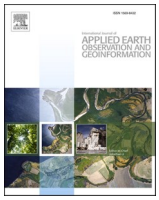
International Journal of Applied Earth Observations and Geoinformation 112 (2022) 102875

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Mask R-CNN based automated identification and extraction of oil well sites

Hongjie He [a], Hongzhang Xu [a], Ying Zhang [b], Kyle Gao [a], Huxiong Li [c], Lingfei Ma [d,*], Jonathan Li [a,e]

[a] Geospatial Sensing and Data Intelligence Laboratory, Department of Geography and Environmental Management, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada
[b] Canada Centre for Remote Sensing, Natural Resources Canada, 560 Rochester Street, Ottawa, Ontario K1S 5H4, Canada
[c] Department of Computer Science and Engineering, Shaoxing University, Shaoxing 312000, China
[d] School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 102206, China
[e] Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada

ARTICLE INFO

Keywords:
Land disturbance
Oil well sites
OWS Mask R-CNN
Multi sensors
RCAN

ABSTRACT

Fine-scale land disturbances due to mining development modify the land surface cover and have cumulative detrimental impacts on the environment. Understanding the distribution of fine-scale land disturbances related to mining activities, such as oil well sites, in mining regions is of vital importance to sustainable mining development. For efficient mapping, automated identification and extraction of the oil well sites using high-resolution satellite images are required. In this work, we proposed the Oil Well Site extraction (OWS) Mask R-CNN based on the original Mask R-CNN (Region-based Convolutional Neural Networks), to accurately extract well sites using multi-sensor remote sensing images. For improvement of mapping efficiency, two modifications were made to Mask R-CNN: (1) replacing the backbone of Mask R-CNN with D-LinkNet, and (2) adding a semantic segmentation branch to Mask R-CNN to force the whole network to focus on the relationship between line objects and oil well sites. As imagery data were from multiple sensors (RapidEye 2/3 and WorldView 3), a pre-trained Residual Channel Attention Network (RCAN) was applied to super-resolve the images with different resolutions. Several key spatial features, such as nearby roads and area size, have also been used in the oil well site mapping process. The experimental results indicate that our OWS Mask R-CNN considerably improves the average precision (AP) and the F_1 score of Mask R-CNN from 51.26% and 25.7% to 60.93% and 61.59%, respectively.

1. Introduction

Oil or natural gas production is important for economic development around the world. However, oil and gas mining developments inevitably have detrimental impacts on the environment. The most used technology for oil and gas production is through the construction of in-situ drilling sites, where the development of a series of well sites and other facilities is needed (Zhang et al., 2018a). The in-situ mining with footprints of well sites and resource roads in a massive number can certainly lead to landscape transformation, landscape fragmentation, and result in cumulative impacts on the environment (Yang et al., 2018). Taking oil sands production in Alberta, Canada as an example, the in-situ oil sands mining development in Alberta has undergone a period of rapid expansion in the past few years, which plays a key role in Alberta's, even Canada's economy (Gosselin et al., 2010). However, there is a rise of concern about environmental impacts resulting from land disturbances related to oil sands production. A number of studies have demonstrated that land disturbances can have long-term effects on the population of various species, ranging from songbirds to carnivores (Bayne et al., 2005; Dyer et al., 2001; Machtans, 2006; Nielsen et al., 2007). Therefore, the identification, mapping and monitoring of land disturbances and potential risks associated with mining activities at local and regional scales becomes a requirement for sustainable mining development (Erzurumlu and Erzurumlu, 2015).

Earth observation and remote sensing techniques have provided an effective approach to extracting information about land surface disturbance footprints to support environmental assessment and risk analysis related to mining activities of oil sands production (Yang et al., 2018). In

* Corresponding author.
E-mail addresses: h69he@uwaterloo.ca (H. He), h389xu@uwaterloo.ca (H. Xu), ying.zhang@NRCan-RNCan.gc.ca (Y. Zhang), junli@uwaterloo.ca (K. Gao), jsj_lhx@126.com (H. Li), l53ma@cufe.edu.cn (L. Ma), junli@uwaterloo.ca (J. Li).

https://doi.org/10.1016/j.jag.2022.102875
Received 5 January 2022; Received in revised form 15 June 2022; Accepted 16 June 2022
Available online 24 June 2022

Fig. 1. High-definition images acquired from Google Earth.

previous operational mapping, visual interpretation became a conventional method to extract oil well sites from remote sensing images. However, this method is time-consuming and heavily relies on professional knowledge (Yang et al., 2018). Hence, automated extraction of oil well sites was urgently needed. Moreover, the manual extraction method was inefficient for processing a massive volume of data. In recent years, machine learning techniques have prevailed in the remote sensing community due to their high processing performance and low reliance on human knowledge. Methods like decision trees (DT) and other traditional machine learning methods were widely used in remote sensing applications for classification and regression tasks (Schulz et al., 2018). Features used in these methods were commonly constructed using expert knowledge. In other words, feature learning in deep learning (DL) instead of feature engineering in conventional machine learning automated the entire process (Ok, 2013). Moreover, with sophisticated innovations, deep learning produced end-to-end architectures to solve complex problems with higher accuracy when trained on a huge amount of data (Ball et al., 2017). Deep learning-based algorithms are widely used in a variety of remote sensing applications, such as building detection (Sun et al., 2017), road segmentation (Zhang et al., 2018c), and topological map generation (Ma and Zhao, 2017). Therefore, deep learning-based methods are promising for oil site extraction.

Different from building footprints, road networks, and other small objects, such as vehicles, ships, and oil tanks, oil well sites are hard to extract because of the diversity of geometric and spectral features. Because of oil and gas production, the oil well sites are always shaped irregularly as shown in Fig. 1. In addition, because of vegetation recovery after oil and gas production, vegetation will cover the oil well sites, which makes it harder to detect and extract oil well sites. Therefore, while DL methods were widely used in remote sensing applications, there were very few works for the applications of deep learning-based methods in oil industrial facility extraction. In the work of Zhang et al. (2018a), the method You Look Only Once (YOLO) v2 was applied to detect the location of oil well sites and achieved an accuracy of 92% in bounding box detection results. Faster R-CNN (Regional Convolutional Neural Network), a two-stage detector, was also used for detecting oil wells from remote sensing images (Wang et al., 2021), which achieved a high recall of 92.4% (Song et al, 2020). However, the detailed mask of well site footprint is also important for scientific analysis of land disturbance. Therefore, in addition to locating oil well sites, extracting masks of oil well sites should be taken into account in automated oil well sites detection and extraction. Instance segmentation methods from computer vision fit perfectly for the purpose, which detect the locations of target objects and extract the masks of target objects; but these methods have not yet been applied in oil well site detection and extraction. In our work, we developed an oil well site extraction method based on Mask R-CNN, which detects not only the location (bounding boxes) of oil well sites but also their masks. The motivations of this work are 1) to develop a new instance segmentation-based oil well sites extraction and 2) to build the connection between oil well sites and road networks in the detection process.

The contribution of this work is two folds:

(1) A new algorithm, OWS Mask R-CNN, for oil well site identification and extraction from remote sensing images.
(2) A new framework to extract oil well sites including the fusion and post-processing of multi-modality images.

In this paper, a review of commonly used deep learning-based methods for object extraction is given in Section 2. In Section 3 the study area and data used in this work are described. In Section 4, our proposed OWS Mask R-CNN method and metrics used for performance evaluation are described in detail. The results from the experiments are presented in Section 5. In Section 6, we conclude the paper with our findings from experiments.

2. Literature review

2.1. The development of instance segmentation.

Automated extraction of oil/gas well sites from satellite images falls into the tasks of object detection and instance segmentation. The former

aims at localizing and classifying objects in images and output bounding boxes, confidence score of localization, and class of each object. The latter in addition produces segmentation masks for each detected object. Both object detection and instance segmentation are key tasks in computer vision. In this section, we briefly introduce object detection and instance segmentation methods recently proposed in the computer vision field. Deep learning-based object detection methods and instance segmentation are commonly classified into two types: one-stage or regression-based methods, and two stage methods.

After the proposal of R-CNN (Girshick et al., 2014), deep learning-based object detection methods overtook traditional object detection methods. In R-CNN, the selective search method was used first to generate candidate regions, which was followed by feature extraction, classification, and localization. The existence of region proposal step divided object detection and instance segmentation into two stages. To improve the performance of R-CNN, Spatial Pyramid Pooling Network (SPP-Net) was proposed (He et al., 2015). By employing the idea from SPP-Net and Region Proposal Networks (RPN), Fast R-CNN and Faster R-CNN were proposed to show high performance (Girshick, 2015; Ren et al., 2016). Mask R-CNN was developed based on Faster R-CNN by adding a mask segmentation branch and replacing Region of Interest (RoI) pooling with RoI Align in feature tailoring after feature extraction (He et al., 2017). In 2018, cascade strategy was applied in object detection and instance segmentation (Cai and Vasconcelos, 2018). As explained by its name, in Cascade R-CNN (Cascade Mask R-CNN), two intermediate stages for object detection (and mask segmentation) were added between feature extraction and object detection (instance segmentation) heads of Faster R-CNN (Mask R-CNN), which resulted in Cascade R-CNN (Cascade Mask R-CNN). After Cascade R-CNN, there were other methods, such as Dynamic R-CNN (Zhang et al., 2020), but the improvement of these methods was limited. For two-stage instance segmentation, Hybrid Task Cascade (HTC) was proposed by involving interleaved execution, mask information flow, and semantic segmentation branch on top of Cascade Mask R-CNN (Chen et al., 2019). Both Cascade Mask R-CNN and HTC showed higher performance compared to Mask R-CNN, while both require more time in the training and testing phases. In our work, our modification is similar to those in HTC, while we use semantic features from the CNN backbone instead of shared features.

