cost and timely management of farming land in remote areas in the desert.

In this context, this article investigates the use of deep learning for crop circle detection in the desert. We selected Mask R-CNN and YOLOv3 models, owing to their performance in various applications. Mask R-CNN offers the possibility to carry out instance segmentation of the target and has demonstrated plausible performance in several applications. For instance, in [32], U-Net and Mask R-CNN were combined for nuclei segmentation in fluorescence and histology images, where a precision of 0.72 and a recall of 0.6 were obtained. In [33], Mask R-CNN was adopted for strawberry detection, and a precision of 0.96 and a recall of 0.95 were achieved. In [34], vehicle damage detection was addressed by means of Mask R-CNN with a detection accuracy of 94.53%. Knee meniscus tear detection was carried out via Mask R-CNN in [35], scoring a weighted AUC score of 0.906. In [36], melanoma skin cancer detection was tackled by combining Mask R-CNN for initial segmentation, followed by DenseNet for feature extraction, topped by least square support vector machines to draw the final result, where the accuracy ranged between 88.5% and 96.3%. YOLOv3 has been widely considered for object detection. For instance, in [37], it was compared with Faster R-CNN for car detection in unmanned aerial vehicle images. It turned out that YOLOv3 is more accurate than R-CNN, with a precision and recall of 0.99 and 0.99, respectively. Apple lesion detection was addressed by means of a YOLOv3-based architecture with an accuracy of 95.57% in [38]. Apple detection was also achieved based on YOLOv3-dense, with an F1 score of 0.817 [39]. YOLOv3 was applied in [40] for traffic flow detection, with an accuracy of 95.01%. YOLOv3 was adopted for collapsed building detection in [41] with a precision of 0.88 and a recall of 0.78. Thus, owing to their generalization potential in various applications, Mask R-CNN and YOLOv3 models were opted for in this work for crop circle detection in the desert.

For evaluation purposes, we built a large dataset of 2511 crop circle samples from images obtained from Google Earth over the desert in the south of Egypt. Satisfactory results were obtained on a small training set (i.e., mimicking a real scenario where abundant training data are often inaccessible).

This paper is outlined as follows. Section 2 presents the adopted crop circle detection models. Section 3 describes the dataset and discusses the results. Section 4 draws conclusions.

2. Proposed Methodology

The next two subsections provide a brief description of the adopted Mask R-CNN and YOLOv3 architectures.

2.1. Mask R-CNN

The implemented Mask R-CNN extends the well-known Faster R-CNN, which in turn is an extension of the Region-based CNN but introduces an attention mechanism through a region proposal network (RPN). This latter suggests candidate bounding boxes pertaining to the object of interest. Afterward, features are extracted from each candidate bounding box using RoIPool, followed by classification and bounding box regression [42].

In Mask R-CNN, the RPN part is retained, allowing potential object bounding boxes to be detected [43]. In the second stage, classification and bounding box regression are supplemented with a binary mask for each region of interest. Thus, for training purposes, the network follows a global loss, which sums up the classification loss, the bounding box loss and the mask (segmentation) loss:

$$L = L_{class} + L_{bbox} + L_{mask} \quad (1)$$

The first two terms of the total loss are the same as in Faster R-CNN [21]. Therefore, Mask R-CNN can be regarded as a Faster R-CNN reinforced with an instance segmentation feature. In brief, Mask R-CNN incorporates the following key steps:

Feature maps are drawn from the image presented at the input with a CNN model (we selected ResNet-50).

The region proposal network is adopted to generate multiple regions of interest by incorporating a CNN and a binary classifier. In return, object scores are obtained. In order to eliminate the anchor boxes that likely belong to the background, intersection over union mechanism is invoked, where only the anchor boxes with an intersection above 0.5 are considered. In order to tackle the bottleneck of picking up multiple bounding boxes over the threshold of 0.5, non-max suppression is introduced by selecting the bounding box with the highest intersection over union and discarding the remaining candidate boxes.

A region of interest aligns the network and then produces multiple boxes of evidence of the same object and warps them into a fixed dimension.

Afterward, the warped features are fed into fully connected layers to carry out the classification using a Softmax. The boundary box is further refined with the regression model.

Further, for instance segmentation, the warped features are also fed into a mask classifier that consists of two CNNs to produce a mask for each region of interest. The pipeline of Mask R-CNN is depicted in Figure 1.

