IJEGEO

International Journal of Environment and Geoinformatics 10(1):009-015 (2023)

Research Article

An Assessment of YOLO Architectures for Oil Tank Detection from SPOT Imagery

Tolga Bakırman

Yildiz Technical University, Faculty of Civil Engineering, Geomatics Engineering, Istanbul, TURKIYE

E-mail: bakirman@yildiz.edu.tr

Received 31.10.2022
Accepted 21.01.2023

How to cite: Bakırman, T. (2023). An Assessment of YOLO Architectures for Oil Tank Detection from SPOT Imagery, *International Journal of Environment and Geoinformatics (IJEGEO)*, 10(1):009- 000. doi.10.30897/ijegeo.1196817

Abstract

The inventory of oil tanks is essential for both economy and military applications, since it can be used to manage and estimate oil reserves. Knowing that the oil tanks contain valuable materials required for transportation and industrial production, they are a significant type of target. Oil tank detection techniques have several use-cases, including monitoring disasters, preventing oil leaks, designing cities, and assessing the damage. A huge amount of satellite imagery has recently been available and it is used in both military and civil applications. The new spaceborne sensors' higher resolution enables the detection of targeted objects with varying sizes. Therefore, remote sensing instruments provide ideal tools for oil tank detection tasks. Conventional approaches for oil tank detection from high resolution remote sensing imagery generally relies on geometric shape, structure, contract differences and color information of the boundary or hand-crafted features. However, these methods come along with vulnerabilities and hence it can be challenging to obtain accurate detection in the presence of a number of disturbance elements, particularly a wide range of colours, size variations, and the shadows that view angle and illumination create. Therefore, deep learning-based methods can provide a big advantage for the solution of this task. In this regard, this study employs four YOLO models namely YOLOv5, YOLOX, YOLOv6 and YOLOv7 for oil tank detection from high-resolution optical imagery. Our results show that YOLOv7 and YOLOv5 architectures provide more accurate detections with mean average precision (mAP) values of 68.11% and 69.69%, respectively. The experiments and visual inspections reveal that these models provide efficient, generalized and transferable results.

Keywords: Oil Tank, Detection, Deep Learning, YOLO

Introduction

Oil tanks are essential energy storage devices and critical infrastructures which are widely used in petroleum, natural gas, petrochemical industries and transportation (Xu et al., 2022; Yu et al., 2021). Rapid and accurate detection of oil tanks is substantial in terms of disaster management, risk evaluation and monitoring (Ok and Başeski, 2015; Xu et al., 2022).

With the advanced ability to acquire high resolution imagery, remote sensing has become an ideal tool for this task. There have been many approaches in the literature proposed for oil tank detection from remote sensing imagery. The first studies applied template matching and Hough transform methods for automatic oil tank detection which are generally based on structure and geometric shape information (Zhang et al., 2005; Zhu et al., 2012). These methods can be vulnerable to various factors such as scale, rotation, small targets, complex backgrounds and false detections. Saliency based methods were also exploited using synthetic aperture radar (Zhang and Liu, 2020; Zhang et al., 2019) and optical imagery (Liu et al., 2019). However, saliency methods rely on contrast differences and color information of the boundary which can also lead to false positives and false negatives. Zhang et al. (2015) have proposed machine learning based support vector machines (SVM) classification for extracted features using high resolution optical imagery. The drawback of machine learning methods is that they require manual feature extraction and feature selection processes.

