Aarhus Universitet

Geografisk dataanalyse (147201E002)

S24-o Aarhus Hjemmeopgave

Predefined Information

Start date:	24-05-2024 14:00 CEST
End date:	07-06-2024 14:00 CEST
Term:	S24-o
Grading scale:	Danish 7-point scale
Flow code:	254967.123970.149.o.1.137417
Internal assessor:	Adéla Sobotkova

Participant

Name:	Márton Kardos
WAYF ID:	689890@au.dk
Studienummer:	202105399
AU ID:	au689890

Information from participant

Characters - paper *:	13641

Can the submission be used anonymously for educational purposes?:
Yes

Declaration of Honesty *:
Yes

Group

Group name:	Márton Kardos
Group number:	18
Other members:	The participant has submitted as a one-person group

Only Looking Once, from Afar

Pretraining YOLO on Satellite Imagery Improves Burial Mound Recognition

Márton Kardos
(Studienr. 202105399)

Faculty of Arts
Aarhus University
June, 2024

Abstract

Preserving cultural heritage sites is of high importance for archaeological and historical studies. This is, however, no easy task, especially in the case of Thracian and Thraco-Roman burial mounds in Bulgaria, where it is expected that a sizeable portion of tumuli are left to be discovered and recorded by the research community, but are often looted or damaged by agricultural activity. Researchers have explored the possibility of using satellite imagery for manual and automated mound detection measures. Sobotkova et al., 2024 finetuned a pretrained convolutional neural network classifier to identify whether small windows of satellite images contain a burial mound, with limited success. This study explores whether, and to which extent, the performance deficit observed in previous work can be attributed to modelling choices. It is demonstrated that pretraining on satellite imagery greatly improves models' ability to learn meaningful patterns in burial mound recognition in comparison to vision models pretrained in other domains. Utilizing the YOLO architecture (Redmon et al., 2016) also eliminates false detections due to confounding features, thereby significantly reducing false positive rates. Three models are released under a non-commercial open licence, along with the `burial-mounds` Python package (https://github.com/x-tabdeveloping/burial-mounds-object-recognition) for ease of use and utilities for preprocessing and model training. The trained models demonstrate no signs of overfitting, thereby suggesting that results could be improved with more training or larger models. While the presented evidence is promising, even the best model performs quite poorly quantitatively, only correctly identifying 3 mounds out of 25 in the validation set, and proposing 6 false positives. Reasons for poor performance are discussed, and future directions are proposed.

1 Introduction

1.1 Burial Mounds in Bulgaria

Bulgaria is home to tens of thousands of burial mounds, mostly of Thracian origin, some of which are UNESCO World Heritage sites (UNESCO World Heritage Centre, 1979). While Bulgaria hosts some of the world's most culturally significant and well known tumuli, such as the Kazanlak and Sveshtari tombs, a large portion of mounds have never been adequately studied or recorded. These, mostly unknown sites, often times fall prey to looters due to a lack of or inadequate law enforcement and limited information about their locations in authorities' hands (Loulanski and Loulanski, 2017).

The TRAP project (Ross et al., 2018) has tried utilizing satellite imagery for easier detection of burial mounds. It was, however, found that many of the mounds found by annotators were false positives, hardly visible mounds often times remained undiscovered, and that training made errors less likely (Sobotkova et al., 2024).

This prompted Sobotkova et al. (2024) to investigate whether computer vision could yield better and more effective mound detection measures due to less human bias and higher cost-effectiveness with unsatisfactory results.

1.2 Object Recognition

Object recognition is a machine learning task that entails two crucial steps: (1) Identifying bounding boxes for objects in an image (2) identifying which class is present in the bounding box (Zhao et al., 2019).

A naive paradigm for object recognition is to use a classification model (typically a convolutional neural network) to identify whether a sought object occurs in evenly sized and spaced windows over the target image and what class the object is (Felzenszwalb et al., 2008). More involved approaches, such as region-based convolutional neural networks, use a separate model for proposing potential **regions of interest** in an image (Girshick et al., 2014).

Redmon et al. (2016) proposed a novel method for object detection, termed YOLO. YOLO models predict both object locations, proposed label and confidence with the same convolutional neural network. YOLO models can also be used with **oriented bounding boxes** or OBB (Zand et al., 2022), where the model also predicts the angle of the box in which the object is located.