Object detection methods in the YOLO family are well-known one-stage methods. By taking the object detection task as the regression task, they are faster compared to two-stage detectors. Bounding boxes and class probabilities were directly output from images, not candidate regions. The latest YOLO detector is PaddlePaddle YOLO v2 (Huang et al., 2021), while after YOLO v3 (Farhadi and Redmon, 2018) there were few modifications made compared to previous algorithms. Starting from YOLO v4 (Bochkovskiy et al., 2020), newly developed optimization methods, such as better activation and better loss function, rather than major modifications on architectures were introduced to YOLO to improve the detection performance. Those optimization methods are supposed to further improve the performance of our method, but we leave these for future studies. You Only Look At CoefficienTs (YOLACT) and YOLACT++ were instance segmentation methods developed based on YOLO v3 (Bolya et al., 2019a,b; Bolya et al., 2020). Although these methods gave the highest accuracy among one-stage instance segmentation methods, their accuracy was lower than two-stage instance segmentation methods. Therefore, Mask R-CNN serves as the baseline in this work.

2.2. The applications of instance segmentation in remote sensing.

The application of instance segmentation in remote sensing is limited by the availability of training data. Even though the iSAID (Waqas Zamir et al., 2019) and MWPU VHR-10 (Su et al., 2020) were released, only limited object types were included, and the publicly available datasets cannot fulfill all requirements in remote sensing for instance segmentation.

From the existing literature, part of the research focused on single object detection, such as building footprint extraction and vehicle detection, while others focused on multiple object detection. For object detection, Zhao et al. (2018) used Mask R-CNN to extract building footprint from high spatial resolution satellite images and then processed extraction results using building boundary regularization. To extract building footprints in an end-to-end manner, Marcos et al. (2018) developed Deep Structured Active Contours (DSAC) by involving the Active Contour Model in a convolutional neural network. The method gave a high performance for building footprint extraction from aerial images in their experiment. Li et al. (2019) developed an end-to-end building footprint extraction method named PolyMapper, which was re-examined and improved by Zhao et al. (2021). In their methods, a recurrent neural network was added after a CNN network. Both of them showed high performance on aerial images compared to Mask R-CNN. For vehicle detection, Mou and Zhu (2018) applied a semantic boundary-aware multi-task learning network to extract vehicles from aerial images, which resulted in better performance compared to using semantic segmentation methods. There are also applications detecting multiple objects, which mainly use large datasets, such as iSAID and MWPU VHR-10.

For multiple object detection, Su et al. (2020) proposed a new method based on Cascade Mask R-CNN and named it High-Quality Instance Segmentation Network (HQ-ISNet). In the method, sophisticatedly designed RoI pooling was adopted. In addition, High Resolution Feature Pyramid Pooling (HRFPN) and ISNet v2 were applied to preserve high-resolution features and improve the accuracy of mask segmentation. The experiment on the Synthetic Aperture Radar (SAR) Ship Detection Dataset (SSDD) and iSAID dataset showed the high performance of the proposed method compared to the state-of-the-art methods. Zhang et al. (2021) proposed a new method, Semantic Attention and Scale Complementary Network, by adding a Semantic Attention module (SEA) and a Scale Complementary Mask Branch (SCMB) to Mask R-CNN. The SEA makes the network focus on the objects of interest and the SCMB improves the accuracy of mask segmentation. The experiment on iSAID and MWPU VHR-10 showed the new method processes competitive performance against the state-of-the-art methods. Zeng et al. (2021) proposed a Consistent Proposals of Instance Segmentation Network (CPISNet) based on Cascade Mask R-CNN. They developed an Adaptive Feature Extraction Network (APEN) for multi-level feature extraction, the Proposal Consistent Cascaded (PCC) architecture for bounding boxes refinement, and the Elaborated RoI Extractor (ERoIE) for mask RoIs extraction. The experiment on iSAID and MWPU VHR-10 showed the high performance of CPISNet compared to the state-of-the-art methods.

All applications mentioned above do not consider multi-source data, which is quite common in remote sensing. All Mask R-CNN based methods proposed in these publications modify FPN, RoI extractor, or head part which works for mask segmentation and bounding boxes detection. However, few works focussed on the feature extractor, while it is important for later modules and the performance of the whole network. In addition, oil well site detection is rarely explored, while it is an important task. Therefore, we conducted the work to explore oil well site detection with the newly proposed instance segmentation method using multi-source satellite images.

3. Study area and data

3.1. Study area

Alberta is one of the provinces in Canada ranging from 49°N to 60°N and from 110°W to 120°W (Fig. 1). Whereas the western part of the province borders the Rocky Mountains, the eastern part is occupied by the Great Plains; the latter predominates in Alberta, Canada, accounting for about 90% of ~ 66000 km^2 in total. The Alberta province is well-

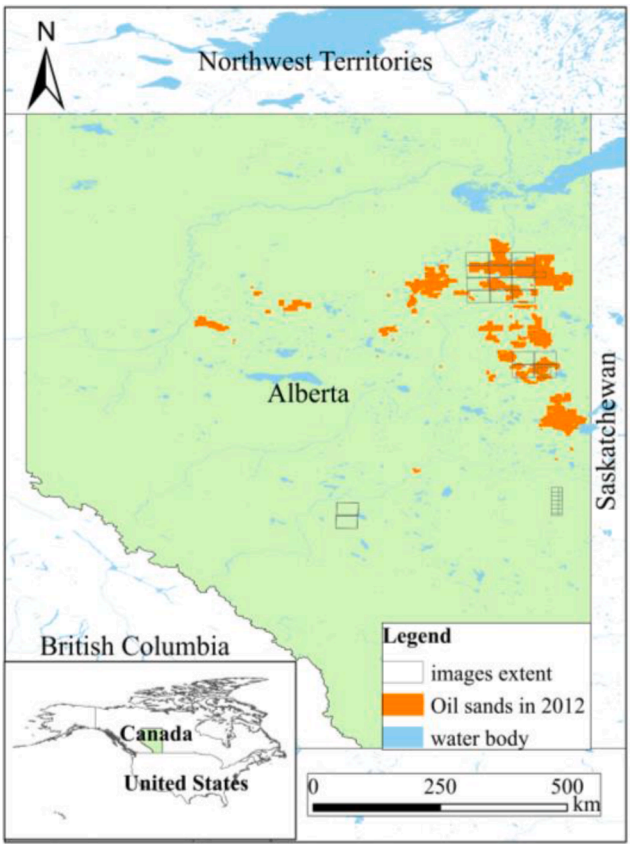


Fig. 2. Map of the extent of oils sands in Alberta, Canada and the coverage of collected images (Administrative area map comes from Hijmans et al., 2012).

known for abundant mineral resources, especially oil sands. Fig. 2 shows the extent of the oil sand deposition in Alberta, Canada. In the oil sands mining, the in-situ mining developments in forested land and farmland in Northern Alberta leave massive footprints on the land, including those of well sites and resources roads connecting to the well sites. In this work, we define those land disturbances related to mining development as oil well sites, and roads connected to oil well sites as resource roads.

3.2. Datasets

Remote sensing satellites acquire images with different spatial resolutions ranging from sub-meter level to kilometers level per pixel. The spatial resolution will limit the size of objects that can be extracted from those images. According to the Nyquist-Shannon theorem (Shannon, 1949), in order to identify the target object in images, the pixel size should be 1/2 the size of the smallest target object. As most oil well sites in Alberta, Canada has an area of about 100 m^2 (10 x10 m) in size, it is reasonable that meter-level images are feasible to be used in oil well site extraction. Lower resolution cannot capture effectively oil well sites. However, too much useless detailed information consumes much compute resources resulting in low computation speed when the spatial resolution is too high.

In this work, we collected 18 RapidEye 2/3 multi-spectral (MS) images (Image © 2021 Planet Labs PBC) and 15 WorldView-3 images with MS bands and PAN bands (Satellite imagery © 2021 Maxar Technologies). The coverage of the data is shown in Fig. 1. Each RapidEye image has an area of 25 × 25 km^2 with an image size of 5000 × 5000 pixels, and with a spatial resolution of 5 m. Each WorldView-3 in use in this work has different sizes. The resolutions of PAN and MS images are 0.5 m and 2 m, respectively.

Given different spectral and spatial resolution in our collected images, we selected Blue, Green, Red, Red Edge, and NIR (NIR1 in WorldView-3 images) bands to make full use of collected data, as well as 2 x2 m pixel size as the spatial resolution of our dataset used for oil well sites extraction for full use of the dataset. In addition, we used a pretrained RCAN (Zhang et al., 2018b) model[1], one of the state-of-the-art super-resolution methods, to improve the spatial resolution of RapidEye to 2 m/pixel spatial resolution to match the resolution of WorldView 3 multispectral data.

To generate the training dataset, oil well sites and line objects connected to them were manually annotated on these images with the aid of Google Maps. Labeled images and processed images were further cropped into 512 × 512-pixel patches. Consequently, we obtained 11,250 and 845 image tiles for RapidEye 2/3 and WorldView 3, respectively, i.e., a total of 12,095 paired mask tiles for both roads and well site labels. Fig. 3 depicts several examples of paired images and mask tiles. Those patches were further split into train, validation, and test with a ratio of 6:1:3.