Recent developments in artificial intelligence have also put forward deep learning-based methods which automatize these processes. In the literature, the implementation of deep learning is limited as an end-to-end solution for oil tank detection by researchers. Wang et al. (2019) have applied state-of-the-art Faster R-CNN and R-FCN architectures with increasing scales of the anchors. The authors obtained a recall value of around 80%. For boundary extraction of oil tanks, Yu et al. (2021) have developed Res2-Unet+ architecture that replaces the convolution block in the encoder of U-Net in order to decompose the feature map. In order to generalize their proposed model, the authors trained the model with high spatial resolution images from three different sensors. They obtained better accuracy values compared to state-of-the-art segmentation models such as U-Net, SegNet and PSPNet. Xu et al. (2022) present an improved version of EfficientDet architecture using 3-D deformable convolution, attention mechanism and focal loss function. They have achieved better accuracy results compared to various models which are generally region proposal-based architectures. Wu et al. (2022) proposed an improved version of YOLOX using transformers and VGG-like blocks specifically for SAR imagery. The authors have trained and evaluated their proposed model using Geofen-3 imagery. They have achieved 60.8% and 94.8% for mAP and mAP at 50% intersection over union threshold, respectively. Qi (2022) applied RetinaNet for the created a detection dataset composed of TripleSAT imagery which includes wind turbines, airplanes and oil tanks. For the oil tank class, they obtained 96.55% value for mean average precision value at 50% intersection over union threshold.

Literature review shows that there is no study-to-date that implements state-of-the-art YOLO architectures for oil tank detection using high resolution optical imagery. In this study, we aim to apply four versions of YOLO architectures namely YOLOX (Ge et al., 2021), YOLOv5 (Jocher, 2022), YOLOv6 (Li et al., 2022) and YOLOv7 (Wang et al., 2022) in order to investigate comparative performances for oil tank detection from high resolution optical imagery.

Materials and Methods

Most deep learning architectures requires huge amounts of data in order to be trained. In this study, we use the open access Airbus Oil Storage Detection (AOSD) dataset (AirbusGeo, 2021). The dataset consists of 98 SPOT RGB imagery around the world with roughly 1.2 meters of geometric resolution, 8-bit radiometric resolution and 2560 x 2560 pixels size. The dataset also provides 13,593 bounding boxes for oil tanks. A sample image and corresponding bounding boxes can be seen in Figure 1.

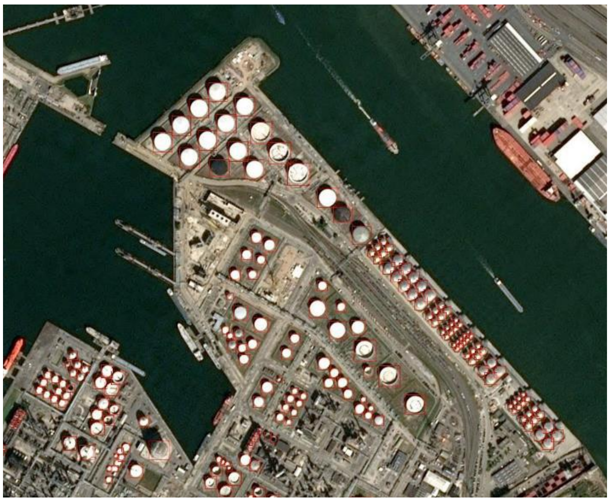


Fig. 1. A cropped image and bounding boxes (red) from the AOSD dataset.

This study employs four versions of YOLO architecture. YOLO architecture was first introduced by Redmon et al. (2016) which approaches the detection task as a regression problem based on Darknet architecture. In despite of popular Region Proposal Networks (RPN), YOLO would both predict bounding boxes and class probabilities in a single network. The main idea is based on a user-defined size grid cell in which responsible for detecting the object if it falls into the cell.

In order to decrease training time and produce a more generalized network, YOLO9000 (Redmon and Farhadi, 2017) has been proposed with some modifications to original YOLO such as batch normalization, higher resolution classifier and anchor boxes with convolutions (Atik et al., 2022).

In the next version of YOLO, the authors proposed a deeper network which is referred as Darknet-53. YOLOv3 (Redmon and Farhadi, 2018) uses 53 convolutional layers in combination with skip connections. Another advancement in the YOLOv3 is the multi-scale detector at it performs detections as 3 different scales using the feature pyramid network concept.