Predicting bounding boxes and labels with the same model not only increases model throughput and training efficiency, but also allows models to have lower false positive rate, due to having access to more visual context. It has been observed that classification-based recognition can propose false positives due to correlated features. Sobotkova et al. (2024), who utilized a CNN classifier over a sliding window, have, for instance, reported that many false mound detections were roads, possibly due to their frequent occurrence in proximity of tumuli. The YOLO Model architecture might provide remedy to this issue.

1.3 Transfer Learning

Transfer learning or knowledge transfer refers to a deep learning model's capability to better adapt to novel tasks, which are similar to tasks it has been previously trained on (Weiss et al., 2016). It is common practice in contemporary machine learning to publish **foundation models** (Bommasani et al., 2022). These are parameter-rich models that have typically been trained on large datasets, which are thought to represent much of the variation in a given domain. These models learn to effectively represent useful aspects of the training data in dense numerical representations, commonly referred

to as **embeddings** (Liu et al., 2020, Kiela and Bottou, 2014). For particular applications, especially with small amounts of data, it is advisable to either **fine-tune** a foundation model, or utilize its embeddings as input features.

Sobotkova et al. (2024) have utilized the VGG16 foundation model (Simonyan and Zisserman, 2015), which was pretrained on image classification. While it is conceivable that VGG16's ability to better represent low-level image features would be helpful for mound recognition, a model pretrained on satellite imagery might adapt easier to the problem domain.

Furthermore, Sobotkova et al. (2024) have merged the infrared and red colour channels, leading to unnatural looking images, that might be hard for foundation models to interpret.

1.4 Data Augmentation

Data augmentation refers to the practice of introducing random alterations to the training data in order to avoid overfitting and improve models' generalization capabilities and robustness to noise (Mumuni and Mumuni, 2022). In the case of computer vision, this typically takes form as rotations, flips, zooming, panning, stretching, introducing noise, cutting out areas, mosaicking and colour shifting (Shorten and Khoshgoftaar, 2019). Data augmentation has been demonstrated effective, especially in the case of small datasets (Brigato and Iocchi, 2021), such as the annotated Bulgarian mounds studied by Sobotkova et al., 2024.

1.5 Contribution

This study is dedicated to investigating whether (1) using a more appropriate architecture, (2) pretraining on remote sensing imagery and (3) using more extensive data augmentation can improve detection of burial mounds in satellite imagery without improvement to data quality.

2 Methodology

2.1 Datasets

The base models released by Ultralytics were originally pretrained on large image datasets. OBB models were pretrained on the **DOTA** dataset (Xia et al., 2019), which contains oriented bounding boxes for various vehicles and landmarks in satellite imagery from Google Earth, as well as JL-1 and GF-2 satellites. Detection models were pretrained on **OpenImages V7** (Benenson and Ferrari, 2022), a dataset containing annotations of 600 common objects in approximately 9 million images. Notably, the dataset does not contain remote sensing imagery.

Two datasets were utilized for training models in this study. The **Burial Mounds** dataset contains bounding box annotations of 773 mounds in the Kazanlak Valley collected by the TRAP project (Sobotkova and Ross, 2018). Remote sensing imagery originates from IKONOS satellites with a 1m/pixel resolution. All models presented in this study were finetuned on this dataset for mound detection.

xView is a large dataset containing bounding boxes for 60 classes of landmarks and vehicles in high-resolution remote sensing imagery (0.3m/pixel). Imagery was recorded by WorldView3 satellites. The xView dataset was utilized to fine-tune *detection* base models on satellite imagery before fine-tuning on burial mound detection.

Dataset	Resolution	# Classes	# Images	# Instances	Task
OpenImages V7	*NA*	600	9M	74.7M	Detection
DOTA	Mixed	15	2806	188282	OBB
xView	0.3m/pixel	60	859	601937	Detection
Burial Mounds	1m/pixel	1	1	773	Detection

Table 1: Dataset overview

2.2 Preprocessing

Satellite images in the Mounds dataset were preprocessed with the following steps to make them adequate for YOLO model training: The GeoTIFF satellite image was split into smaller images of size 2048x2048. Due to the relatively large size, a train-test split emulates domain shift effectively, which is likely to occur in an applied setting. 20% percent of the images were reserved for validation. The red, green and blue colour channels stacked into images and were min-max normalized to gain images that are close to natural colour. The infrared channel was ignored. For training, all annotated mounds that had bounding boxes smaller than 10 pixels were removed, as these boxes might not contain enough information to elicit a detection. All labels were converted into a YOLO-compatible format in the xView and Mounds datasets.