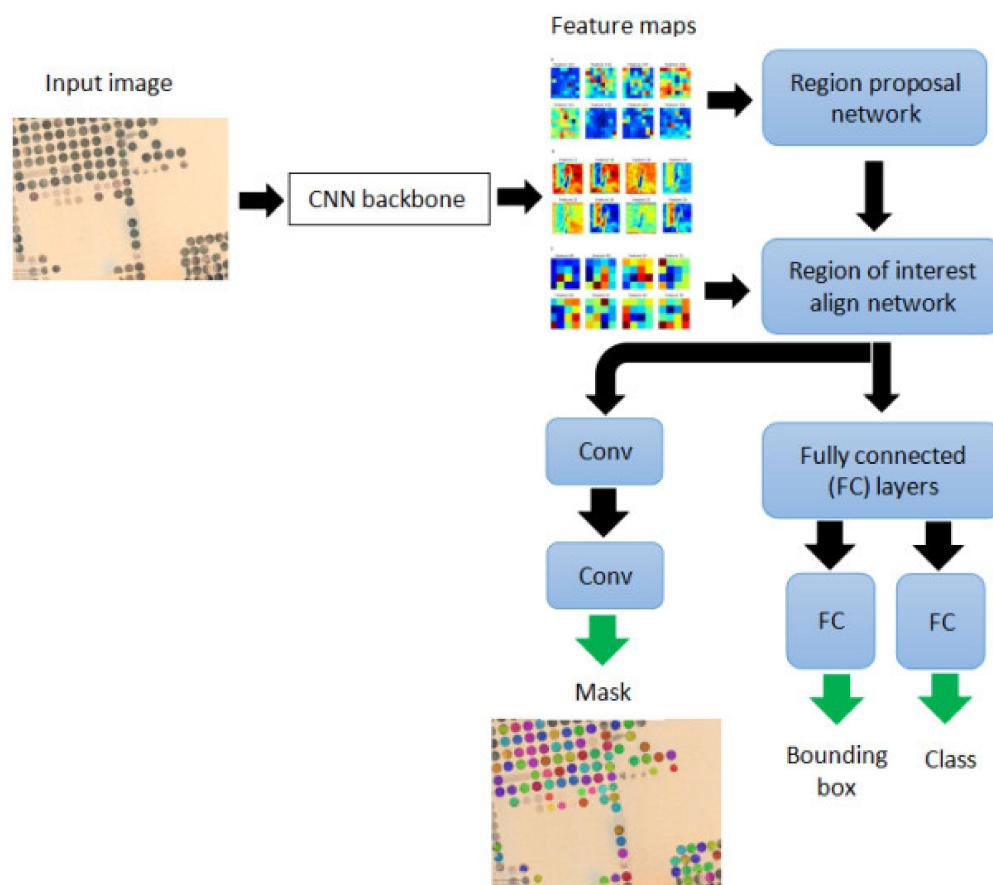


Figure 1. Mask R-CNN architecture for crop circle detection.

2.2. YOLOv3

You Only Look Once is a family of single-shot object detectors. YOLOv3 improves over the second version in the sense that it incorporates multi-scale detection, improves the loss function and makes use of a stronger feature extraction network [44].

YOLOv3 runs over two blocks, namely a feature extractor and an object detector. Object scale change is tackled by extracting the features at three different scales.

Feature extraction is performed via Darknet-53, which consists of 53 layers. Features from the last three residual blocks are considered for further multi-scale object detection. This latter is formed of several 1×1 and 3×3 Conv layers, topped with a 1×1 Conv layer to produce the final output. Moreover, large-scale features are concatenated with medium-scale features, and these latter are concatenated with small-scale features, which enables the small-scale features to avail the results of large-scale features.

Object search is addressed by means of a set of anchor boxes over a grid of the image, which all share the same centroid, and the box that shares the highest intersection over union with the ground truth is selected.

As per the loss function, four terms are envisioned, namely the objectness loss, width and height loss, centroid loss, and classification loss. We choose the architecture depicted in Figure 2.