In the following years, YOLOv4 has been proposed by Bochkovskiy et al. (2020) in order to achieve the optimal balance between the number of convolutional layers and parameters, the input network resolution and the number of layer outputs. The Cross Stage Partial (CSP) (Wang et al., 2020) is integrated into Darknet-53 and the backbone is renamed as CSPDarknet-53 in which residual blocks are replaced with dense blocks. CSP allows to manage features better and to decrease the number of parameters. Additionally, Spatial Pyramid Pooling (SPP) module (He et al., 2014) is also integrated in order to expand the network’s receptive field. SPP removes the most important context elements while minimizing performance losses on the network.

Similar to YOLOv4, YOLOv5 uses CSP backbone and path aggregation network as neck. YOLOv5 is proposed in PyTorch environment with 4 different sizes in terms of network depth. Even though the developers have not published a paper regarding the details of YOLOv5, the main improvements in this version are mosaic data augmentation and auto learning anchors (Nelson and Solawetz, 2020). In this study, YOLOv5x version has been exploited.

YOLOX architecture has chosen a different approach and the developers have switched the YOLO detector with an anchor-free based detector. By doing this, the predictions for each location are reduced and only four values are predicted which are two offsets from grid corners, and the height and weight of the box. For the detection part, the couple head which performs classification and localization is decoupled with a lite head in order to improve converging speed. In this study, YOLOX-x version has been used.

YOLOv6 is a renovated version of YOLO in terms of network design which is based on RepVGG (Ding et al., 2021). It also includes VariFocal Loss (Zhang et al., 2021) and a combination of SIoU (Gevorgyan, 2022) and GIoU (Rezatofighi et al., 2019) for classification and regression loss, respectively. In this study, YOLOv6-L version has been exploited.

YOLOv7 is currently the latest version of the YOLO series. The main improvement is the backbone which is Extended Efficient Layer Aggregation. The architecture uses the cardinality of extending, mixing and combining to continuously improve the learning ability of the network without disturbing the original gradient path. The architecture uses group convolution to expand the channel of computational blocks. YOLOv7 concatenates layers together considering the scaling of depth and width of the network. The architecture also exploits gradient flow propagation paths in order to determine which modules need re-parameterization. Finally, an auxiliary head is integrated which is settled on a coarse-to-fine definition for better predictions. In this study, YOLOv7-X version has been implemented.

Results

In this study, oil tank detection performances of YOLOv5, YOLOX, YOLOv6 and YOLOv7 have been investigated using the AOSD dataset. Investigation of the dataset shows that raw image sizes are huge for training in terms of GPU memory. Moreover, almost half of the images have more than 100 oil tanks of varying sizes. This could lead to data loss during image resizing, especially for small objects. Therefore, we have decided to crop the dataset into 640 x 640 pixels sized images with a 10% overlap between chips within the image. In the end, we obtained 1021 images which are then split as 70% (714 images), 20% (205 images) and 10% (102 images for training, validation and testing, respectively.

All experiments were performed on the Google Colab environment. All architectures were trained for 100 epochs with 640 x 640 sized images using pre-trained weights obtained with the MS COCO dataset.

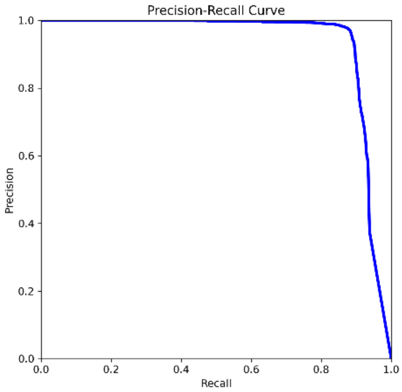
For the evaluation of the architectures, we have used average precision (AP) metrics of MS COCO evaluation. These are mean AP (mAP) at 50% intersection over union (IoU) threshold, mAP at 75% IoU threshold and mAP which is calculated as the average of 10 AP values for 10 IoU thresholds between 50% and 95% with 5% increments.

The calculated accuracy metrics for all YOLO models are given in Table 1. The values in Table 1 are calculated using the test dataset which has not been used during the training.

