**Dissertation Draft**

**Topic:** To leverage deep learning techniques for analyzing satellite imagery to monitor environmental changes.

Abhinav Kumar

Btech CSE, VIT Bhopal

Table of Content

1. Introduction……………………………………………………………………………………….……III
   1. Background……………………………………………………………….……………….………III
   2. Aim and Objectives……………………………………………………..……………………..III
   3. Research Question……………………………………………..………………………………IV
   4. Structure and Dissertation……………………………………………...................….IV
2. Literature Review…………………………………………………………......................……..IV
3. Methodology……………………………………………………………....................……,,.……V
   1. System Design…………………………………………………………....................….……V
   2. Tools And Technology Used/Model Used…………....................………………VI
   3. Evaluation……………………………………………....................………………………....VI
4. Implementation………………………………………………………....................…………....VI
5. Results………………………………………………………………....................………………....VII
6. Conclusion………………………………………………....................……………………….……IX
   1. Achievements……………………………………………………....................……….……IX
   2. Limitations………………………………………………………....................…………..….IX
   3. Answer to Research Question…………………………....................……………….X
   4. Future Work……………………………………………………....................…………….…X
7. References………………………………………………………....................…………………...XI

**1.Introduction**

* 1. **Background**

Land cover and land use (LULC) monitoring is crucial for environmentally sustainable urban planning and climate change mitigation. With the deployment of numerous Earth-observation satellites, the amount of high-resolution remote sensing images become available, which is an ideal source of information for the task. However, the largeness and the complexity of this data render the traditional manual analysis infeasible and ineffective. Over past decade, deep learning has become a disruptive technology that automates the analysis of satellite data. In particular, Convolutional Neural Networks (CNNs) have shown excellent performance in learning hierarchical features from raw pixel data automatically. This study investigates the use of a compact, lightweight CNN to create a sustainable, scalable model for automated LULC classification, a critical task in environmental monitoring.

* 1. **Aims and Objectives:**

The primary aim of the research was to design, implement, and evaluate an efficient deep learning model for the automated classification of land use and land cover from Sentinel-2 satellite imagery.

To achieve this aim, the following objectives were established:

* To collate and organize the EuroSAT public dataset for a supervised ML problem, with training and validation splits.
* To develop a lightweight and low computational cost Convolutional Neural Network framework, namely MobileNetV2, using transfer learning with the weighed pre-trained ImageNet weights.
* To facilitate the development and testing of the model in a resource-restricted, cloud-based environment (Google Collab with a T4 GPU), speedup of the development process.
* Both the quantitative and qualitative measures based on a wide range of metrics: overall accuracy, a confusion matrix, a classification report, illustrating good performance of the model and comparing with other works, the visual inspection of predictions.
  1. **Research Questions**

The research questions that the study attempts to address are:

* How well can a pretrained, lightweight CNN model such as MobileNetV2 classify various land cover types using Sentinel-2 satellite data?
* What are the accuracy and computational performance trade-offs when applying a speed-oriented model such as MobileNetV2 to a remote sensing task?
* Which specific land cover classes within the EuroSAT dataset are the most challenging for the model to distinguish, and what might be the underlying reasons for this confusion?
  1. **Structure of the Dissertation**

The dissertation is structured as follows: Chapter Two is a review of the related work that covers remote sensing methods from traditional mapping to more modern deep learning methods; Chapter Three covers the methodology, including the system design, datasets, model architecture, and evaluation metrics; Chapter Four describes the actual implementation of the model in the Google Collab environment; Chapter Five provides and discusses the experimental results, and Chapter Six summarizes and concludes the dissertation, including findings of the research study, the limitations of the work, and directions for future research.

**2.Literature Review**

Remote sensing has experienced tremendous changes in methodology for analyzing satellite data. Originally, methodologies were based on manual photo-interpretation, which had high subjectivity, high human input, and was very time consuming. Digital imagery allowed us to develop the automated pixel-based methods such as NDVI that are fairly basic methods but offer limited ability to capture complex spatial context. OBIA improved upon this methodology by segmenting an image and creating meaningful objects in an image before classifying them, however it often relied on a significant human investment for rule-setting and feature engineering.

The recent revolution of deep learning has revolutionized the landscape for satellite imagery analysis. CNNs take advantage of the spatial hierarchical structure to learn a pool of features from raw pixel inputs through the learning process, without requiring explicit feature extraction. First generation architectures like AlexNet, VGG and ResNet established new state-of-the-art in general computer vision and were rapidly adapted for remote sensing tasks, exhibiting superior performance.

