MACQUARIE University
SYDNEY·AUSTRALIA

Automated Identification of Burial Mounds in Soviet Maps of Bulgaria Project

---

Analysis Report

Prepared By:

Rodulfo Dela Paz Alfafara II

January 21, 2024

1. Project Overview

1.1. Objective:

The primary objective of this project is to develop a Proof-of-concept for an automated process for identifying potential burial mounds from a given dataset of old Soviet maps of Bulgaria. The project aims to provide a baseline for a study on whether an ML-based automated approach can outperform the manual process. One of the central questions driving the project is understanding the relationship between the size of the training set and the precision of the model. This also involves investigating how many labeled points are required to achieve a certain level of precision in identifying burial mounds.

1.2 Scope:

The project focused on leveraging machine learning models, specifically the YOLO model to identify specific target markings on the maps, represented by hairy brown circles. These were not present on the original maps and were manually added during the dataset creation. Given that the project focused on using YOLO, the models were evaluated based on metrics prescribed by YOLO such as precision, recall, and accuracy.

1.3 Limitations:

The project only made use of 20 Geotiff images as the dataset. GIS-generated shapefiles were used to provide metadata information on the respective images such as the location of the symbols present on the map, coordinate systems used, and image dimensions among other things.

2. Analysis Framework

To address the Project objective, given the project duration of only 1 month, delivery of the MVP(Minimum Viable Product), which is to be able to provide a proof-of-concept ML model that can be trained and evaluated on Soviet maps with burial mounds labels was prioritized. The activities have been grouped into two: MVP phase and Post-MVP. Post-MVP phase aimed to explore other methods that could further improve/tune the output of the MVP phase.

MVP Phase

PREPARE

During this phase, all pertinent sources of information related to the project have been gathered and supplementary research were done for the following:

- Project Kickoff
- YOLO setup

- Python packages setup
- Gathering of Data: Download and access of data files
  - 20 Geotiff image files
  - GIS-generated Shapefiles containing information on the supplied Geotiff image files
- Gathering of Knowledge Enhancement References
  - YOLO
  - GIS

ANALYSIS

After gathering all the important information mentioned above, the following analysis activities have been performed:

- YOLO Familiarization
- GIS Familiarization
- Data Exploration
  - What data were available? Format, Size, Quality, Dimensions etc
  - What data was needed? Produced?
  - Dataset Cleanup/Preparation
- Machine Learning Application – how machine learning can be applied to the use-case of identifying potential burial mounds.
- Model Design
- Train and Test Design

RECOMMENDATION EXECUTION

During this phase, the model has been trained and tested based on the findings of the analysis activities performed above. The model design has been iteratively improved or changed based on the results of the model evaluation and feedback from the team.

<u>POST- MVP Phase</u>

PREPARE

- Machine Learning/Other Automation Techniques Applications on Similar Use-Cases. During this phase, we tried to explore other approaches that can be used to improve the models such as splitting/cropping, rotating of images.

ANALYSIS

- Supplementary Data Analysis
  - Relationship Analysis
  - Consideration of the symbols other than the hairy brown circles
- Analysis of Techniques/Methods gathered from the prepare phase
  - Can it be applied to our project
  - How it can be applied

RECOMMENDATION EXECUTION

- A series of models applying different variations of splits, rotations, symbols included, dataset size and availability of background images have been trained and

tested. The models have been iteratively improved or changed based on the results of the model evaluation and the feedback of the team.

3. Analysis Findings Summary

Model Variations (Best – Worst)

BEST TRAINED MODEL (36 Splits)

The model was tested on the **test set (unseen data)** and had the following performance scores:

```
val: New cache created: /content/drive/.shortcut-targets-by-id/1fs0dPYE2LSV-MIvSUUs2UEqqm
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95):
                   all         69        138      0.957      0.975      0.965      0.664
```

Here, we can see that the model has high P, R, and MAP 50 values which means that it is performing well on objects that are easy to be detected but we can see that the MAP 50-95 values are significantly and relatively lower which means that the model's performance diminishes as the complexity of predictions increase.