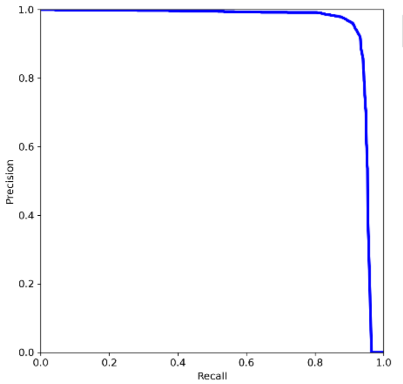
Table 1. Accuracy results for all models. The best values for each metric have been indicated with bold.

Model	mAP@0.50	mAP@0.75	mAP
YOLOv5x	93.40%	**79.43%**	**69.69%**
YOLOX-x	85.60%	61.90%	55.60%
YOLOv6-L	90.20%	74.70%	65.60%
YOLOv7-X	**94.45%**	77.24%	68.11%

The results show that YOLOv7 has the highest accuracy with 94.45% according to mAP@0.50 with only a marginal difference from YOLOv5 with 93.40%. However, it seems that YOLOv5 performs marginally better with increasing IoU threshold according to mAP@0.75 and mAP with 79.43% and 69.69%, respectively. This can also can be seen in the precision-recall curve presented in Figure 2.



(a)

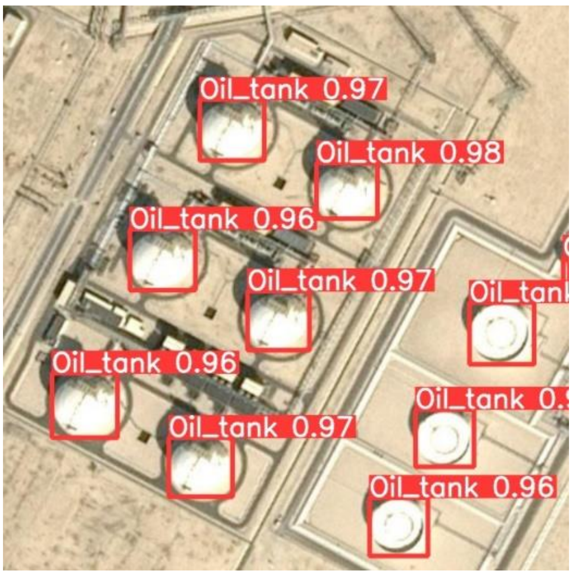


(b)

Fig. 2. Precision-Recall curve from test dataset for (a) YOLOv5 and (b) YOLOv7.

Accuracy values for YOLOv6 are slightly worse compared to YOLOv5 and YOLOv7. However, it can be said that YOLOv6 results are still efficient compared to YOLOX. mAP value of YOLOX is almost 15% lower than the most successful network. Additionally, it can be said that an mAP value of around 55% is quite low for a single class detection task.

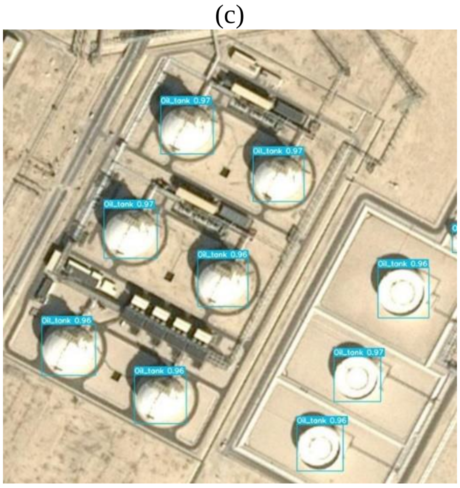
Figure 3 shows a detection sample for each network for the same image from the test dataset.



(a)

(b)

(c)



(d)

Fig. 3. A sample oil tank detection result for (a) YOLOv5, (b) YOLOX, (c) YOLOv6 and (d) YOLOv7.