Unfortunately, these deep architectures can be expensive to compute in practice, especially for large scale applications. This has led to the emergence of lightweight, efficient CNNs that were designed for resource constrained environments. MobileNetV2, introduced by Sandler et al. (2018), is one such architecture that makes use of inverted residuals and linear bottlenecks. MobileNetV2 effectively reduces the number of parameters and computations while still maintaining high accuracy. The success of these models is based on the principle of transfer learning, where a model is fine-tuned for a specific task using a model pre-trained on a large dataset, such as ImageNet. This allows for the use of the general features learned from a large dataset, such that a (relatively) small target dataset can achieve faster convergence and improved performance.

The EuroSAT dataset proposed by Helber et al. (2019) has emerged as a standard benchmark for LULC classification - it is composed of Sentinel-2 satellite images encompassing 10 distinct Land Cover classes, and provides a clean, structured foundation for model evaluation and comparison. The research draws from these advancements in the space of LULC classification by specifically focusing on the real-world application of a lightweight mobile convolutional neural network, MobileNetV2 in combination with transfer learning, on the EuroSAT benchmark dataset aiming to develop an accurate but also highly efficient and accessible model.

**3.Methodology**

**3.1 System Design:**

The proposed imagery classification algorithm flows through a multi-step process for supervised image classification. We begin with data acquisition; specifically we will acquire the EuroSAT dataset that is publicly available. Next follows data preparation, where we systematically partition the dataset into two subsets: training (x85%) and validation (x15%). In this case, we will use the 15% validation subset for an unbiased evaluation of the model's generalization performance. The crux of the system is model development; we will choose MobileNetV2, which is pre-trained on the ImageNet, for this classification task and adapt it to the specific classification task. We then train the model on the training subset and assess its performance with the validation subset. Lastly we subject the best model to further evaluation and review its accuracy and its confusion.

**3.2 Tools and Technologies Used/Models used**

Dataset: The EuroSAT dataset forms the basis of this research. It contains 27,000 labelled Sentinel-2 images with a size of 64x64 pixels. The dataset has been balanced over ten LULC classes: 'AnnualCrop', 'Forest', 'HerbaceousVegetation', 'Highway', 'Industrial', 'Pasture', 'PermanentCrop', 'Residential', 'River', and 'SeaLake'.

Model: The basis of the system is a MobileNetV2 architecture. This is a lightweight Convolutional Neural Network (CNN) model that strikes a great balance between performance and computational efficiency. The way the model was implemented uses a version pre-trained on the ImageNet dataset, which used pre-learned visual features. The last classification layer of the neural network was removed and replaced with a new linear layer with 10 outputs, consistent with the number of classes represented within the EuroSAT dataset.

Software & hardware: The project was developed in Python 3, using the PyTorch deep learning framework. Training and evaluation were conducted in Google Colaboratory, running an NVIDIA T4 GPU, which is offered by Google within the tier free of charge to users to enhance and accelerate computational workflow. The torchvision library was used for data loading, and model implementation.

**3.3 Evaluation**

The model's efficacy was evaluated on the held-out validation set. All models were evaluated with the overall accuracy as the main saving best model criterion during training. Finally, ten alternatives were left in a confusion matrix for better evaluatory purposes. Additionally, a classification report was generated to study the per-class precision, recall, and F1-score, which were capable of detailing the model’s performance on each land cover type.

**4.Implementation**

The entire work was done in one Google Collab notebook for reproducible convenience. The first step was to download the EuroSAT dataset from a direct URL with the wget command and unzip the contents of the dataset in the collab environment. Then, a Python script was used to split the 27,000 images into a train and val directory.

Data loading was simplified using the PyTorch torchvision.datasets.ImageFolder class and the images also underwent a standard set of transformations, including conversion to a PyTorch tensor and normalization of the pixel intensity values.