Approach: Part of the success of this model is based on the analysis finding that a model performs best when trained and detected on the default image size of 640 and the 36 splits made the cropped images' size really close to 640. Optimizing the bounding box dimensions to encapsulate the symbols more precisely and aligning the training images with the default size of the model (640) which was achieved by segmenting the original map image into 36 smaller, focused sections during the training phase while a dynamic sliding window method was employed, systematically traversing the map(all 36 cropped images) and seamlessly consolidating the outcomes. This approach ensured enhanced model compatibility with default image sizes, facilitating more accurate symbol segmentation and comprehensive training and testing coverage.

Out of all the model variations that have been trained, the model found to perform the best was trained with the following characteristics:

- YOLOv8 parameter settings
    - 640 image size (default setting)
    - 100 epochs
    - Default confidence and IOU threshold rates
    - Default learning rates

- Bounding Box – adjusted to encompass only the hairy brown circles

- Dataset - 19 images with 1409 labels of hairy brown circles have been split/segmented into 36 partitions/subsets.

- Training Dataset
    - Train set with 553 images
    - Validation set with 62 images

- Evaluation Dataset
    - Test set with 69 images

You can find all the details on the model's performance in this google drive link: https://drive.google.com/drive/folders/1FFPlMIN47F0iwyldeDS7xT5lOFC7AI8v?usp=drive_link

36 Splits with 45, 90, 135, 180, 225, 270, 315 degree rotations

The model was tested on the **test set (unseen data)** and had the following performance scores:

```
val: New cache created: /content/drive/.shortcut-targets-by-id/1fs0dPYE2LSV-MIvSUUs2UEqqmD_
                Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 1
                  all        548        999      0.931      0.957      0.948      0.599
Speed: 0.9ms preprocess, 5.9ms inference, 0.0ms loss, 3.1ms postprocess per image
Results saved to runs/detect/val
```

Here, we can see that the model with no rotations performed slightly better. This means that rotations, in our case, did not contribute to significant improvements for the model.

The model was trained with the following characteristics:

- YOLOv8 parameter settings
    - 640 image size (default setting)
    - 100 epochs
    - Default confidence and IOU threshold rates
    - Default learning rates

- Bounding Box – adjusted to encompass only the hairy brown circles

- Dataset - 19 images with 1409 labels of hairy brown circles have been split/segmented into 36 partitions/subsets with copies of the partitions/subsets rotated in the following degree rotations: 45, 90, 135, 180, 225, 270, 315

- Training Dataset
    - Train set with 4432 images
    - Validation set with 492 images

- Evaluation Dataset
    - Test set with 548 images

You can find all the details on the model's performance in this google drive link:
https://drive.google.com/drive/u/1/folders/1UBDr4t1ZmTIn634tlq2zED3P0CeqR3gh

36 Splits With No Background Images

The model was tested on the **test set (unseen data)** and had the following performance scores:

```
val: New cache created: /content/drive/.shortcut-targets-by-id/1fs0dPYE2LSV-MIvSUUs2UEqqm
                Class     Images  Instances      Box(P          R      mAP50  mAP50-95):
                  all         69        139      0.941      0.916      0.934      0.613
Speed: 0.2ms preprocess, 7.3ms inference, 0.0ms loss, 12.5ms postprocess per image
Results saved to runs/detect/val
```

Based on this and on the visual detection results, it can be observed that the model trained with background images performed better, and had higher confidence values on the predictions made. Moreover, it was observed that this model missed predicting hairy brown circles that do not look like the typical hairy brown circles. The model trained with background images was able to consistently predict these.