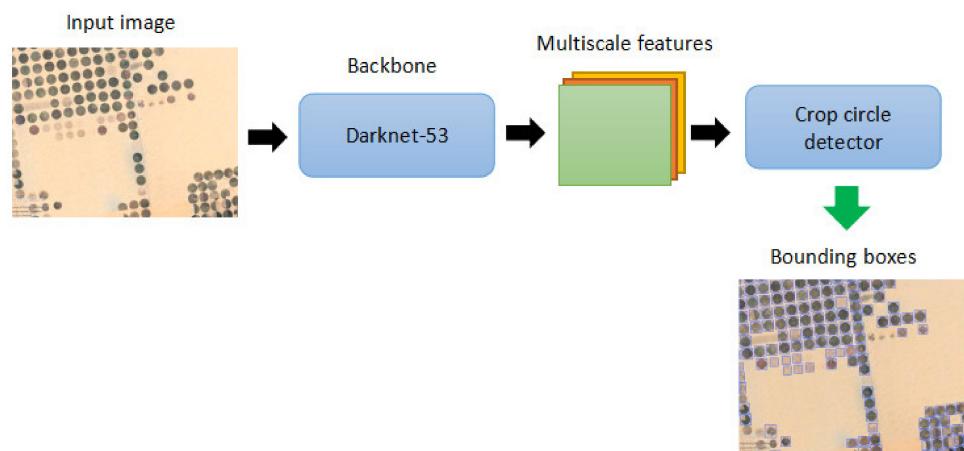


Figure 2. Adopted YOLOv3-based crop circle detection network.

3. Experimental Results

The experiments were carried out on an image dataset corresponding to the East Oweinat, in the South Western Desert of Egypt ($22^{\circ}39'00.7''$ N $28^{\circ}48'47.7''$ E). The images were exported from Google Earth Pro at an altitude of 20 km. The overall view of the study area and sample images from the dataset are illustrated in Figure 3.

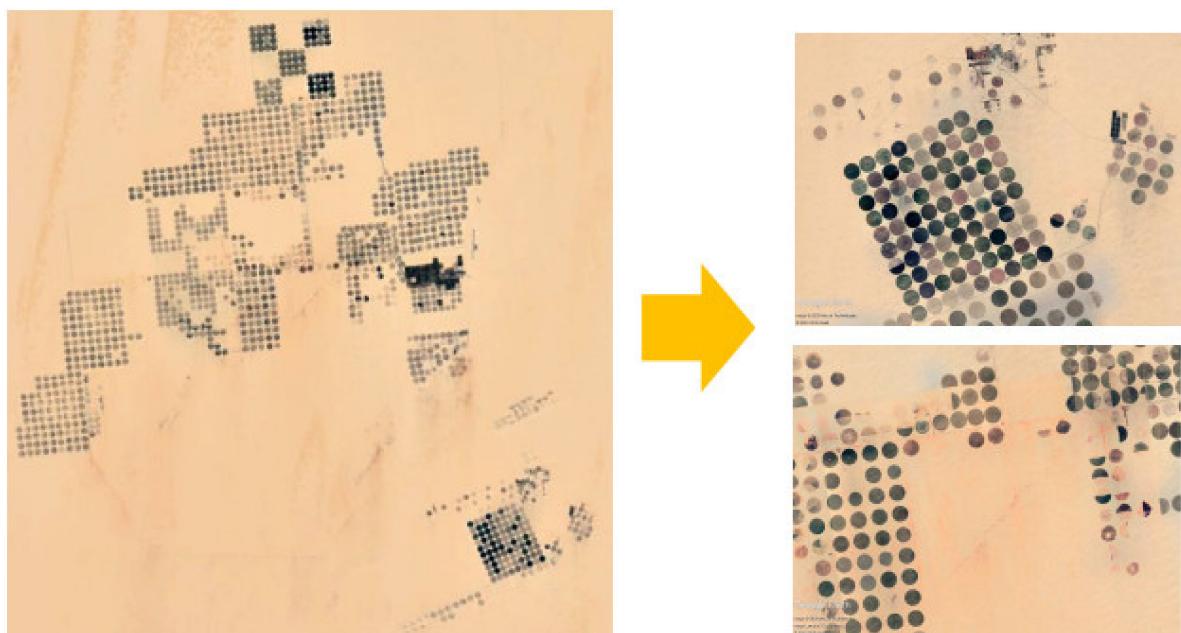


Figure 3. Left: study area. Right: Sample images at 20 km altitude (Source: Google Earth Pro).

The dataset totals 24 images, which were split into 4 images for training, 4 images for validation and 16 images for testing purposes. In terms of crop circles, it contains 437 training samples, 519 validation samples and 1555 test samples, making up 17%, 21% and 62% of the total count, respectively. It is worth noting that we adopt a large test set and a small training set in order to mimic a real scenario, where the number of data are limited.