4. Method

4.1. Oil well site extraction Mask R-CNN

As the well sites are always connected to resource roads, we took the connection relationship between road network and oil well sites into account in the development of our method. As D-LinkNet is a state-of-the-art method in road network segmentation, we selected it as the feature extractor in our method. Furthermore, the U-shape and the dilation convolution of D-LinkNet made it good at preserving multi-level features and context information, which was important to improve the network performance. In addition, a semantic segmentation loss, calculated using the nonlinear transformation (sigmoid activation layer in this work) of features from D-LinkNet and ground truth masks, was added to the total loss, which made the CNN feature extractor focus on both oil well sites and connected road network and served as a new branch exclusively for semantic segmentation. Adding the new loss was expected to generate high-quality features with context information of oil well sites for later modules of the new method and improve the performance of the model (Li et al, 2021). The task of semantic segmentation can act as an auxiliary task to help learn the features needed for well site extraction. The architecture of our process workflow is shown in Fig. 4, in which the dashed line represents the newly added branch.

Mask R-CNN is proposed on top of Faster R-CNN by replacing RoI pooling with RoI Align and adding a mask segmentation branch. Both RoI pooling and RoI Align work for processing and resizing objects RoIs to the same size. The processed and resized RoIs are taken as input for followed FC layers. Here, RoI pooling first mapped RoIs to feature maps and then resized mapped RoIs. For example, if an object has a RoI with a size of 100 × 80 and our input images have a size of 512 × 512, to map the RoI to the feature map P3, which has a size of 64 × 64, we have to do processing on our RoIs. As shown in Fig. 5, we get the mapped RoI with a size of (12,10). To explain RoI pooling and RoI align clearly, we set the pooling size as 3 × 3 here. It means that we have to resize the RoI from (12, 10) to (3,3). Quantization and downsampling are used in both steps in RoI pooling, which results in information loss. For RoI Align, the original RoIs are divided into 3 × 3 bins first. For pooling purposes, four points are then evenly generated in each bin. Finally, the value of the whole bin is determined by four points' values using bilinear interpolation (as shown in Fig. 6). To construct our OWS Mask R-CNN, we selected D-LinkNet (as shown in Fig. 6) instead of the original ResNet 101 (as shown in Fig. 7) as the CNN, which is developed based on LinkNet (Chaurasia and Culurciello, 2017). By utilizing ResNet blocks in encoder and decoder blocks of U-Net (Ronneberger et al., 2015),

[1] Pretrained model and codes can be found at https://github.com/hehongjie/Oil-well-pads-extraction.

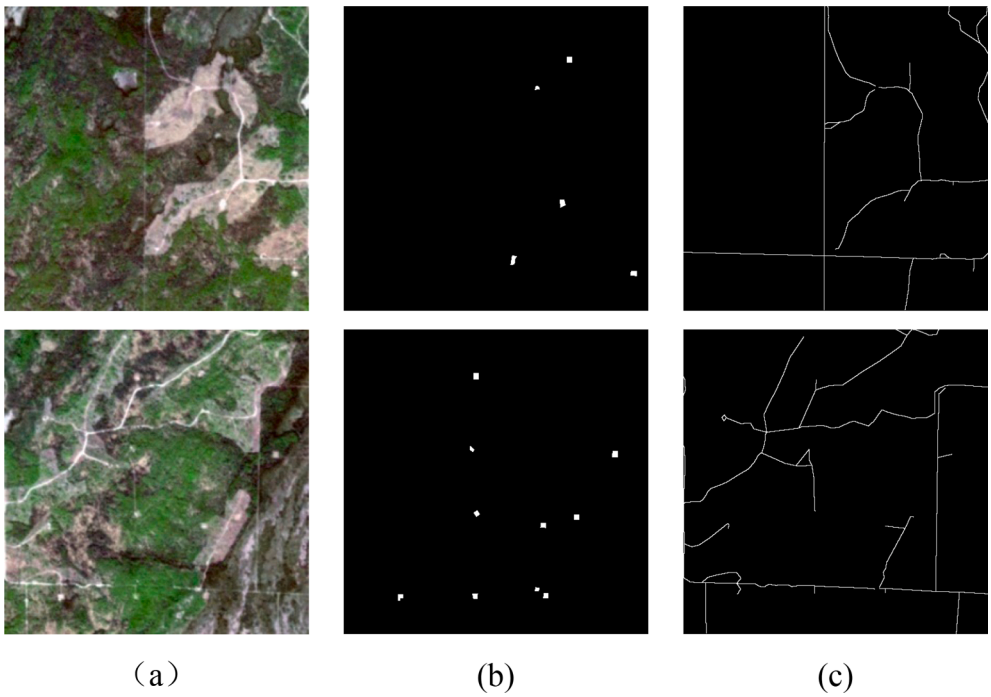


Fig. 3. Example of generated roads and oil well sites dataset. (a) Satellite image tiles, (b) mask image tiles for well site labels, and (c) mask image tiles for road labels.

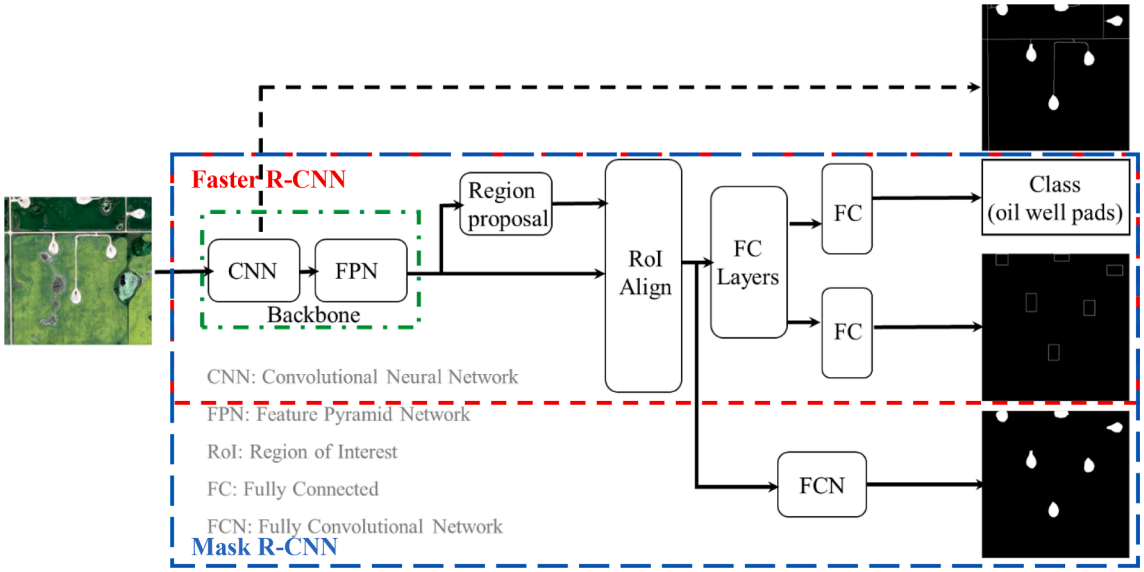


Fig. 4. Architecture of proposed oil well sites extraction Mask R-CNN model (ResNet 101 is replaced by D-LinkNet as new CNN for feature extraction).

LinkNet was proposed as an efficient semantic segmentation. To extract line objects, such as road networks, dilated convolution layers were applied between the encoder and decoder parts of LinkNet, which enlarge the receptive field and output high spatial resolution features for semantic segmentation. The encoder, central and decoder parts make up the D-LinkNet (Zhou et al., 2018). Considering the imbalance distribution of positive objects and negative objects (pixels of oil well sites and background in this work) in ground truth masks, we apply soft Jaccard loss (Yuan et al., 2017) to supervise the training of the newly added semantic segmentation branch. Other loss functions include classification loss and bounding box detection loss from region proposal network, classification loss, bounding box detection loss, and mask segmentation loss from the output of the method following the original setting in Mask R-CNN. Categorical cross-entropy and binary cross-entropy loss functions are selected as classification and mask segmentation loss. For bounding box regression loss, the L1 loss function is adopted. Other modules of Mask R-CNN, such as FPN, RPN, and head parts are detailed in Figs. 8 and 9.

4.2. Evaluation metrics

To evaluate the performance of the DL models oil well site extraction, two types of quantitative evaluation metrics were utilized: (1) pixel-level metrics, and (2) object-level metrics. The former includes six commonly used metrics: pixel accuracy, Intersection of Union (IoU), precision, recall, and F_1 score. To start with, pixel accuracy simply reports the percentage of pixels in the image which were correctly segmented. IoU refers to the intersection of predicted masks and ground

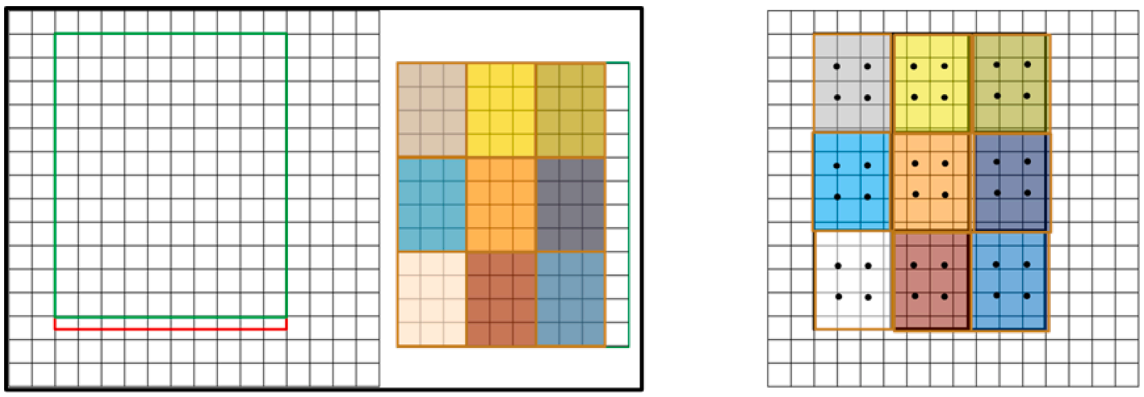


Fig. 5. RoI pooling (left) and RoI align (right) (only 16 × 16 instead of 64 × 64 grids are shown here for clear representation).