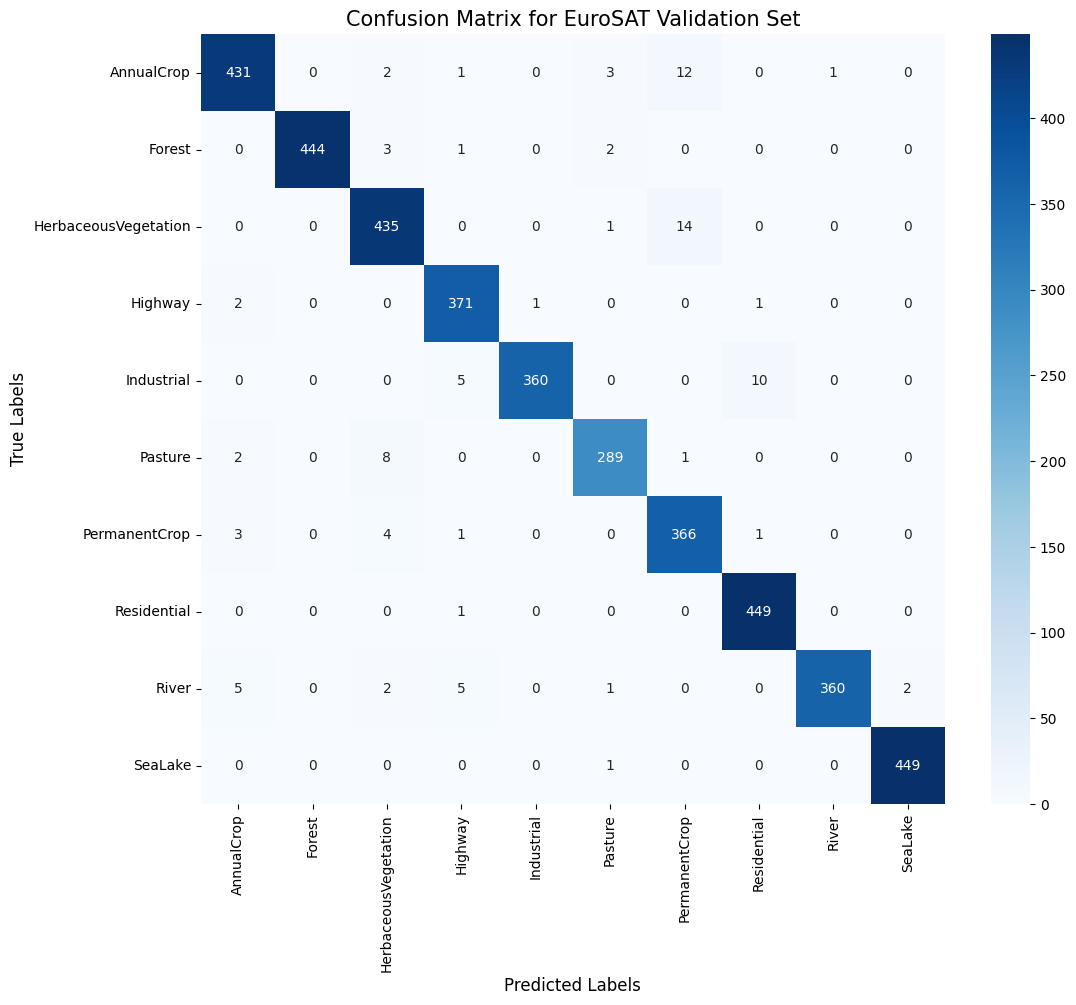
The MobileNetV2 model was read in from the torchvision.models library using default ImageNet pre-trained weights. The fully connected layer at the end of the model called classifier[1] was replaced with a torch.nn.Linear layer adjusted for 10 output classes.

The training loop was run for a total of 15 epochs. The Adam optimizer was selected due to its adaptive learning rate properties. The Cross-Entropy Loss Function was used because it is standard practice for multi-class classification tasks. The training process was accelerated significantly with Automatic Mixed Precision (AMP) through PyTorch's torch.cuda.amp.GradScaler which enabled computation to take place at faster speeds through the usage of lower precision floating point numbers where it is possible and without degrading the model performance. At the end of every epoch, the model's accuracy was evaluated on the validation set.