Image labeling was carried out by means of the Visual Geometry Group (VGG) image annotator [45]. Note that the crop circles that are not visually clear to the naked eye, or that are partially occluded (e.g., only half of the crop circle is visible, or that the crop circle lies at the border of the image), are omitted during the labeling process.

The training was performed via the Google Colab platform on a Tesla T4 GPU with 16GB memory and a 16 GB RAM; the training rates were set to 0.001 and the remaining parameters were set as mentioned in the original papers of both models. In order to quantify the results, we followed the state-of-the-art object detection metrics, namely the Precision and the Recall. The precision measures the ratio of detected objects that are relevant (i.e., how many crop circles are detected among all the detected objects), whilst the recall measures the ratio of relevant crop circles that were successfully detected, as follow:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (2)$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (3)$$

where *True Positives* indicates the crop circles samples that were indeed identified as crop circles, *False Positives* refers to non-crop circle objects that were mistakenly identified as crop circles and *False Negatives* refers to crop circles that were not identified.

It can be seen from Table 1 that both models yield satisfactory results. However, while Mask R-CNN tends to be more precise, YOLOv3 favors detection over precision. This can be explained by the fact that Mask R-CNN inherently integrates the mask attribute into the learning process, which allows it to learn finer details of the target, yet it is less likely to infer irrelevant objects. However, this comes at the cost of more processing overheads as it runs nearly three times slower. YOLOv3, on the other hand, emphasizes object search over several scale grids, paying more attention to the existence of the target object in the context of the input image, which is translated into a high Recall and a relatively lower precision, while cutting down sharply on the inference time.

Table 1. Crop circles detection results.

	Precision	Recall	Inference Time/Image (s)
Mask R-CNN	1.00	0.82	3
YOLOv3	0.88	0.94	0.9

Detection instances are depicted in Figures 4 and 5, for Mask R-CNN and YOLOv3, respectively. It can be observed that both of the models become less accurate when the crop circles share nearly the same color with the background which constitutes a bottleneck for most object detection endeavors. Further, the detection seems to be achieved rather easily when the crop circle manifests uniform texture/color cues. However, an advantage of Mask R-CNN is that it performs an instance segmentation, providing thereby the precise segment of the crop circle, which suggests a potential farming area measurement. This property can not be availed with the YOLO model as the bounding box does not fit perfectly the crop circle. In such a case, tying the multiscale object detection of YOLO with the mask learning property of Mask R-CNN suggests an ad hoc solution for accurate as well as precise crop circle detection (e.g., fusion of detection priors of both models based on their confidence [46]).