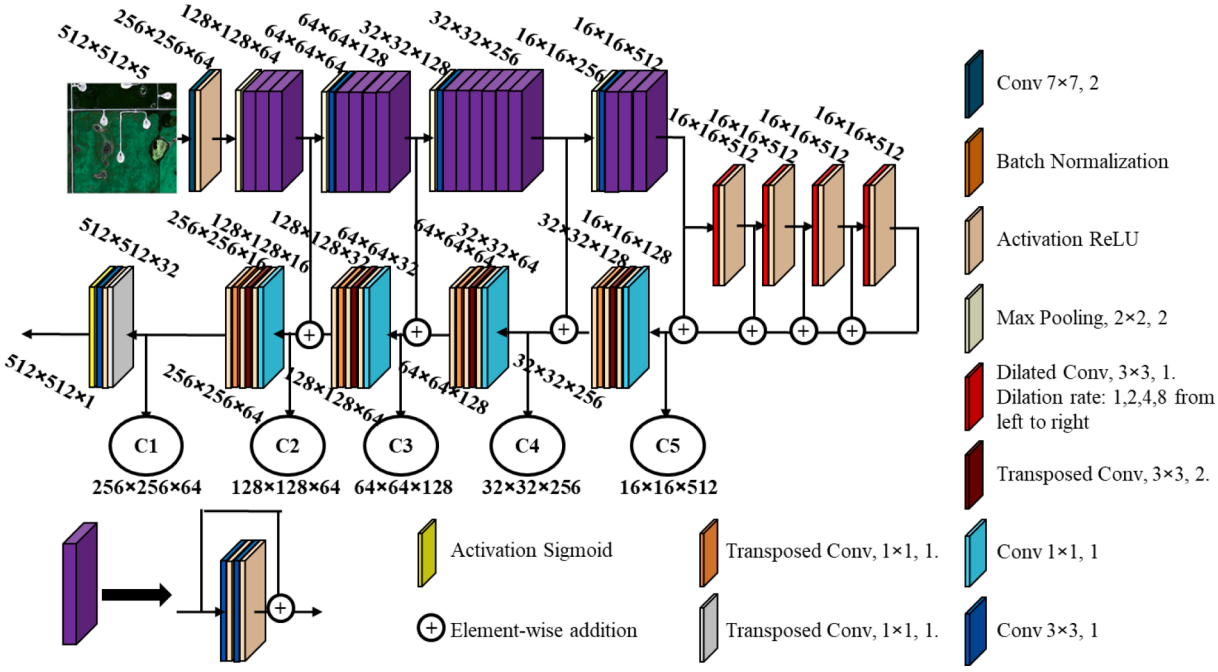


Fig. 6. The architecture of D-LinkNet.

truth to the union of them. Precision represents the ratio of correctly extracted masks to all predicted ones; recall refers to the ratio of accurately detected masks to ground truth. Finally, F_1 score is the harmonic mean of precision and recall. Equations for calculating these metrics are listed below:

$$\text{Pixel Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} = \frac{2TP}{2TP + \text{FP} + FN} \tag{5}$$

where True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) represent cases when the model predicts the positive class as positive(i.e., TP) or as negative (i.e., FN) and predicts the negative class as positive (i.e., FP) or as negative (i.e., TN), respectively.

On the other hand, the object-level metrics used in this study are Average Precision (AP), AP_{75}, and AP_{50}. All three metrics were calculated based on the precision-recall curve under certain IoU thresholds. For example, given an IoU threshold required to deem the object detection as a positive detection, a pair of recall and precision can be calculated, which constructs the precision-recall curve. In most cases, the calculated AP indicates averaged AP value when IoU equals 0.50, 0.55, 0.60, 0.65, 0.70. 0.75, 0.80, 0.85, 0.90 and 0.95. AP_{75} and AP_{50} represent the AP value with IoU equaling 0.75 and 0.50. Both pixel-level and object-level metrics are used to evaluate the well sites extraction results including bounding boxes and masks. The higher AP value represents the higher accuracy of extraction or detection results.

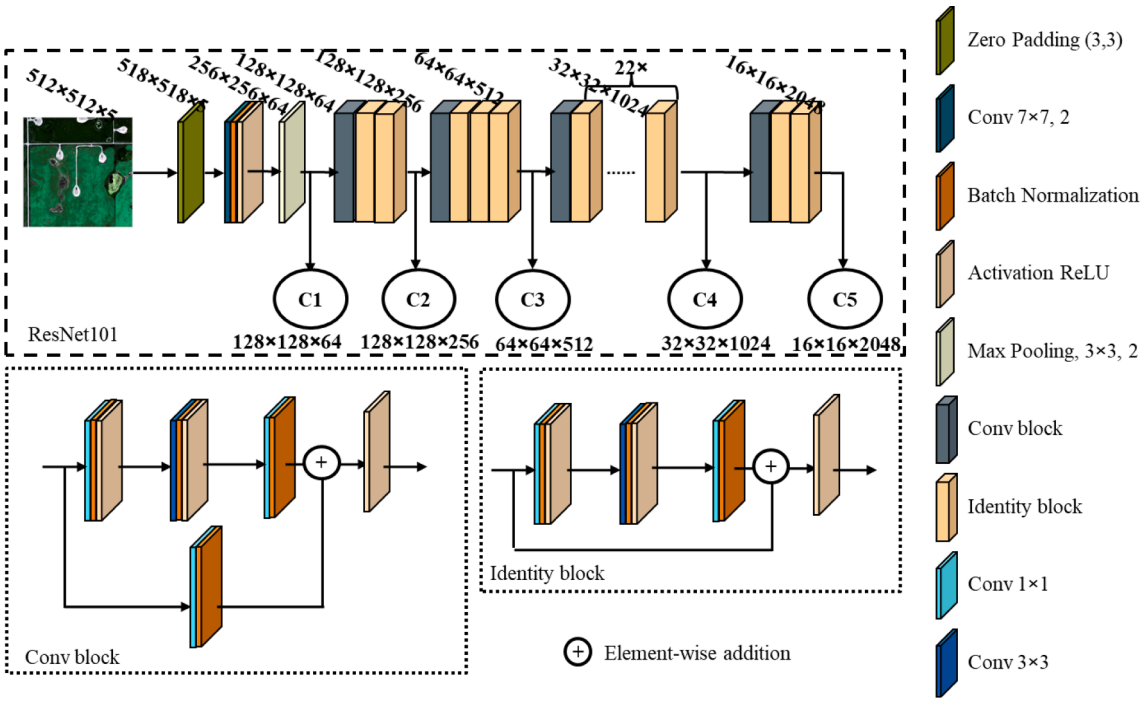


Fig. 7. The architecture of ResNet 101 (the first Conv 1 × 1 layer in each branch of each Conv block, except the first block, has a stride of 2).

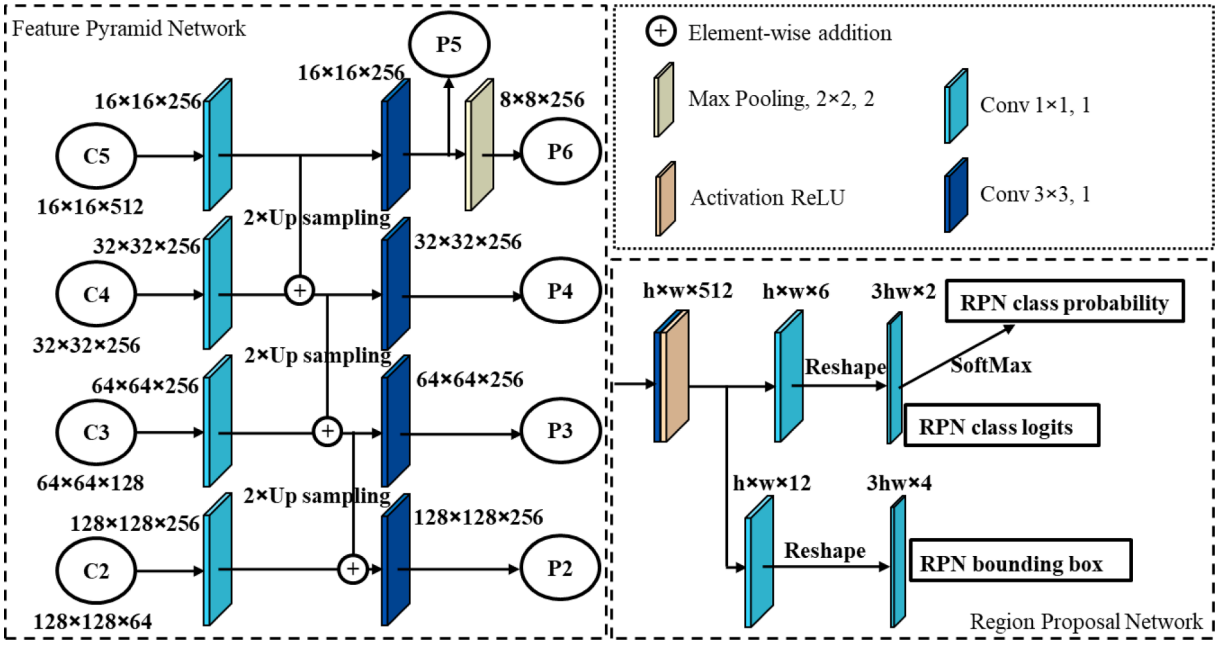


Fig. 8. The architecture of FPN (left) and RPN (bottom right).

