3rd Place Solution Report
Arm UNICEF Disaster Vulnerability Challenge

Team Members:

KSOURI Azer

TUO Muhamed

BOUAZIZ Mohamed

HAKIM Ahmed

Contents

List of Figures

Abstract

Malawi has faced numerous natural disasters and climatic shocks in recent years, including droughts, floods, and landslides. These events, along with the economic repercussions of COVID-19 and other global issues, have severely impacted the health and well-being of most Malawians, especially those in rural areas, which comprise over 80% of the population.

Significant progress has been made globally in mapping flood extents and the resulting damages using satellite imagery. However, there are still gaps in accurately determining the number of affected people, particularly in Malawi's rural regions. Many rural homes are built with traditional grass-thatched roofs, which are often overlooked by algorithms that use satellite or aerial imagery to identify populations or buildings impacted by floods.

1 Overview and Objectives

1.1 Solution Description and Purpose

The project aims to develop an advanced machine-learning algorithm designed to accurately count various types of roofed houses in aerial (drone) imagery of rural areas in Malawi. This initiative responds to the critical need for precise population estimation and disaster impact assessment in regions vulnerable to frequent natural disasters such as droughts, floods, and landslides. These events, compounded by global challenges like the COVID-19 pandemic, dispro portionately affect Malawi's rural population, which comprises over 80Traditional rural housing in Malawi often features grass-thatched roofs, which are challenging to identify accurately using conventional satellite and aerial imaging techniques. The proposed solution leverages state-of-the-art machine learning models trained on high-resolution drone imagery to distinguish between different roof types, including thatch, tin, and others.

1.2 Objectives and Expected Outcomes

1. **Algorithm Development**: Develop and deploy a machine-learning algorithm capable of automatically identifying and counting thatch, tin, and other roof types in aerial imagery with high accuracy and reliability.

2. **Enhanced Disaster Response**: Improve the speed and accuracy of disaster impact assessments, enabling quicker and more effective responses from humanitarian organizations and government agencies.

3. **Commmunity Resilience**: Empower rural communities in Malawi by providing accurate data for disaster preparedness planning, enabling proactive measures to mitigate risks and reduce vulnerabilities.

2 Method

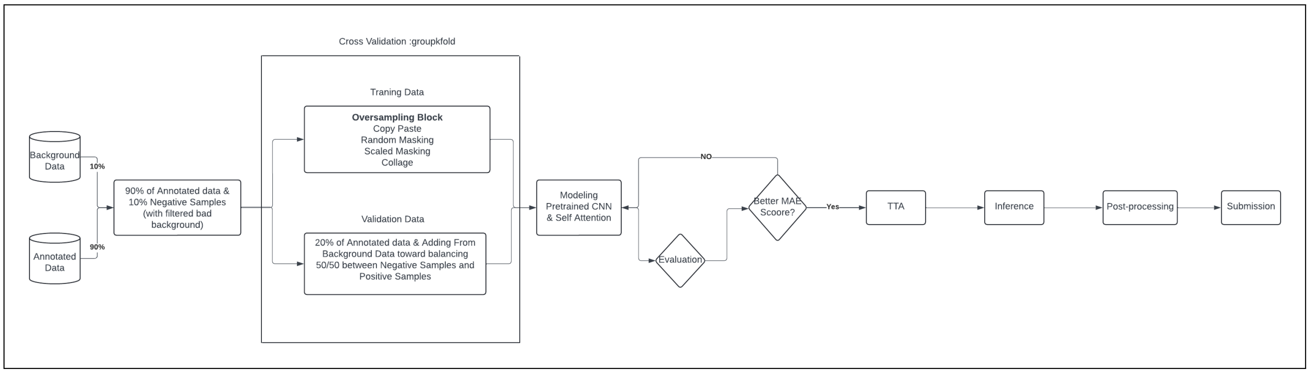

Figure 1: Pipeline Diagram

3 ETL process

3.1 Extract

3.1.1 Data Sources:

The primary data source for this project is aerial imagery captured over various regions of Malawi. These images serve as the basis for labeling different types of roofed houses, including those with grass-thatched roofs, tin roofs, and other roof types.

3.1.2 Data Formats:

The aerial imagery is typically stored in standard image formats such as tif.

3.1.3 Considerations:

- **Data Volume:** The dataset comprises 4,772 training images and 2,045 testing images, totaling 6,817 images. This volume is manageable for initial model training and evaluation.
- **Frequency of Extraction:** Data extraction occurs once initially to acquire the dataset for model development. Additional extractions may be necessary to update the dataset or expand it based on ongoing project needs.

3.2 Transform

3.2.1 Transformation Logic:

- **Data Cleaning:** Initially, background images are filtered, retaining only 10% for training to focus on relevant housing structures.
- **Data Preprocessing:** Images undergo preprocessing steps such as resizing and normalization to ensure uniformity and enhance model performance.

3.2.2 Augmentation Techniques:

- **Copy-Paste:** Inserts annotated objects onto varied background scenes to diversify training data.

Figure 2: Image before CopyPaste

Figure 3: Image after CopyPaste

- **Random Image Masking:** Applies geometric shapes to obscure parts of images, simulating occlusions and environmental variability.

Figure 4: Image before Random Image Masking

Figure 5: Image after Random Image Masking

- **Masked Scale:** Resizes and positions images within a defined canvas to simulate different object scales and viewpoints.

Figure 6: Image before Masked Scale

Figure 7: Image after Masked Scale

- **Collage:** Assembles smaller images into composite visuals, enriching the dataset with diverse visual compositions.