The model was trained with the following characteristics:

- YOLOv8 parameter settings
    - 640 image size (default setting)
    - 100 epochs
    - Default confidence and IOU threshold rates
    - Default learning rates

- Bounding Box – adjusted to encompass only the hairy brown circles

- Dataset - 19 images with 1409 labels of hairy brown circles have been split/segmented into 36 partitions/subsets

- Training Dataset
    - Train set with 343 images
        - No Background Images
    - Validation set with 62 images
        - Consists of Background Images

- Evaluation Dataset
    - Test set with 69 images
        - Consists of Background Images

You can find all the details on the model's performance in this google drive link:
https://drive.google.com/drive/u/1/folders/1cKR0WQyLv2nq9TYY3FAPi0NOpUOm7pDA

49 Splits

The model has not been tested on an unseen test dataset since during the time this was tested, the aim was to look for the number of splits that will give the best results to a performance of a model. The result of having 49 splits was close to the results of having 36 splits but the one with 36 splits was slightly better, hence, the decision to conduct more testing activities on 36 splits.

The model was trained with the following characteristics:

- YOLOv8 parameter settings
  - 640 image size (default setting)
  - 100 epochs
  - Default confidence and IOU threshold rates
  - Default learning rates

- Bounding Box – encompasses a lot more space beyond the bounds of the hairy brown circles

- Dataset - 20 images with 1494 labels of hairy brown circles have been split/segmented into 49 partitions/subsets.

- Training Dataset
  - Train set with 980 images
  - Validation set with 980 images

- Evaluation Dataset
  - No Evaluation Set

You can find all the details on the model's performance in this google drive link:
https://drive.google.com/drive/u/1/folders/1qjMKxSv9YgKP5YqBXy0jcacJyZlJ2J4N

36 Splits With Other Symbols

The model was tested on the **test set (unseen data)** and had the following performance scores:

```
cache created: /content/drive/.shortcut-targets-by-id/1fs0dPYE2LSv-M1vS0oS2oEqqm
        Class     Images  Instances      Box(P          R      mAP50  mAP50-95):
          all         23        106      0.716      0.759      0.813      0.488
       mound          23         41      0.534      0.976      0.845      0.515
 Black diamond        23         65      0.898      0.543      0.782      0.461
.2ms preprocess, 22.8ms inference, 0.0ms loss, 29.9ms postprocess per image
saved to runs/detect/val
```

Here, we can see that the model is able to predict mounds however, it also predicts other symbols to be mounds, which made the results poor. Training the model using symbols other

than the hairy brown circles made the model perform worse. The addition of the other symbols confused the model and began predicting black symbols as mounds. It is not ideal to include symbols other than the hairy brown circles especially considering that we do not have a lot of data available for training.

The model was trained with the following characteristics:

- YOLOv8 parameter settings
    - 640 image size (default setting)
    - 100 epochs
    - Default confidence and IOU threshold rates
    - Default learning rates

- Bounding Box – adjusted to encompass only the symbols

- Included All Symbols other than hairy brown circles

- Dataset - 19 images have been split/segmented into 36 partitions/subsets.

- Training Dataset
    - Train set with 553 images
    - Validation set with 62 images

- Evaluation Dataset
    - Test set with 69 images

You can find all the details on the model's performance in this google drive link:
https://drive.google.com/drive/u/1/folders/1rRiydqFyM4F8ffvzFG_T33mTPQw7cYcc

16 Splits

The model has not been tested on an unseen test dataset since during the time this was tested, the aim was to look for the number of splits that will give the best results to a performance of a model.

The model was trained with the following characteristics:

- YOLOv8 parameter settings
    - 1600 image size and 640 image size (default setting)
    - 100 epochs
    - Default confidence and IOU threshold rates
    - Default learning rates

- Bounding Box – encompasses a lot more space beyond the bounds of the hairy brown circles

- Dataset - 20 images with 1494 labels of hairy brown circles have been split/segmented into 16 partitions/subsets.