4.3. Implementation details

In this work, we used an image tile size of 512 × 512, a batch size of 2[2], an anchor ratio of 0.5, 1, and 2, a minimum confidence score of 0.9, a learning rate of 1e-3, and equal weight for all losses. We used ResNet 101 as the initial backbone, which we then replaced. The rest of the parameters are the same as the widely used implementation[3]. All models were trained for 200 epochs and tested under TensorFlow 2.4.1 on a single Nvidia TITAN XP with CUDA 11.4.

5. Experiments results and discussion

In this section, first, both qualitative and quantitative evaluations of these methods are presented; and in the following section, the results

[2] Batch size is suggested to be set with larger value with more computational resources or smaller size of input.

[3] https:// github.com/matterport/Mask_RCNN.

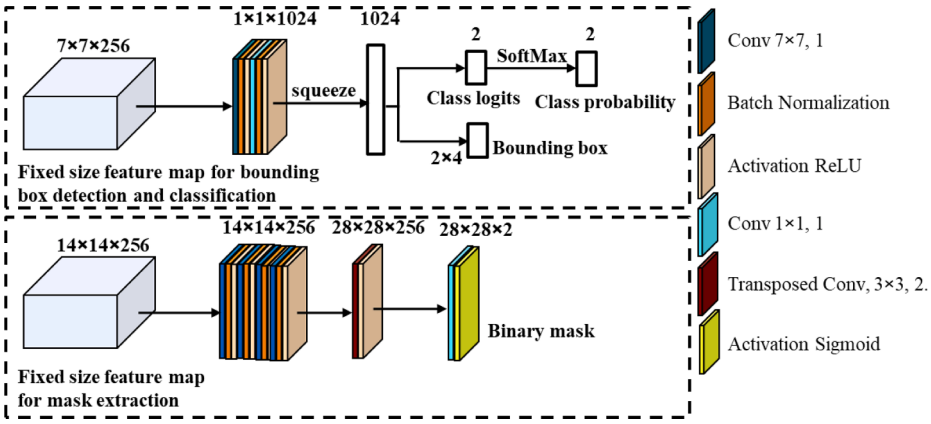


Fig. 9. Fully Connected layers for classification and bounding box detection (top), and fully convolutional layers for mask extraction (bottom).

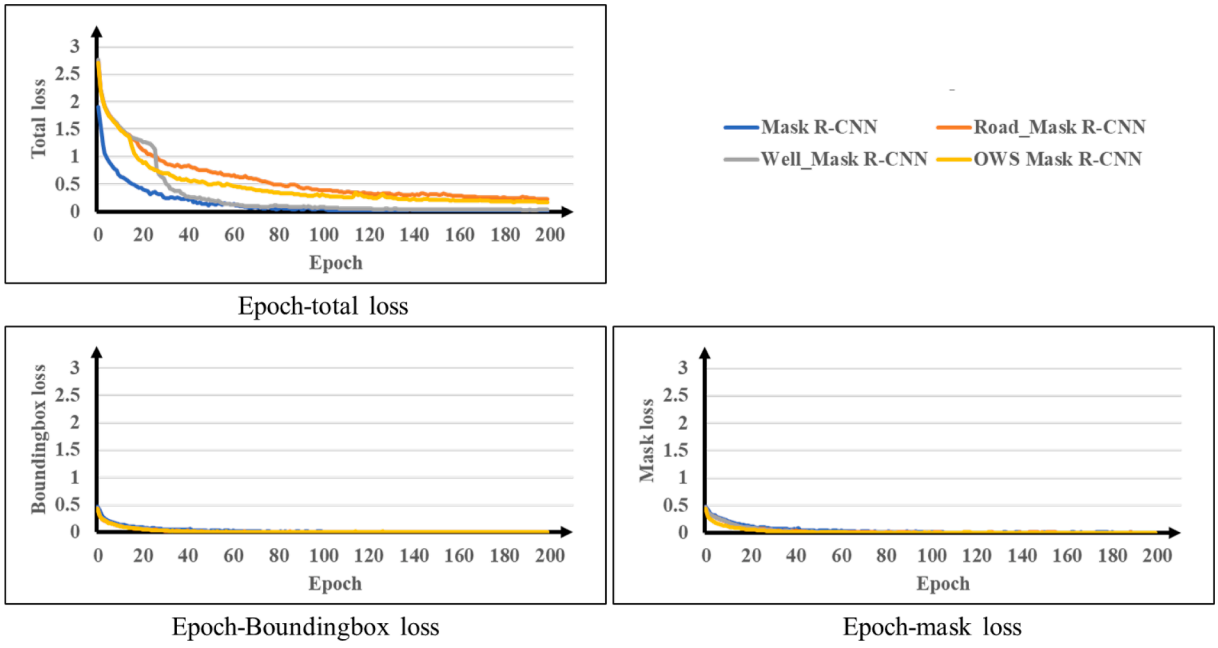


Fig. 10. Learning curves.

presented from studies on the impacts of replacing backbone, as well as different tasks in semantic segmentation branch, involving NDVI images and adding a new branch, on the performance of OWS Mask R-CNN on oil well sites extraction.

5.1. Qualitative evaluation

Before diving into the qualitative evaluation, we visualize the training process with learning curves in Fig. 10. Examples of the mask extraction results based on Mask R-CNN and our proposed OWS Mask R-CNN model are provided in Figs. 11 and 12. In these two figures, the first and second rows are images and ground truth examples. From the third to the last row, extraction results of Mask R-CNN, Road_Mask R-CNN (OWS Mask R-CNN considering only road network in semantic segmentation branch), Well_Mask R-CNN (OWS Mask R-CNN considering only well in semantic segmentation branch), OWS Mask R-CNN are tabulated. As shown in Fig. 11, from Mask R-CNN to OWS Mask R-CNN, misclassified pixels decrease, which is obvious in the first four examples. The improvement from Mask R-CNN to Road_Mask R-CNN is significant among all methods. The instance-level extraction results are provided in Fig. 12. As shown in Fig. 12, Mask R-CNN recognizes all anomaly small patches as oil well sites, while Road_Mask R-CNN, Well_Mask R-CNN, and OWS Mask R-CNN can accurately recognize oil well sites. We attribute it to the positive impact of the new backbone (D-LinkNet + FPN) and the new semantic segmentation branch. It is interesting to notice that although the mislabelled oil well sites exist near the center of example 2, all methods can well detect and extract the real object. We believe only limited mislabelled oil well sites exist in our dataset as we checked several rounds after annotating. The mislabelled objects may also serve as noise in model training and help deal with generalization errors in test or deployment phases (Zhou et al., 2019). The qualitative evaluation results, to some extent, confirm the success of our proposal.

5.2. Quantitative evaluation

Table 1 summarizes the quantitative evaluation metrics' values of mask extraction using Mask R-CNN, Road_Mask R-CNN, Well_Mask R-CNN, and our OWS Mask R-CNN model.

As shown in Table 1, the OWS Mask R-CNN model achieves better performance compared to other models in terms of all metrics except Recall. Specifically, AP is gradually increased from 20.98% of Mask R-CNN to 25.35% of OWS Mask R-CNN. It is more than 20% improvement.