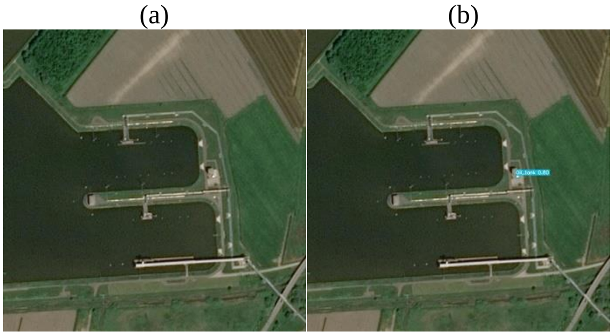
Discussion and Conclusion

In this study, state-of-the-art YOLO architectures were tested for oil tank detection from high resolution optical imagery. It can be said that YOLOv7 have provided the best solutions considering mAP@0.50 as Redmon and Farhadi (2018) suggested. Additionally, it is safe to say that a mAP value of 94.45% for 50% IoU threshold should be sufficient for this task in terms of usability and implementation.

It should be noted that oil tank detection is a challenging task due to varying shapes and especially sizes. Some oil tanks can be as small as 10 x 10 pixels in the image (Figure 4). Even though small objects in object detection applications can be problematic, YOLOv7 seems to successfully detect these small oil tanks. Inspection of the predictions of the test dataset also shows that confidence values for detections are quite high. Moreover, visual inspections show that medium and large-sized oil tanks are detected in most cases by all used architectures. Therefore, small targets are more decisive to determine which architecture is superior. Figure 5 shows a tiny oil tank example for all YOLO architecture. As can be seen in the figure, the oil tank could only be detected by YOLOv5 and YOLOv7 which are also the best architectures in terms of mAP.

Fig. 4. Detected small oil tanks by YOLOv7.

(a) (b)



(c) (d)

Fig. 5. A tiny oil tank detection result for (a) YOLOv5, (b) YOLOX, (c) YOLOv6 and (d) YOLOv7.

Fig. 6. Partly visible oil tank examples (top left).

Fig. 7. Extra image with snow cover from the AOSD dataset

Since the original images were divided into smaller chips, in some cases oil tanks are also split into multiple images. Nevertheless, YOLOv7 is still able to detect partly visible oil tanks (Figure 6). The AOSD dataset also provides a couple of extra images without ground truth. An image with a snow cover has been run for prediction with YOLOv7 (Figure 7). Considering that the training dataset does not contain any images with snow cover, the prediction results seem sufficient. However, it can be seen that shadows of the oil tanks are detected as False-Positive. Recently, deep learning-based methods have emerged as a key methodology for resolving issues with remote sensing. Deep learning-based techniques are crucial for improved, accurate, extensive, and quick data generation.

In this study, YOLOv7 and YOLOv5 have provided the best results among four YOLO models for oil tank detection using high resolution optical imaging. The accuracy and visual inspection results show that both architectures can be used efficiently for this task.

In future work, it is aimed to implement more MS COCO accuracy metrics to investigate the effect of object size more deeply and apply instance segmentation on oil tanks in order to extract also their boundaries. Moreover, the performance of the very recently released YOLOv8 for oil tank detection from high resolution optical imagery is worth investigating.

References

Atik, M. E., Duran, Z., Özgünlük, R. (2022). Comparison of YOLO Versions for Object Detection from Aerial Images, *International Journal of Environment and Geoinformatics*, 9(2), 87-93. doi: 10.30897/ijegeo.1010741.

AirbusGeo. (2021). *Airbus Oil Storage Detection*. https://www.kaggle.com/datasets/airbusgeo/airb us-oil-storage-detection-dataset

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

Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., Sun, J. (2021). Repvgg: Making vgg-style convnets great again. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,

Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J. (2021). Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*.

Gevorgyan, Z. (2022). SIoU Loss: More Powerful Learning for Bounding Box Regression. *arXiv preprint arXiv:2205.12740*.

He, K., Zhang, X., Ren, S., Sun, J. (2014, 2014//). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. Computer Vision – ECCV 2014, Cham.

Jocher, G. (2022). ultralytics/yolov5: v6. 2-YOLOv5 Classification Models, Apple M1, Reproducibility, ClearML and Deci. ai integrations. *Zenodo. org*.

Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W. (2022). YOLOv6: a single-stage object detection framework for industrial applications. *arXiv preprint arXiv:2209.02976*.