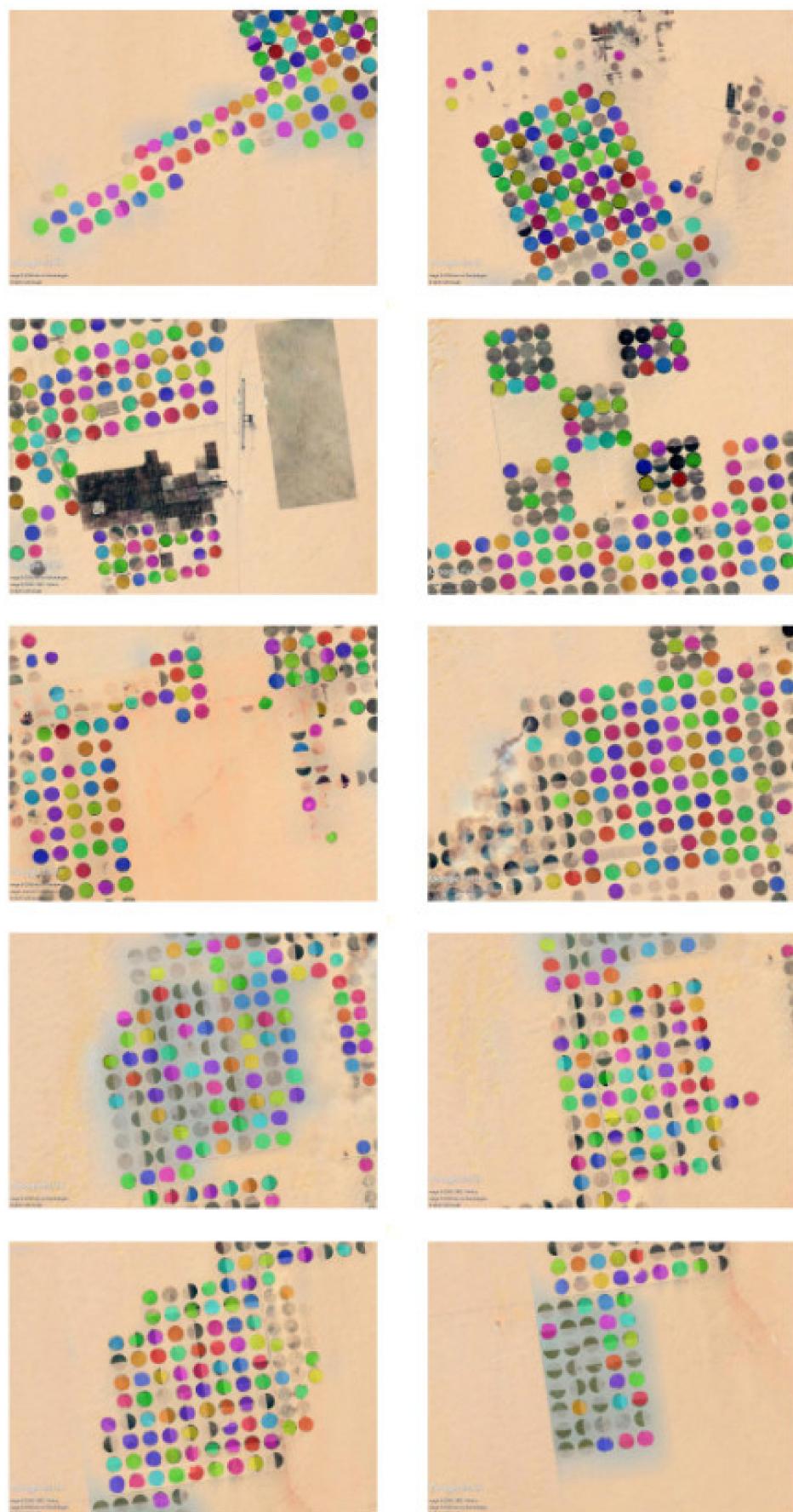


Figure 4. Several inference instances using Mask R-CNN.