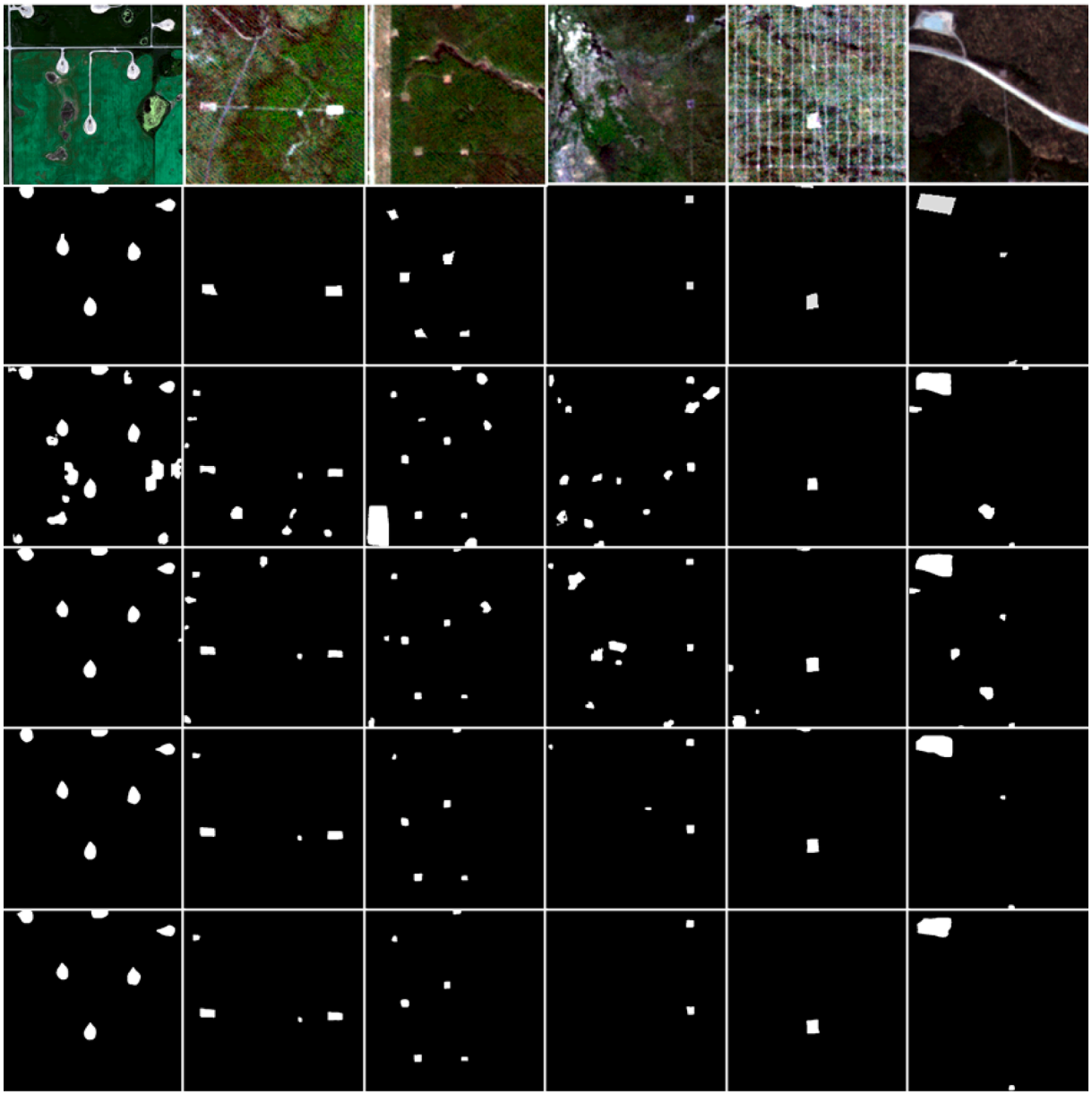


Fig. 11. Examples of masks detection results (visualization in semantic segmentation manner).

Pixel-level metrics also increased from Mask R-CNN to OWS Mask R-CNN except for Recall. Because our data has significant data imbalance, we omit the pixel accuracy metric in Tables 1, 2, and 3 to avoid misleading the readers. Given the results, we can conclude that our proposed OWS Mask R-CNN with a new backbone and a new semantic segmentation branch is more successful than Mask R-CNN in oil well site detection and extraction.

5.3. Comparative study

In this section, we conduct a comparative study to examine the performance of our OWS Mask R-CNN compared to other instance segmentation methods in oil well site detection and extraction. As Mask Scoring R-CNN (Huang et al., 2019) and Cascade Mask R-CNN are commonly selected in the comparative study (Zeng et al., 2021; Zhang et al., 2021; Chen et al., 2019), we also selected them in our experiment.

As shown in Table 2, our OWS Mask R-CNN has the best performance compared to the other three methods in all metrics except Recall. Compared to Mask R-CNN, Cascade Mask R-CNN give high scores in pixel-level metrics except for Recall but low scores in object-level metrics. The Mask Scoring R-CNN is supposed to give higher performance compared to Mask R-CNN, but it does not. Mask Scoring R-CNN adds a Mask IoU head and scoring loss to Mask R-CNN, which considers the IoU of predicted masks and ground truth masks in model training. Cascade Mask R-CNN can be seen as a multi-stage extension of Mask R-CNN. For cascade Mask R-CNN, current stage detection and mask extraction take as input the detection results of the last stage. The Mask Scoring R-CNN and Cascade Mask R-CNN are proposed after Mask R-CNN and are supposed to give higher performance compared to Mask R-CNN, but they do not in our experiment. We explained the results as that given limited data the advantage of two more complicated methods cannot be fully exploited.

Because the oil well site mask is the focus of our work, we also compare our OWS Mask R-CNN with the state-of-the-art semantic segmentation methods. As shown in Table 3, among 4 methods, our OWS Mask R-CNN shows the best performance with high scores in all metrics. Compared to our previous work (He et al., 2022), both DeepLab v3+, HRNet v2, and Mask R-CNN show worse performance in oil well site

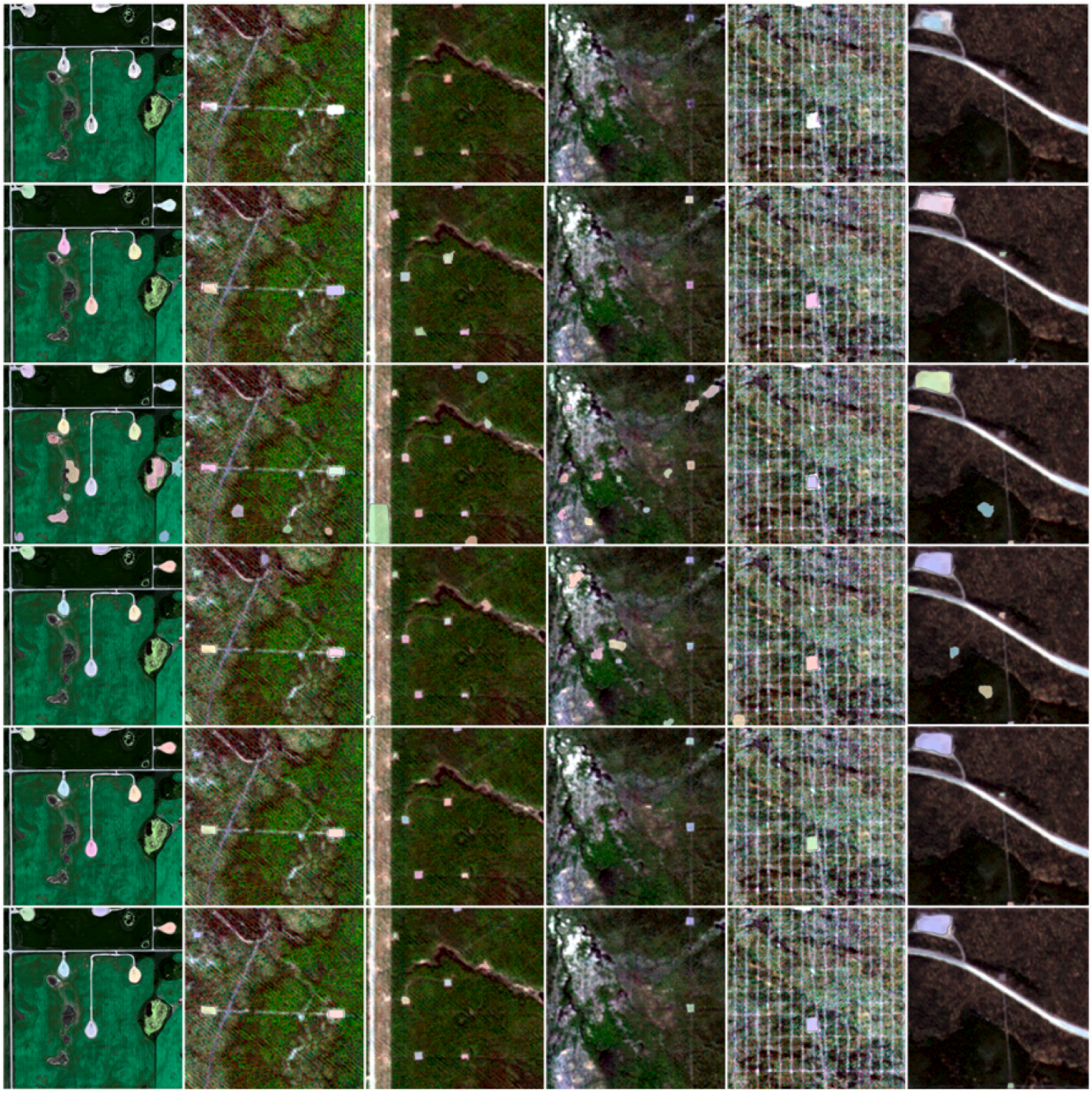


Fig. 12. Examples of masks detection results (visualization in instance segmentation manner).

Table 1
Evaluation metrics (%) for the mask extraction.

Models	AP	AP_{75}	AP_{50}	IoU	Precision	Recall	F_1
Mask R-CNN	20.98	13.70	53.38	15.56	16.42	**74.74**	26.93
Road_Mask R-CNN	23.00	15.17	56.60	19.06	20.40	74.49	32.02
Well_Mask R-CNN	24.84	15.27	62.42	42.53	52.03	69.96	59.68
OWS Mask R-CNN	**25.35**	**15.92**	**63.02**	**47.15**	**59.63**	69.26	**64.09**

detection and extraction than that in building footprint extraction. To some extent, the results show the challenge and the failure of existing methods in oil well site detection and extraction.

5.4. Ablation study

(1) The impact of the backbone replacement and the new semantic segmentation.

To develop OWS Mask R-CNN, an improvement was made by

Table 2
The results of instance segmentation methods comparison.

Models	AP	AP_{75}	AP_{50}	IoU	Precision	Recall	F_1	Trainable parameters
Mask R-CNN	20.98	13.70	53.38	15.56	16.42	74.74	26.93	64,151,966
Cascade Mask R-CNN	18.20	10.8	46.38	22.62	25.52	66.59	36.90	97,222,582
Mask Scoring R-CNN	16.14	6.47	47.90	11.64	11.95	81.85	20.86	79,830,176
OWS Mask R-CNN	**25.35**	**15.92**	**63.02**	47.15	59.63	69.26	64.09	**54,851,198**

Table 3
The results of semantic segmentation methods comparison.

Models	IoU	Precision	Recall	F_1
DeepLab v3+	10.21	72.39	**10.62**	18.52
HRNet v2	21.27	83.51	22.20	35.08
Mask R-CNN	15.56	16.42	74.74	26.93
OWS Mask R-CNN	**47.15**	**59.63**	**69.26**	**64.09**

replacing the original ResNet + FPN backbone with D-LinkNet + FPN and adding a semantic segmentation branch by considering the relationship between oil well sites and connected road networks. In Table 4, the impacts of two modifications on the accuracy of extraction results are presented. Here, we denote Mask R-CNN which takes D-LinkNet and FPN as the backbone as D-Mask R-CNN.