2.3 Model Training

Three models were trained to assess the efficacy of object detection models for burial mound recognition, all of them based on the same architecture, but with different training procedures:

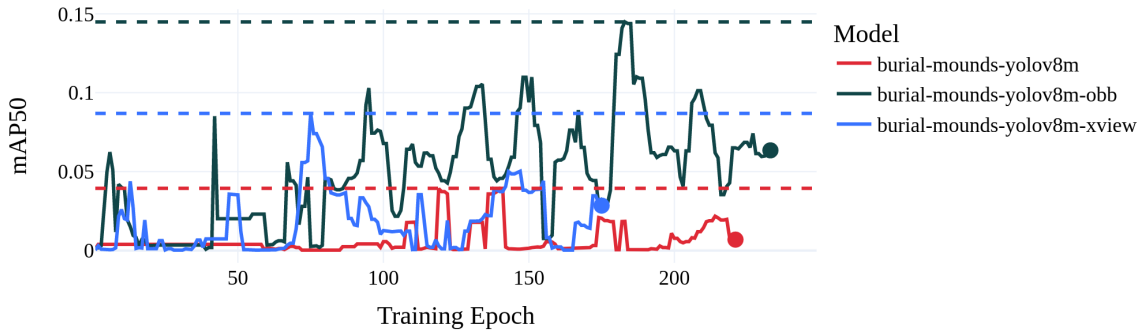
1. **burial-mounds-yolov8m**: As a baseline, a YOLOv8 medium model pretrained on OpenImagesV7 was fine-tuned on the Mounds dataset with 300 training epochs.
2. **burial-mounds-yolov8m-xview**: To see whether pretraining on satellite images could enhance performance, the same base model was also fine-tuned in two steps: (1) 25 epochs on xView (2) 300 epochs on the Mounds dataset.
3. **burial-mounds-yolov8m-obb**: A YOLOv8 medium model with an OBB head, which was previously pretrained on the DOTA dataset, was fine-tuned for 300 epochs on the Mounds dataset.

The same data augmentation was used on all datasets across all experiments. This included: randomly scaling images, flipping images vertically with a 30% probability and horizontally with 50% probability, random rotations, colour shifts, mosaicking, hue, saturation and brightness shifts and random translations. Training was stopped early if the models did not improve over 100 epochs. All models were optimized using ADAM with a learning rate of 0.01.

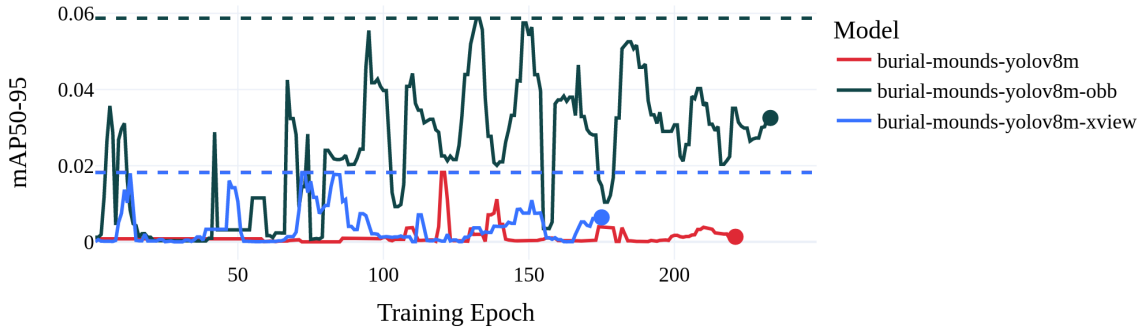
Further evaluation was conducted on the best performing models over the 300 epochs. Models' weights were saved at the best performing epochs, and were released under a non-commercial open licence.