- Training Dataset
    - Train set with 320 images
    - Validation set with 320 images

- Evaluation Dataset
    - No Evaluation Set

You can find all the details on the model's performance in this google drive link: https://drive.google.com/drive/u/1/folders/14-gvqwG_NGKwfWep8n31XltV_l9IU9QM

9 Splits

The model has not been tested on an unseen test dataset since during the time this was tested, the aim was to look for the number of splits that will give the best results to a performance of a model.

The model was trained with the following characteristics:

- YOLOv8 parameter settings
    - 1600 image size and 640 image size (default setting)
    - 100 epochs
    - Default confidence and IOU threshold rates
    - Default learning rates

- Bounding Box – encompasses a lot more space beyond the bounds of the hairy brown circles

- Dataset - 20 images with 1494 labels of hairy brown circles have been split/segmented into 16 partitions/subsets.

- Training Dataset
    - Train set with 180 images
    - Validation set with 180 images

- Evaluation Dataset
    - No Evaluation Set

You can find all the details on the model's performance in this google drive link: https://drive.google.com/drive/u/1/folders/1sqbl0yld92Q_x0xldcQ_8TPnT8Co1u-u

No Splits

The model has not been tested on an unseen test dataset since during the time this was tested, the aim was to resolve the issues encountered when running the YOLO model. It was not significant anyway since the model performed really poorly.

Issues Resolved During This Phase:

- Model was unable to make predictions:
  - Training Set & Image Size – It has been found that training on just 3 images with a small image size does not really help the model learn anything. When Training Set & Image Size have both been increased, the model was able to learn to include bounding boxes while training.
- Missing MAP (Mean Average Precision) values
  - YOLOv5 V.S. YOLOv8 Comparative Analysis
    - Troubleshooting Steps:
      - downgrading CUDA version of GPU in Colab
      - tweaking the amp setting within YOLOv5 train.py
      - tweaking of parameters thresholds, size, learning rates, etc
      - use of YOLOv5

    After performing all these steps and training using YOLOv5, MAP values were still missing, and it has been confirmed that the issue does not lie in bugs coming from different YOLO versions. Upon training on YOLOv5, a warning was received stating that all the 1494 labels are less than 3 pixels in size, which is one of YOLO's limitations. This led to a realization that the reason as to why MAP values are missing is because the labels are so small in size that the YOLO model is not learning enough for it to produce these computation outputs. To confirm this, changes have been applied on how the labels were annotated. This time, following Steve's suggestion to ensure that the boxes are enclosing the hairy brown circles and not just its center point. After applying said changes, labels already became larger than the YOLO pixel size requirement. Consequently, running the model has successfully produced MAP values, which confirms that the issue really lies on label annotations pixel size. The model was not learning enough for it to generate MAP value computations.

4. Other Findings

5.1 YOLO version Power - In our comparison, it has been found that yolov8 is actually more powerful than yolov5 because when the same training parameter settings have been used, despite using label annotations that are less than the prescribed 3 pixel size, the YOLOv8 model was still able to learn as its box loss scores were not stuck at 0, contrary to YOLOv5. Though highly inaccurate, the model was able to make predictions whereas the trained YOLOv5 could not.

5. Recommendations For Future Enhancement

- Colab's GPU: With regards to the compute units consumptions in Google Colab, the following have been observed.

  One run of training on ~500-600 dataset with 100 epochs using YOLO's default image size of 640:

  - T4 GPU used up 0.91 compute units
  - A100 GPU used up 1.99 compute units
  - V100 GPU used up 0.84 compute units

  Shared this information for future reference in case the project would subscribe to Colab's subscription plans with compute units restriction.

6. Appendix

I. Github Repository : https://github.com/stevecassidy/burial-mounds-ml/tree/main

This repository consists of all the notebooks used for data exploration, data preprocessing, model training, model detection and evaluation. This also consists of the dataset and the shapefiles used in the project.

II. Training Results:
https://drive.google.com/drive/u/1/folders/1KHbHDAV_gaeNuT2bGdt4mmjYAZwB6mKi

This drive contains the results of all model variations discussed above.