As shown in Table 4, object-level metrics' values are significantly improved from Mask R-CNN to D-Mask R-CNN. AP and AP_{75} of D-Mask R-CNN are even higher than OWS Mask R-CNN. However, the improvement of the pixel-level metrics' values (except recall) from Mask R-CNN to OWS Mask R-CNN comes from the involvement of a new semantic segmentation branch. We explain it as 1) D-LinkNet brings high spatial resolution features to OWS Mask R-CNN and results in masks refinement; 2) the new semantic segmentation branch filters out most FP pixels and results in the improvement of pixel-level metrics.

(2) The impact of road network connection information.

Oil well sites are always connected to roads for transportation. Fig. 1, Fig. 11, and Fig. 12 show some examples. In our OWS Mask R-CNN, we replaced ResNet 101 with D-LinkNet which is proposed for road network segmentation. In addition, we considered both road network segmentation and oil well sites mask extraction in the new semantic segmentation branch to force the network to focus on their relationship and improve the accuracy of oil well site extraction. To test the impact of the relationship on oil well site extraction, we summarized related experiments in this section.

As shown in Table 5, from D-Mask R-CNN to OWS Mask R-CNN, all scores are increased significantly except Recall. The reason for the increase is the new semantic segmentation branch, which considers both road networks and oil well sites. In addition, by considering road network information, OWS Mask R-CNN surpasses Well_Mask R-CNN in all metrics values except Recall. The results confirm the positive impact of involving road network information and considering the connection relationship between oil well sites and road networks in model training.

(3) The impact of NDVI on the performance of different models.

Normalized Difference Vegetation Index (NDVI) is commonly used in land disturbance detection (Goetz et al., 2006; Yang et al., 2018). In this work, we also tested the impact of involving NDVI in input on the final extraction results. Table 6 summarizes evaluation metrics' values of extraction results from models with or without NDVI. Models that take the original 5-band as input are denoted as Mask R-CNN and D-Mask R-CNN, while models that take the original 5-band and NDVI as input are denoted with + NDVI below the correspondent model for simplification.

As shown in Table 6, by adding NDVI features in input, almost all metrics' values decreased from their baselines. The NDVI features enhanced the contrast features of oil well sites from their neighbor pixels. Therefore, adding NDVI features was supposed to increase the performance of the instance segmentation models, but opposite results

Table 4
Evaluation metrics (%) for the models with different modifications.

Models	AP	AP_{75}	AP_{50}	Pixel Accuracy	IoU	Precision	Recall	F_1
Mask R-CNN	20.98	13.70	53.38	97.34	15.56	16.42	74.74	26.93
D-Mask R-CNN	25.19	15.82	60.63	97.83	19.02	20.12	**77.78**	31.96
OWS Mask R-CNN	25.35	15.92	63.02	99.49	47.15	59.63	69.26	64.09

Table 5
Evaluation metrics (%) for the impact of road network connection information.

Models	AP	AP_{75}	AP_{50}	Pixel Accuracy	IoU	Precision	Recall	F_1
Mask R-CNN	20.98	13.70	53.38	97.34	15.56	16.42	74.74	26.93
D-Mask R-CNN	25.19	15.82	60.63	97.83	19.02	20.12	**77.78**	31.96
Road_Mask R-CNN	23.00	15.17	56.60	97.93	19.06	20.40	74.49	32.02
Well_Mask R-CNN	24.84	15.27	62.42	99.38	42.53	52.03	69.96	59.68
OWS Mask R-CNN	**25.35**	**15.92**	**63.02**	**99.49**	**47.15**	**59.63**	69.26	**64.09**

Table 6
Evaluation metrics (%) for the models with/without NDVI.

Tasks	AP	AP_{75}	AP_{50}	Pixel Accuracy	IoU	Precision	Recall	F_1
Mask R-CNN	20.98	13.70	53.38	97.34	15.56	16.42	74.74	26.93
+NDVI	19.65	11.82	**51.28**	94.83	9.15	9.37	**79.41**	16.77
D-Mask R-CNN	**25.19**	**15.82**	**60.63**	**97.83**	**19.02**	**20.12**	**77.78**	**31.96**
+NDVI	22.56	14.62	55.36	97.62	17.63	18.57	77.69	29.98

Table 7
Precision and recall along with different prediction scores.

Scores rank	Ground truth	Precision	Recall
1	OWS	1.00	0.17
2	Other	0.50	0.17
3	OWS	0.67	0.33
4	OWS	0.75	0.50
5	Other	0.60	0.50
6	Other	0.50	0.50
7	OWS	0.57	0.67
8	OWS	0.63	0.83
9	OWS	0.67	1.00
10	Other	0.60	1.00

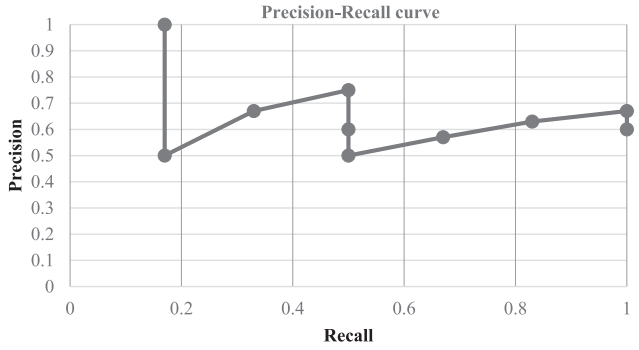


Fig. 13. The precision-recall curve of the example.

were generated. We would like to study this phenomenon further.

6. Conclusion

Automated extraction of well sites from satellite imagery is important for providing efficiently the information about footprints of mining development, which is essential for research on the cumulative impacts due to mining activities. In this paper, we presented a new method, named OWS Mask R-CNN, for automated extraction of oil well sites from multi-modality satellite images. In the new algorithm, we replaced the backbone from ResNet101 + FPN to D-LinkNet + FPN to preserve high spatial resolution features. In addition, a new semantic segmentation branch was added to Mask R-CNN to help the network focus on the relationship between road networks and oil well sites. The proposed OWS Mask R-CNN was shown to be successful in this work given its high performance in oil well site extraction from RapidEye 2/3 and WorldView-3 images. The involvement of super-resolution provided a solution to the problem brought by different spatial resolutions among different sensors, which also relieved the lack of training samples in oil well site extraction by ensuring the use of all available satellite images. Size-filtering was also confirmed as a useful step to improve the accuracy of extraction results in certain circumstances. According to the pixel accuracy of extraction results from the OWP model, there was probably no room for accuracy improvement in terms of pixel-level metrics, although our proposal can be transplanted to Cascade R-CNN, HTC, and even the latest method, QueryInst (Fang et al., 2021). Future works could consider mask refinement by using sophisticated designed architectures, advanced optimization methods, and more context information, such as road networks. We would focus on these directions in our future studies.

CRediT authorship contribution statement

Hongjie He: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Hongzhang Xu:** Investigation, Visualization, Data curation. **Ying Zhang:** Investigation, Data curation, Funding acquisition. **Kyle Gao:** Writing – review & editing. **Huxiong Li:** Investigation. **Lingfei Ma:** Writing – review & editing, Supervision, Funding acquisition. **Jonathan Li:** Resources, Writing – review & editing, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was partially funded by the Emerging Interdisciplinary Project of Central University of Finance and Economics, and also partially funded by the Remote Sensing for Cumulative Effects program of Canada Centre for Remote Sensing, National Resources Canada. The first author also acknowledges the China Scholarship Council for their support via a doctoral scholarship (No. 201906180088). The group of the students at the Geospatial Sensing and Data Intelligence Laboratory, Faculty of Environment, University of Waterloo, including Siyu Li, Wenxuan Zhu, Yiqing Wu, Yuxiang Fang, Longxiang Xu, and Charlotte Pan are acknowledged for their contributions to labeling data. We also acknowledge the Planet Inc and the Maxar Technologies Inc for providing satellite images used in this work.

Appendix A

AP indicates averaged AP value when IoU equals 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90 and 0.95. AP_{75} and AP_{50} represent the AP value with IoU equaling 0.75 and 0.50. We provide a fictitious example below to calculate AP_{75}. In practice, the number of predicted masks is too large to present here as an example.

With the IoU threshold of 0.75, we detected 10 masks for 6 objects. We can first sort the masks along with the predicted scores from the highest score to the lowest score. With different classification score thresholds, we can recalculate the precision and recall from object level, as shown in Table 7. In calculation, the predicted masks overlapping the ground truth over 75% (IoU larger than 0.75), as well as possessing classification scores larger than the threshold, are recognized as "Oil Well Sites (OWS)" otherwise "Other". According to Table 7, we can plot the precision-recall curve, as shown in Fig. 13.

AP_{75} is calculated as the area under to the precision-recall curve. In our experiment, we adopted the equations used by Pascal Visual Object Classes (VOC) 2010 (Everingham et al., 2012), which can be calculated as follow:

$$AP = \sum_{1 \le i \le n} (r_i - r_{i-1}) * p_i$$

where r_i *and* p_i are all recall and precision value under a certain classification score. r_1, r_n is the smallest and the largest recall value. If two precision

values match one recall value, the larger precision will be preserved for the curve. In the example, AP_{75} = (1–0.83) *0.67+(0.83–0.67) *0.63+(0.67–0.50) *0.57+(0.5–0.333) *0.75+(0.33–0.17) *0.67 = 0.55. AP with other IoU thresholds can be calculated in the same way.