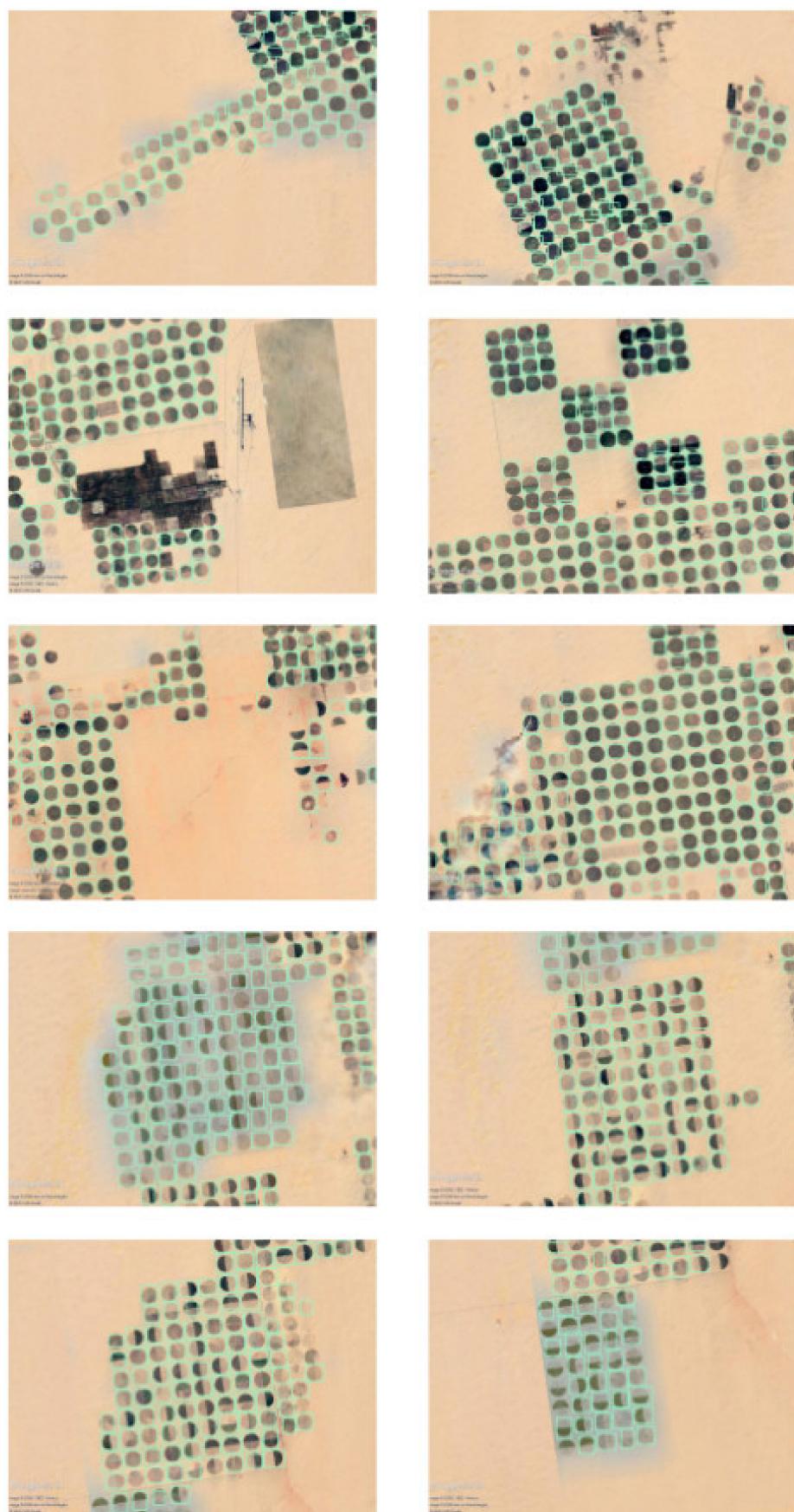


Figure 5. Several inference instances using YOLOv3.

4. Conclusions

This paper investigated the use of deep learning for crop circle detection in the desert. A crop circle dataset was built from Google Earth images at 20 km altitude over the East Oweinat area in the South of Egypt. In particular, we opted for Mask R-CNN and YOLOv3 as detection baselines.

The experiments showed that YOLOv3 tends to detect more crop circles while compromising on precision, while Mask R-CNN is by far more precise, although less accurate. In terms of inference time, YOLOv3 remains much faster.

A potential improvement would be for the mask learning property of the Mask R-CNN to be coupled with the object detection tendency of YOLOv3.

We are currently working on increasing the size of the dataset and possibly supplementing it with a second dataset for an in-depth analysis of crop circle detection.

Author Contributions: M.L.M., Conceptualization, methodology, validation, formal analysis, writing—review and editing, supervision. C.N., resources, data curation, validation, formal analysis. Y.B., Conceptualization, formal analysis, investigation, supervision. M.M.A.R., formal analysis, review and editing, investigation. E.A.M., formal analysis, review, visualization. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Deanship of Scientific Research, King Saud University, grant number RG-1441-502.

Data Availability Statement: Not Applicable.

Acknowledgments: The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding this work through research group No. RG-1441-502.