3 Results

All models performed poorly at identifying mounds in the validation set. While the OBB model correctly identified 3 previously unseen mounds, it proposed 6 false positives and missed 22 mounds. The base model made no proposals for mounds thereby missing all tumuli in the validation set. The xView-finetuned model correctly identified one mound, while missing all others, but not proposing any false positives.



(a) mAP50



(b) mAP50-95

Figure 1: Mean Average Precision Achieved by Models During Training
Dots indicate early stopping; Highest performance indicated by dashed lines

Mean average precision values were recorded both with an intersection over union (IoU) threshold over 50% (mAP50) and an average over thresholds ranging from 50 to 95 (mAP50-95).

None of the models' performance was satisfactory. The best results were achieved by the **burial-mounds-yolov8m-obb** model, with a maximum mAP50 score of 0.145 and maximum mAP50-95 score of 0.059. To put this into context, the same model architecture's best mAP50-95 score on

OpenImages V7 was 0.336, where the number of classes was much higher (Ultralytics, 2024).

The models' performance was incredibly volatile during training. Adjusting the learning rate would probably decrease volatility.

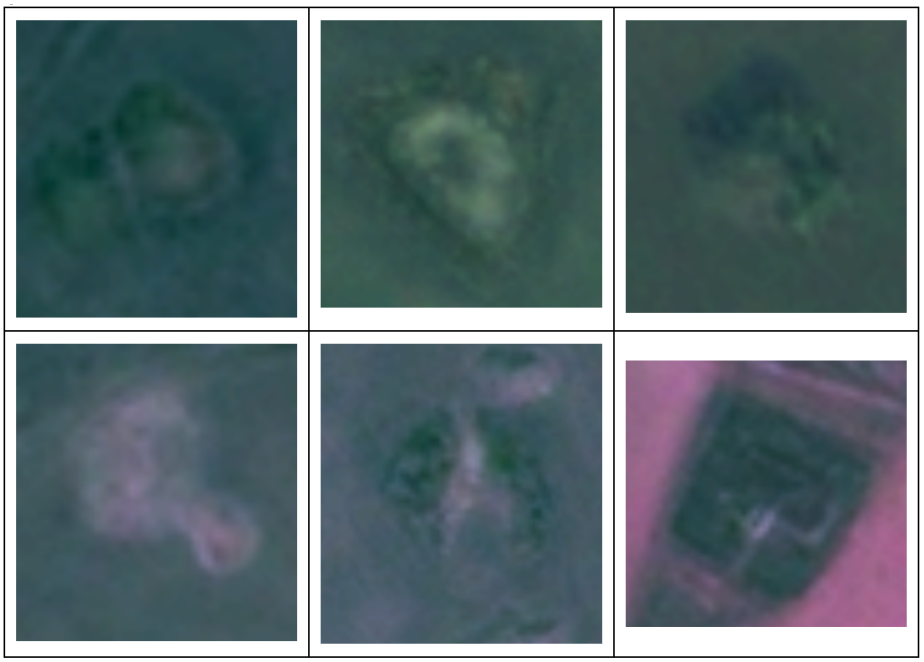


Figure 2: Examples of False Positives Detected by the **burial-mounds-yolov8m-obb** Model

The only model substantially improving during training was the OBB model, demonstrating the utility of pretraining on satellite images. Note that the xView model achieved higher mAP50 scores than the baseline, even though it was only trained on xView for 25 epochs. The baseline's inability to increase performance throughout training indicates that pretraining on remote sensing data is important to facilitate transfer learning, and simply having a good representation of low-level image features is not enough.

Qualitative investigations of the models' predictions on the training and validations sets reveal the following patterns:

1. The models did not overfit on the training set, still proposing false positives and not recognizing all mounds in training images. (see images with yellow background in Figure 3)
2. Some of the largest mounds did not get recognized by the models in the training set. This could probably be improved with more aggressive image scaling.
3. False positives proposed by the OBB model typically do resemble mounds, with most of them having a circular shape with some lighter patterns in the middle. (see Figure 2)
4. Most of the mounds not recognized by **burial-mounds-yolov8m-obb** were covered by vegetation or were hardly visible on the images.