References

Ball, J.E., Anderson, D.T., Chan Sr, C.S., 2017. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. J. Appl. Remote Sens. 11(4), p.042609.

Bayne, E.M., Van Wilgenburg, S.L., Boutin, S., Hobson, K.A., 2005. Modeling and field-testing of Ovenbird (Seiurus aurocapillus) responses to boreal forest dissection by energy sector development at multiple spatial scales. Landsc. Ecol. 20 (2), 203–216.

Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M., 2020. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

Bolya, D., Zhou, C., Xiao, F., Lee, Y.J., 2019. Yolact: Real-time instance segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9157-9166.

Bolya, D., Zhou, C., Xiao, F., Lee, Y.J., 2019. Yolact++: Better real-time instance segmentation. arXiv preprint arXiv:1912.06218.

Bolya, D., Zhou, C., Xiao, F., Lee, Y.J., 2020. YOLACT++: Better Real-time Instance Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. https://doi.org/10.1109/TPAMI.2020.3014297.

Cai, Z., Vasconcelos, N., 2018. Cascade r-cnn: Delving into high quality object detection. In: Proc. CVPR, pp. 6154-6162.

Chaurasia, A., Culurciello, E., 2017, December. Linknet: Exploiting encoder representations for efficient semantic segmentation. In: Proc. 2017 IEEE Visual Communications and Image Processing (VCIP), pp. 1-4.

Chen, K., Pang, J., Wang, J., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Shi, J., Ouyang, W., Loy, C.C., 2019. Hybrid task cascade for instance segmentation. In: Proc. CVPR, pp. 4974-4983.

Dyer, S.J., O'Neill, J.P., Wasel, S.M., Boutin, S., 2001. Avoidance of industrial development by woodland caribou. J. Wildlife Manage. 65 (3), 531–542.

Erzurumlu, S.S., Erzurumlu, Y.O., 2015. Sustainable mining development with community using design thinking and multi-criteria decision analysis. Resources Policy 46, 6–14.

Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2011. The pascal visual object classes challenge 2012 (voc2012) results (2012). In: URL http://www.pascal-network.org/challenges/VOC/voc2011/workshop/index.html.

Fang, Y., Yang, S., Wang, X., Li, Y., Fang, C., Shan, Y., Feng, B., Liu, W., 2021. Instances as queries. In: Proc. ICCV, pp. 6910-6919.

Farhadi, A., Redmon, J., 2018, April. Yolov3: An incremental improvement. In: Proc. CVPR, pp. 1804-02767.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proc. CVPR, pp. 580-587.

Girshick, R., 2015. Fast R-CNN. In: Proc. CVPR, pp. 1440-1448.

Goetz, S.J., Fiske, G.J., Bunn, A.G., 2006. Using satellite time-series data sets to analyze fire disturbance and forest recovery across Canada. Remote Sens. Environ. 101 (3), 352–365.

Gosselin, P., Hrudey, S.E., Naeth, M.A., Plourde, A., Therrien, R., Van Der Kraak, G., Xu, Z., 2010. Environmental and health impacts of Canada's oil sands industry. Royal Society of Canada, Ottawa, ON, p. 10.

He, H., Jiang, Z., Gao, K., Narges Fatholahi, S., Tan, W., Hu, B., Xu, H., Chapman, M.A., Li, J., 2022. Waterloo building dataset: a city-scale vector building dataset for mapping building footprints using aerial orthoimagery. Geomatica 75 (3), 99–115.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: Proc. CVPR, 2961-2969.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37 (9), 1904–1916.

Hijmans, R.J., Guarino, L., Mathur, P., 2012. DIVA-GIS. Version 7.5. A geographic information system for the analysis of species distribution data. Bioinformatics 19.

Huang, X., Wang, X., Lv, W., Bai, X., Long, X., Deng, K., Dang, Q., Han, S., Liu, Q., Hu, X. and Yu, D., 2021. PP-YOLOv2: A Practical Object Detector. arXiv preprint arXiv: 2104.10419.

Huang, Z., Huang, L., Gong, Y., Huang, C., Wang, X., 2019. Mask scoring r-cnn. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6409–6418.

Li, X., He, M., Li, H., Shen, H., 2021. A combined loss-based multiscale fully convolutional network for high-resolution remote sensing image change detection. IEEE Geosci. Remote Sens. Lett. 19, 1–5.

Li, Z., Wegner, J.D. and Lucchi, A., 2019. Topological map extraction from overhead images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1715–1724.

Ma, J., Zhao, J., 2017. Robust topological navigation via convolutional neural network feature and sharpness measure. IEEE Access 5, 20707–20715.

Machtans, C.S., 2006. Songbird response to seismic lines in the western boreal forest: A manipulative experiment. Canadian J. Zool. 84 (10), 1421–1430.

Marcos, D., Tuia, D., Kellenberger, B., Zhang, L., Bai, M., Liao, R., Urtasun, R., 2018. Learning deep structured active contours end-to-end. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8877–8885.

Mou, L., Zhu, X.X., 2018. Vehicle instance segmentation from aerial image and video using a multitask learning residual fully convolutional network. IEEE Trans. Geosci. Remote Sens. 56 (11), 6699–6711.

Nielsen, S.E., Bayne, E.M., Schieck, J., Herbers, J., Boutin, S., 2007. A new method to estimate species and biodiversity intactness using empirically derived reference conditions. Biol. Conserv. 137 (3), 403–414.

Ok, A.O., 2013. Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts. ISPRS J. Photogramm. Remote Sens. 86, 21–40.

Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39 (6), 1137–1149.

Ronneberger, O., Fischer, P., Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In: Proc. International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 234-241.

Schulz, K., Hänsch, R., Sörgel, U., 2018, October. Machine learning methods for remote sensing applications: An overview. In: Proc. Earth Resources and Environmental Remote Sensing/GIS applications, IX, vol. 10790, pp. 1079002.

Shannon, C.E., 1949. Communication in the presence of noise. Proc. IRE 37 (1), 10–21.

Song, G., Wang, Z., Bai, L., Zhang, J., Chen, L., 2020, September. Detection of oil wells based on faster R-CNN in optical satellite remote sensing images. In: Image and Signal Processing for Remote Sensing XXVI, Vol. 11533, pp. 114-121. SPIE.

Su, H., Wei, S., Liu, S., Liang, J., Wang, C., Shi, J., Zhang, X., 2020. HQ-ISNet: High-quality instance segmentation for remote sensing imagery. Remote Sens. 12 (6), 989.

Sun, L., Tang, Y., Zhang, L., 2017. Rural building detection in high-resolution imagery based on a two-stage CNN model. IEEE Geosci. Remote Sens. Lett. 14 (11), 1998–2002.

Wang, Z., Bai, L., Song, G., Zhang, J., Tao, J., Mulvenna, M.D., Bond, R.R., Chen, L., 2021. An Oil Well Dataset Derived from Satellite-Based Remote Sensing. Remote Sens. 13 (6), 1132.

Waqas Zamir, S., Arora, A., Gupta, A., Khan, S., Sun, G., Shahbaz Khan, F., Zhu, F., Shao, L., Xia, G.S., Bai, X., 2019. isaid: A large-scale dataset for instance segmentation in aerial images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 28-37.

Yang, Z., Li, J., Zipper, C.E., Shen, Y., Miao, H., Donovan, P.F., 2018. Identification of the disturbance and trajectory types in mining areas using multitemporal remote sensing images. Sci. Total Environ. 644, 916–927.

Yuan, Y., Chao, M., Lo, Y.C., 2017. Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. IEEE Trans. Med. Imaging 36 (9), 1876–1886.

Zeng, X., Wei, S., Wei, J., Zhou, Z., Shi, J., Zhang, X., Fan, F., 2021. CPISNet: delving into consistent proposals of instance segmentation network for high-resolution aerial images. Remote Sens. 13 (14), 2788.

Zhang, H., Chang, H., Ma, B., Wang, N., Chen, X., 2020, Dynamic R-CNN: Towards high quality object detection via dynamic training. In: Proc. pp. 260-275.

Zhang, N., Liu, Y., Zou, L., Zhao, H., Dong, W., Zhou, H., Zhou, H., Huang, M., 2018a, July. Automatic recognition of oil industry facilities based on deep learning. In: Proc. IGARSS, pp. 2519-2522.

Zhang, T., Zhang, X., Zhu, P., Tang, X., Li, C., Jiao, L., Zhou, H., 2021. Semantic attention and scale complementary network for instance segmentation in remote sensing images. IEEE Trans. Cybernet.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y., 2018b. Image super-resolution using very deep residual channel attention networks. In: Proc. ECCV, pp. 286-301.

Zhang, Z., Liu, Q., Wang, Y., 2018c. Road extraction by deep residual u-net. IEEE Geosci. Remote Sens. Lett. 15 (5), 749–753.

Zhao, K., Kang, J., Jung, J., Sohn, G., 2018, June. Building Extraction From Satellite Images Using Mask R-CNN With Building Boundary Regularization. In: CVPR Workshops, pp. 247–251.

Zhao, W., Persello, C., Stein, A., 2021. Building outline delineation: From aerial images to polygons with an improved end-to-end learning framework. ISPRS J. Photogramm. Remote Sens. 175, 119–131.

Zhou, L., Zhang, C., Wu, M., 2018. D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In: Proc. CVPR Workshops, pp. 182-186.

Zhou, M., Liu, T., Li, Y., Lin, D., Zhou, E., Zhao, T., 2019, May. Toward understanding the importance of noise in training neural networks. In: International Conference on Machine Learning. PMLR, pp. 7594-7602.