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

References

1. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [[CrossRef](#)]
2. Alhawiti, H.; Bazi, Y.; Rahhal, M.M.A.; Alhichri, H.; Zuair, M.A. Deep Learning Approach for Multiple Source Classification in Remote Sensing Imagery. In Proceedings of the 2020 3rd International Conference on Computer Applications Information Security (ICCAIS), Riyadh, Saudi Arabia, 19–21 March 2020; pp. 1–4.
3. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep Learning in Environmental Remote Sensing: Achievements and Challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [[CrossRef](#)]
4. Tsagkatakis, G.; Aidini, A.; Fotiadou, K.; Giannopoulos, M.; Pentari, A.; Tsakalides, P. Survey of Deep-Learning Approaches for Remote Sensing Observation Enhancement. *Sensors* **2019**, *19*, 3929. [[CrossRef](#)]
5. Rahhal, M.M.A.; Bazi, Y.; Al-Hwiti, H.; Alhichri, H.; Alajlan, N. Adversarial Learning for Knowledge Adaptation from Multiple Remote Sensing Sources. *IEEE Geosci. Remote Sens. Lett.* **2020**, *1*–5. [[CrossRef](#)]
6. Cheng, G.; Yang, C.; Yao, X.; Guo, L.; Han, J. When Deep Learning Meets Metric Learning: Remote Sensing Image Scene Classification via Learning Discriminative CNNs. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2811–2821. [[CrossRef](#)]
7. Zou, Q.; Ni, L.; Zhang, T.; Wang, Q. Deep Learning Based Feature Selection for Remote Sensing Scene Classification. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2321–2325. [[CrossRef](#)]
8. Das, M.; Ghosh, S.K. Deep-STEP: A Deep Learning Approach for Spatiotemporal Prediction of Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1984–1988. [[CrossRef](#)]
9. Xie, F.; Shi, M.; Shi, Z.; Yin, J.; Zhao, D. Multilevel Cloud Detection in Remote Sensing Images Based on Deep Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3631–3640. [[CrossRef](#)]
10. Wang, S.; Quan, D.; Liang, X.; Ning, M.; Guo, Y.; Jiao, L. A Deep Learning Framework for Remote Sensing Image Registration. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 148–164. [[CrossRef](#)]
11. Gong, M.; Yang, H.; Zhang, P. Feature Learning and Change Feature Classification Based on Deep Learning for Ternary Change Detection in SAR Images. *ISPRS J. Photogramm. Remote Sens.* **2017**, *129*, 212–225. [[CrossRef](#)]
12. 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](#)]
13. Cheng, G.; Zhou, P.; Han, J. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7405–7415. [[CrossRef](#)]
14. Zhang, S.; He, G.; Chen, H.; Jing, N.; Wang, Q. Scale Adaptive Proposal Network for Object Detection in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 864–868. [[CrossRef](#)]