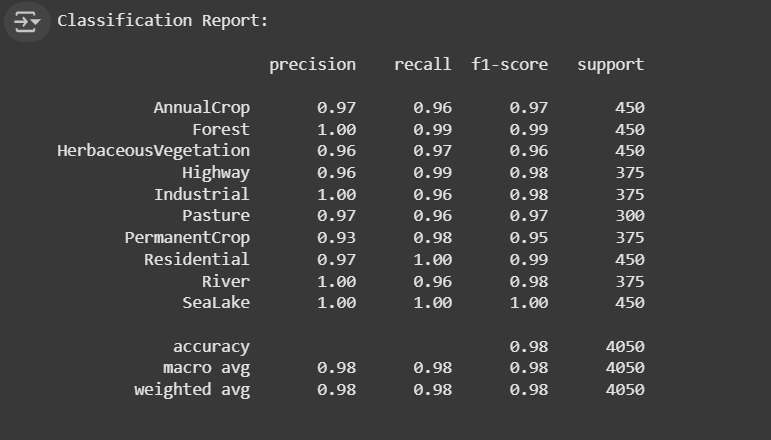
**5.Results**

The MobileNetV2 model performed remarkably well on the LULC classification task. For the held-out validation set, the model achieved a final overall accuracy of 98%, indicating a very high degree of accuracy across the 10 classes. A more thorough review of the confusion matrix provides more relevant detail for our analysis. The results clearly show that the suitability of values recorded in the diagonal, consistent with the factual evidence of high accuracy. However, review of the confusion matrix was able to demonstrate specific areas of systematic confusion. The model tended to confuse classes with similar visual and spectral properties. For example, there were occasions where 'HerbaceousVegetation' was classified as a 'PermanentCrop' (14 occurrences), and 'AnnualCrop' was classified as 'Pasture' (12 occurrences). There was a similar trend with human structures, where 'Industrial' sometimes confused with 'Residential' (10 occurrences). These observations were also further quantified in the classification report, where all the classes had F1-scores of 0.95 or higher. Classes with very identifiable features like 'Forest', 'Industrial', 'River', and 'SeaLake' all achieved F1-scores of 0.98 or above. 'PermanentCrop', at 0.93, was our lowest precision class, which corresponded nicely with the confusion evidenced in the confusion matrix.

A qualitative analysis of misclassified images verified these findings. The model is able to greatly differentiate between high-level categorizations (for example: telling an environment is 'natural' or 'man-made') and the greatest errors occur in either of these categorizations, (mistaking PermanentCrop for HerbaceousVegetation or Industrial for Residential, for example). This leads to the conclusion that the model is learning high-level features of different landscapes, and it is struggling on the lower-level class (more visually similar) subclasses.



***Figure 1: Confusion Matrix of the MobileNetV2 model on the EuroSAT validation set, showing high accuracy along the diagonal.***

****

***Figure 2: Classification Report detailing the precision, recall, and F1-score for each land cover class.***

******

***Figure 3: A sample of misclassified images, highlighting the model's confusion between visually similar sub-categories.***

**6. Conclusion**

**6.1 Achievements**

The study accomplished the objective of proving the effectiveness of a lightweight deep learning architecture, to effectively classify satellite imagery in an automated manner. A MobileNetV2 model was able to be implemented and trained, achieving an overall accuracy of 98% on the EuroSAT benchmark dataset. The efficiency of the overall process was also an achievement; by using a pre-trained and lightweight model, and modern training strategies (Automatic Mixed Precision), we were able to develop and train a model quickly, even in a resource-constrained, free, cloud-based context. The work demonstrates a pragmatic and accessible way to develop LULC classifiers that are accurate.

* 1. **Limitations**

This research is not without its limitations. First, while the EuroSAT dataset is excellent for benchmarking, it included pre-selected, 64x64 image patches. This means that the model has not been evaluated on raw, large-format satellite scenes that include imagery from differing geographic areas, atmospheric distortions, and potentially greater intra-class drift. Second, this research we are used image-level classification methods, where a single label is provided for each image patch. This is a simplification of real-world environmental monitoring, which often requires more sophisticated and can require pixel-based semantic segmentation methods to produce detailed land cover maps. Finally, the model only takes in and evaluates static images; thus, it cannot be used to detect "change" over time.

* 1. **Answer to Research Questions**

The light weight MobileNetV2 model proved to be highly effective with an accuracy of 98%, and demonstrated the ability to appropriately classify a wide range of land cover types present in the Sentinel-2 imagery.

The primary performance trade-off was a very fast training time and the associated low compute cost for what is likely a small degradation of accuracy relative to much larger, more complex Anchored models such as Resnet or Vision Transformers. When considering practical applications, the results suggest this is a very acceptable trade-off.

An evaluation of model performance revealed that the primary limitation of the model was distinguishing between sub-categories of land cover types that appear very similar visually. The model confused 'HerbaceousVegetation' with 'PermanentCrop' (both have natural vegetation), and 'Industrial' with 'Residential' (both have developed or anthropogenic features). This suggests there is an issue with somewhat fine labeled classes that appear very similar visually, rather than the more general recognition of features.

* 1. **Future Work**

Future research could build from this research in a few exciting ways. First, the trained model could be used as a feature extractor in a more complex workflow for analyzing larger, un-curated satellite scenes. Second, the efficient MobileNetV2 architecture could become the encoder backbone of a U-Net or other semantic segmentation model to create detailed, pixel-level land cover maps. Finally, the classification framework could also be extended to be a change detection system by comparing the model's predictions on images of the same location but taken at different times, which would provide a direct means of monitoring changes in the environment for purposes like deforestation or urbanization.

**7. Reference list**

Helber, P., Bischke, B., Dengel, A. and Borth, D. (2019). EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, [online] 12(7), pp.2217–2226. doi:https://doi.org/10.1109/JSTARS.2019.2918242.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.-C. (2018). *MobileNetV2: Inverted Residuals and Linear Bottlenecks*. [online] arXiv.org. Available at: https://arxiv.org/abs/1801.04381.