Figure 8: Image with collage

- **Geometric Augmentations:** Horizontal Flip, Transpose and Random Rotate 90 with probability = 0.5
- **Blur/Noise Augmentations:** Gaussian Blur with probability = 0.15 and Gaussian Noise with probability = 0.25

3.2.3 Purpose:

These transformation techniques aim to enhance dataset diversity, improve model generalization, and prepare the data to effectively train a machine-learning algorithm for accurate roofed house classification.

3.3 Load

3.3.1 Loading Process:

- **Storage Mechanisms:** Transformed and augmented data is stored in a structured format suitable for model training, typically organized into directories or databases.
- **Indexing and Optimization:** Data may be indexed based on categories (e.g., grass-thatched, tin, other roofs) for efficient retrieval during model training and inference.

3.3.2 Data Handling:

- **Inference Environment:** Ensures compatibility between the loaded data format and the model's input requirements for seamless inference.

3.3.3 Optimization Strategies:

- **Batch Processing:** Utilizes batch processing techniques during model training to handle large volumes of data efficiently.
- **Parallel Processing:** Implements parallel processing where feasible to expedite data loading and preprocessing stages.

4 Data modeling

4.1 Description of Data Models

The solution utilizes a combination of multi-label regression model to achieve accurate house roof classification.

4.1.1 Multi Label Regression:

Two models are employed for the multi-label regression component:

- **EfficientNetV2_rw_s**:
  - **Image Size**: 1024
  - **Epochs**: 25
  - **Training Batch Size**: 8
- **ConvNext_base**:
  - **Image Size**: 512
  - **Epochs**: 25
  - **Training Batch Size**: 8

  Both models incorporate a Self-Attention mechanism to capture long-range dependencies in the image data, improving the model's ability to understand complex spatial relationships.

4.2 Assumptions and Theoretical Foundations

- **Self-Attention Mechanism**: This mechanism helps in capturing long-range dependencies within the image data, enhancing the model's capability to recognize and classify different roof types under varying conditions.

- **Pre-trained Models**: Utilizing pre-trained models like EfficientNetV2 and ConvNext_base leverages transfer learning, which allows the models to benefit from prior training on large datasets, thereby improving performance and reducing training time.

4.3 Feature Selection, Engineering, and Normalization

Feature Selection and Engineering:

- **Self-Attention Layers**: Two self-attention layers are used to refine the features extracted by the backbone models.
- **Fully Connected Layer**: A fully connected layer maps the features to the final output, predicting the number of different roof types (grass-thatch, tin, other).

Normalization:

- Images are normalized to have a consistent size (1024x1024 for EfficientNetV2_rw_s and 512x512 for ConvNext_base) to ensure uniformity in input data, which aids in stable and efficient training.

4.4 Model Training

Training Process:

- **Framework**: PyTorch Lightning is used for training and validation, providing a robust and scalable framework.
- **Criterion**: Mean Absolute Error (MAE) is used as the evaluation metric to measure the average magnitude of errors in predictions without considering their direction.
- **Optimizer**:
  - **Learning Rate (lr)**: 1×10^{-4}

  * **Weight Decay**: 1×10^{-4}
  * **Warmup Steps**: 0

4.5 Model Validation

Validation Data: 20% of the annotated data combined with background data to balance the dataset, achieving a 50/50 split between negative and positive samples.
Validation Process:

- **Cross-Validation groupkfold**: Ensures the model is validated on different images of data, reducing the risk of overfitting and ensuring generalizability.
- **Performance Metrics**: Mean Absolute Error (MAE) is the primary metric used to evaluate the model's performance, providing a clear measure of prediction accuracy.

Model Evaluation:

- **Mean Absolute Error (MAE)**: This metric is chosen because it directly reflects the average error magnitude, which is crucial for understanding the model's precision in predicting the number of houses with different roof types.

5 Run time

- **EfficientNetV2_rw_s**:20 hours
- **ConvNext_base**:10 hours
All models are trained on an NVIDIA RTX A6000.

6 Performance metrics

6.1 Model Accuracy

- **Mean Absolute Error (MAE)**: MAE was used as the primary metric to evaluate model performance. It measures the average magnitude of errors in predictions, providing a clear indication of prediction accuracy without considering the direction of errors.
- **Cross-Validation (CV) Score**: The cross-validation score helps in assessing the model's performance stability across different subsets of the training data. It ensures that the model is not overfitting and generalizes well to unseen data.

6.2 Reported Scores

Table 1: Reported Scores

Public Score	Private Score	CV
0.277717391	0.279860302	0.283440860215053

7 Inference

7.1 Input and Output of New Data

Input Data:

- New aerial images are captured using drones and uploaded to the cloud storage.
- The images are preprocessed to match the input specifications of the model, such as resizing and normalization.

- The preprocessed images are then sent to the deployed model via API requests.

Output Interpretation:

- The model outputs the predicted counts of different roof types (thatch, tin, other) for each image.
- These predictions merge the outputs from both models using the harmonic mean formula:

$$\frac{2 \cdot \text{Pred1} \cdot \text{Pred2}}{\text{Pred1} + \text{Pred2} + \epsilon}$$

7.2 Test Time Augmentation (TTA)

Additional augmentations during inference:

- RandomRotate90
- HorizontalFlip

7.3 Post-processing

This post-processing modifies the Target column in a DataFrame based on specific conditions. For values greater than 40, it converts them to integers and increments by 1. For values 40 or below, it rounds them to the nearest integer, ensuring they are at least 0. This ensures that predictions are within a practical range, with no negative values and proper rounding applied.

8 Conclusion

The development and deployment of a machine-learning algorithm to accurately count grass-thatch, tin, and other roofed houses in aerial imagery for rural Malawi addresses a critical need in disaster response and management. By leveraging advanced techniques in data processing, model training, and inference, this project provides a robust solution for estimating affected populations in the event of natural disasters.