This is a notable qualitative improvement over Sobotkova et al. (2024)'s results, as they reported frequent false positives where no mound-like shapes were present. As previously mentioned, YOLO models are less likely to misrecognize correlated features for the target.

4 Discussion

While the performance of the models presented in this study is far from ideal for the problem domain, and would not be sufficient for real-world burial mound detection, evidence suggests that there is room for improvement in this field, even with low-quality data. Careful choices in model architecture and training procedure are crucial for achieving better results, especially with small datasets of low quality.

A number of promising paths could be taken by future endeavours to improve on these results: (1) Using larger pretrained models with higher performance. As no model overfit on the training set, it would be reasonable to assume that a larger model might be better able to learn from the training data. (2) More training epochs on xView and DOTAv1 could render *detect* models more useful for mound recognition, and (3) Adjusting hyperparameters might decrease volatility in training.

It would, however, be most effective to increase the quality of the training data. More accurate bounding boxes, along with higher resolution satellite imagery (possibly obtained from third party APIs), would likely greatly increase performance, especially on smaller mounds.

5 Conclusion

By pretraining on remote sensing object recognition, and choosing a more appropriate model architecture, qualitatively better results could be achieved in burial mound recognition with machine learning models. While results are far from satisfactory, evidence suggests that there is promise in the field. As such, multiple paths for improvement are proposed. Three models are released under a non-commercial open licence, along with the `burial-mounds` Python package (https://github.com/x-tabdeveloping/burial-mounds-object-recognition) for streamlined model training and inference.