15. Li, K.; Cheng, G.; Bu, S.; You, X. Rotation-Insensitive and Context-Augmented Object Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2337–2348. [[CrossRef](#)]
16. Diao, W.; Sun, X.; Zheng, X.; Dou, F.; Wang, H.; Fu, K. Efficient Saliency-Based Object Detection in Remote Sensing Images Using Deep Belief Networks. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 137–141. [[CrossRef](#)]
17. Coulibaly, S.; Kamsu-Foguem, B.; Kamissoko, D.; Traore, D. Deep Neural Networks with Transfer Learning in Millet Crop Images. *Comput. Ind.* **2019**, *108*, 115–120. [[CrossRef](#)]
18. Zhong, L.; Hu, L.; Zhou, H. Deep Learning Based Multi-Temporal Crop Classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [[CrossRef](#)]
19. Chandel, N.S.; Chakraborty, S.K.; Rajwade, Y.A.; Dubey, K.; Tiwari, M.K.; Jat, D. Identifying Crop Water Stress Using Deep Learning Models. *Neural Comput. Appl.* **2020**. [[CrossRef](#)]
20. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using Deep Learning for Image-Based Plant Disease Detection. *Front. Plant. Sci.* **2016**, *7*. [[CrossRef](#)] [[PubMed](#)]
21. Dyson, J.; Mancini, A.; Frontoni, E.; Zingaretti, P. Deep Learning for Soil and Crop Segmentation from Remotely Sensed Data. *Remote Sens.* **2019**, *11*, 1859. [[CrossRef](#)]
22. Mekhalfi, M.L.; Nicolò, C.; Ianniello, I.; Calamita, F.; Goller, R.; Barazzuol, M.; Melgani, F. Vision System for Automatic On-Tree Kiwifruit Counting and Yield Estimation. *Sensors* **2020**, *20*, 4214. [[CrossRef](#)]
23. Ashapure, A.; Jung, J.; Chang, A.; Oh, S.; Yeom, J.; Maeda, M.; Maeda, A.; Dube, N.; Landivar, J.; Hague, S.; et al. Developing a Machine Learning Based Cotton Yield Estimation Framework Using Multi-Temporal UAS Data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 180–194. [[CrossRef](#)]
24. Ashapure, A.; Oh, S.; Marconi, T.G.; Chang, A.; Jung, J.; Landivar, J.; Enciso, J. Unmanned Aerial System Based Tomato Yield Estimation Using Machine Learning. In Proceedings of the Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping IV, Baltimore, MD, USA, 14 May 2019; Volume 11008, p. 110080O.
25. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. *Sensors* **2018**, *18*, 2674. [[CrossRef](#)]
26. Chlingaryan, A.; Sukkarieh, S.; Whelan, B. Machine Learning Approaches for Crop Yield Prediction and Nitrogen Status Estimation in Precision Agriculture: A Review. *Comput. Electron. Agric.* **2018**, *151*, 61–69. [[CrossRef](#)]
27. Crane-Droesch, A. Machine Learning Methods for Crop Yield Prediction and Climate Change Impact Assessment in Agriculture. *Environ. Res. Lett.* **2018**, *13*, 114003. [[CrossRef](#)]
28. Mekonnen, Y.; Namuduri, S.; Burton, L.; Sarwat, A.; Bhansali, S. Review—Machine Learning Techniques in Wireless Sensor Network Based Precision Agriculture. *J. Electrochem. Soc.* **2019**, *167*, 037522. [[CrossRef](#)]
29. Sharma, R.; Kamble, S.S.; Gunasekaran, A.; Kumar, V.; Kumar, A. A Systematic Literature Review on Machine Learning Applications for Sustainable Agriculture Supply Chain Performance. *Comput. Oper. Res.* **2020**, *119*, 104926. [[CrossRef](#)]
30. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep Learning in Agriculture: A Survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [[CrossRef](#)]
31. Drury, B.; Valverde-Rebaza, J.; Moura, M.-F.; de Andrade Lopes, A. A Survey of the Applications of Bayesian Networks in Agriculture. *Eng. Appl. Artif. Intell.* **2017**, *65*, 29–42. [[CrossRef](#)]
32. Vuola, A.O.; Akram, S.U.; Kannala, J. Mask- and U-net ensembled for nuclei segmentation. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging, Venice, Italy, 8–11 April 2019; pp. 208–212.
33. Yu, Y.; Zhang, K.; Yang, L.; Zhang, D. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Comput. Electron. Agric.* **2019**, *163*, 104846. [[CrossRef](#)]
34. Zhang, Q.; Chang, X.; Bian, S.B. Vehicle-damage-detection segmentation algorithm based on improved mask RCNN. *IEEE Access* **2020**, *8*, 6997–7004. [[CrossRef](#)]
35. Couteaux, V.; Si-Mohamed, S.; Nempong, O.; Lefevre, T.; Popoff, A.; Pizaine, G.; Villain, N.; Bloch, I.; Cotten, A.; Boussel, L. Automatic knee meniscus tear detection and orientation classification with Mask-RCNN. *Diagn. Interv. Imaging* **2019**, *100*, 235–242. [[CrossRef](#)] [[PubMed](#)]
36. Khan, M.A.; Akram, T.; Zhang, Y.D.; Sharif, M. Attributes based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework. *Pattern Recognit. Lett.* **2021**, *143*, 58–66. [[CrossRef](#)]
37. Benjdira, B.; Khursheed, T.; Koubaa, A.; Ammar, A.; Ouni, K. Car detection using unmanned aerial vehicles: Comparison between faster r-cnn and yolov3. In Proceedings of the 2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS), Muscat, Oman, 5–7 February 2019; pp. 1–6.
38. Tian, Y.; Yang, G.; Wang, Z.; Li, E.; Liang, Z. Detection of Apple Lesions in Orchards Based on Deep Learning Methods of CycleGAN and YOLOV3-Dense. *J. Sens.* **2019**, *2019*, 1–13. [[CrossRef](#)]
39. Tian, Y.; Yang, G.; Wang, Z.; Wang, H.; Li, E.; Liang, Z. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Comput. Electron. Agric.* **2019**, *157*, 417–426. [[CrossRef](#)]
40. Huang, Y.Q.; Zheng, J.C.; Sun, S.D.; Yang, C.F.; Liu, J. Optimized YOLOv3 algorithm and its application in traffic flow detections. *Appl. Sci.* **2020**, *10*, 3079. [[CrossRef](#)]
41. Ma, H.; Liu, Y.; Ren, Y.; Yu, J. Detection of Collapsed Buildings in Post-Earthquake Remote Sensing Images Based on the Improved YOLOv3. *Remote Sens.* **2019**, *12*, 44. [[CrossRef](#)]
42. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)]

43. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2961–2969.
44. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
45. Dutta, A.; Zisserman, A. The VIA Annotation Software for Images, Audio and Video. In Proceedings of the 27th ACM International Conference on Multimedia—ACM, Nice, France, 15 October 2019; pp. 2276–2279.
46. Mirmahboub, B.; Mekhalfi, M.L.; Murino, V. Person re-identification by order-induced metric fusion. *Neurocomputing* **2018**, *275*, 667–676. [[CrossRef](#)]