Liu, Z., Zhao, D., Shi, Z., Jiang, Z. (2019). Unsupervised Saliency Model with Color Markov Chain for Oil Tank Detection. *Remote Sensing*, *11*(9), 1089. https://www.mdpi.com/2072-4292/11/9/1089

Nelson, J., Solawetz, J. (2020). *Responding to the Controversy about YOLOv5*. https://blog.roboflow.com/yolov4-versus-yolov5/

Ok, A. O., Başeski, E. (2015). Circular Oil Tank Detection From Panchromatic Satellite Images: A New Automated Approach. *IEEE Geoscience and Remote Sensing Letters*, *12*(6), 1347-1351. https://doi.org/10.1109/LGRS.2015.2401600

Qi, W. (2022). Object detection in high resolution optical image based on deep learning technique. *Natural Hazards Research*. https://doi.org/10.1016/j.nhres.2022.10.002

Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition,

Redmon, J., Farhadi, A. (2017). YOLO9000: better, faster, stronger. Proceedings of the IEEE conference on computer vision and pattern recognition,

Redmon, J., Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,

Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*.

Wang, C. Y., Liao, H. Y. M., Wu, Y. H., Chen, P. Y., Hsieh, J. W., Yeh, I. H. (2020, 14-19 June 2020). CSPNet: A New Backbone that can Enhance Learning Capability of CNN. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),

Wang, Y., Zhang, Q., Zhang, Y., Meng, Y., Guo, W. (2019). Oil Tank Detection from Remote Sensing Images based on Deep Convolutional Neural Network. *Remote Sensing Technology and Application*, *34*(4), 727-735. https://doi.org/10.11873/j.issn.1004-0323.2019.4.0727

Wu, Q., Zhang, B., Xu, C., Zhang, H., Wang, C. (2022). Dense Oil Tank Detection and Classification via YOLOX-TR Network in Large-Scale SAR Images. *Remote Sensing*, *14*(14), 3246. https://www.mdpi.com/2072-4292/14/14/3246

Xu, S., Zhang, H., He, X., Cao, X., Hu, J. (2022). Oil Tank Detection With Improved EfficientDet Model. *IEEE Geoscience and Remote Sensing Letters*, *19*, 1-5. https://doi.org/10.1109/LGRS.2022.3183350

Yu, B., Chen, F., Wang, Y., Wang, N., Yang, X., Ma, P., Zhou, C., and Zhang, Y. (2021). Res2-Unet+, a Practical Oil Tank Detection Network for Large-Scale High Spatial Resolution Images. *Remote Sensing*, *13*(23), 4740. https://www.mdpi.com/2072-4292/13/23/4740

Zhang, H., Wang, Y., Dayoub, F., Sunderhauf, N. (2021). Varifocalnet: An iou-aware dense object detector. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,

Zhang, L., Liu, C. (2020). Oil Tank Extraction Based on Joint-Spatial Saliency Analysis for Multiple SAR Images. *IEEE Geoscience and Remote Sensing Letters*, *17*(6), 998-1002. https://doi.org/10.1109/LGRS.2019.2937355

Zhang, L., Shi, Z., Wu, J. (2015). A Hierarchical Oil Tank Detector With Deep Surrounding Features for High-Resolution Optical Satellite Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(10), 4895-4909. https://doi.org/10.1109/JSTARS.2015.2467377

Zhang, L., Wang, S., Liu, C., Wang, Y. (2019). Saliency-Driven Oil Tank Detection Based on Multidimensional Feature Vector Clustering for SAR Images. *IEEE Geoscience and Remote Sensing*

Letters, *16*(4), 653-657. https://doi.org/10.1109/LGRS.2018.2878106

Zhang, W., Zhang, H., Wang, C., Wu, T. (2005, 29-29 July 2005). Automatic oil tank detection algorithm based on remote sensing image fusion. Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05.,

Zhu, C., Liu, B., Zhou, Y., Yu, Q., Liu, X., Yu, W. (2012, 22-27 July 2012). Framework design and implementation for oil tank detection in optical satellite imagery. 2012 IEEE International Geoscience and Remote Sensing Symposium,