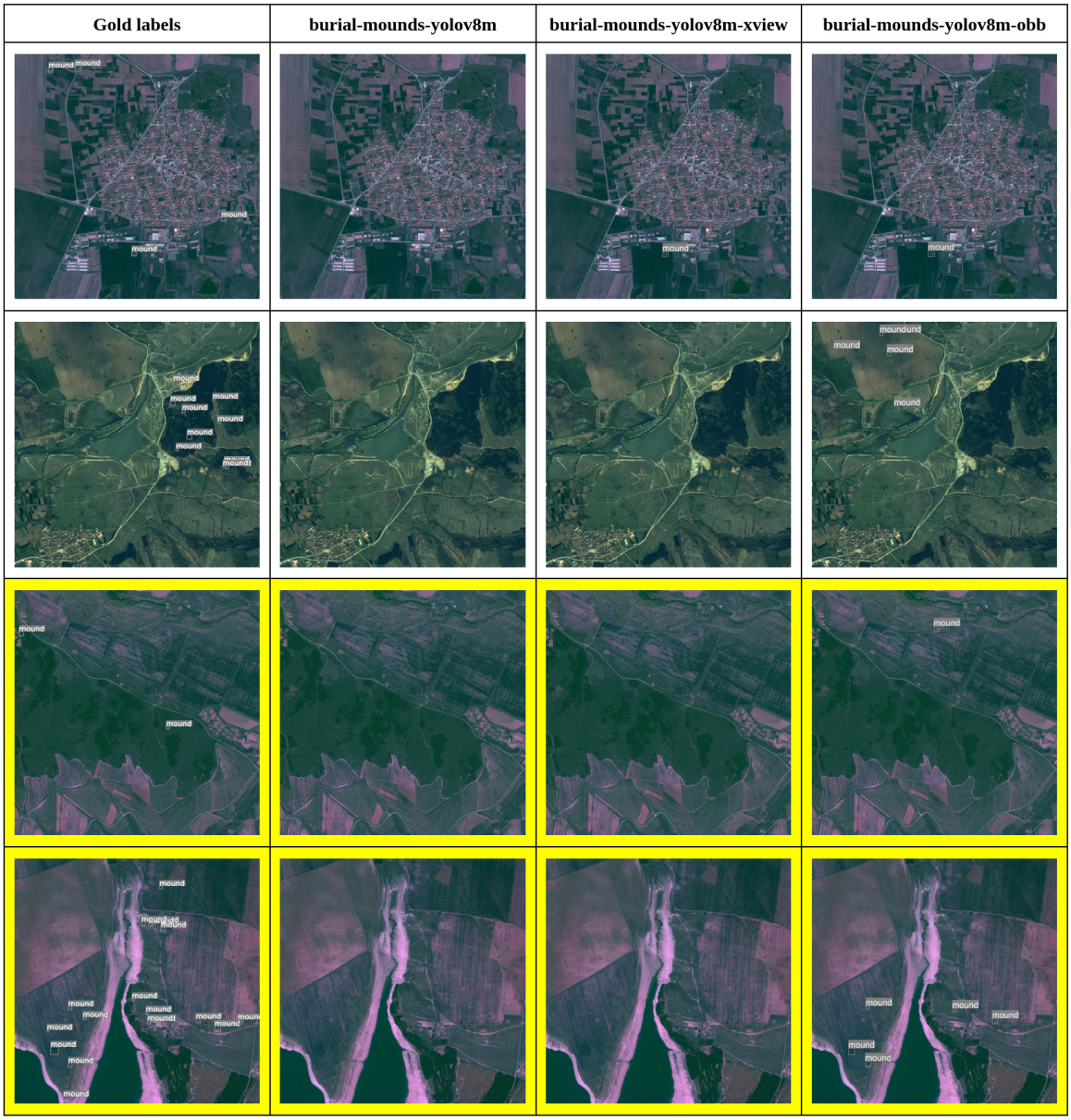


Figure 3: Examples of Gold Labels against Model Predictions
Yellow background indicates that the image is from the training set.

References

Benenson, R., & Ferrari, V. (2022). From colouring-in to pointillism: Revisiting semantic segmentation supervision.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2022). On the opportunities and risks of foundation models.

Brigato, L., & Iocchi, L. (2021). A close look at deep learning with small data. *2020 25th International Conference on Pattern Recognition (ICPR)*, 2490–2497. https://doi.org/10.1109/ICPR48806.2021.9412492

Felzenszwalb, P., McAllester, D., & Ramanan, D. (2008). A discriminatively trained, multiscale, deformable part model. *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8. https://doi.org/10.1109/CVPR.2008.4587597

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation.

Kiela, D., & Bottou, L. (2014). Learning image embeddings using convolutional neural networks for improved multi-modal semantics. *Proceedings of the 2014 Conference on empirical methods in natural language processing (EMNLP)*, 36–45.

Liu, Q., Kusner, M. J., & Blunsom, P. (2020). A survey on contextual embeddings.

Loulanski, T., & Loulanski, V. (2017). Thracian mounds in bulgaria: Heritage at risk. *The Historic Environment: Policy & Practice*, *8*(3), 246–277. https://doi.org/10.1080/17567505.2017.1359918

Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, *16*, 100258. https://doi.org/10.1016/j.array.2022.100258

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection.

Ross, S., Sobotkova, A., Nekhrizov, G., Tzvetkova, J., & Connor, S. (Eds.). (2018). *The tundzha regional archaeology project: Surface survey, palaeoecology, and associated studies in central and southeast bulgaria, 2009-2015 final report*. Oxbow Books.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0197-0

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition.

Sobotkova, A., Kristensen-Mclachlan, R. D., Mallon, O., & Ross, S. A. (2024). Validating predictions of burial mounds with field data: The promise and reality of machine learning. *Journal of Documentation*. https://api.semanticscholar.org/CorpusID:269392634

Sobotkova, A., & Ross, S. (2018). Kazanlak survey results [Version archived for private and non-commercial use with the permission of the author/s and according to publisher conditions. For further rights please contact the publisher.]. In S. Ross, A. Sobotkova, G. Nekhrizov, J. Tzvetkova, & S. Connor (Eds.), *The tundzha regional archaeology project* (pp. 66–81). Oxbow Books.

Ultralytics. (2024, June). Yolov8. https://docs.ultralytics.com/models/yolov8/#supported-tasks-and-modes

UNESCO World Heritage Centre. (1979). Thracian tomb of kazanlak. https://whc.unesco.org/en/list/44

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*(1). https://doi.org/10.1186/s40537-016-0043-6

Xia, G.-S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., & Zhang, L. (2019). Dota: A large-scale dataset for object detection in aerial images.

Zand, M., Etemad, A., & Greenspan, M. (2022). Oriented bounding boxes for small and freely rotated objects. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–15. https://doi.org/10.1109/TGRS.2021.3076050

Zhao, Z.-Q., Zheng, P., Xu, S.-T., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(11), 3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865