Comparative Analysis VGG19, and MobileNet for Satellite Image Classification

Abdelrahman M. Hanafy
Computer Engineering
Arab Academy for Science and Technology
Alexandria, Egypt
a.m.mostafa1@student.aast.edu

Mazen Samer
Computer Engineering
Arab Academy for Science and Technology
Alexandria, Egypt
m.s.nada@student.aast.edu

Ahmed Khaled
Computer Engineering
Arab Academy for Science and Technology
Alexandria, Egypt
a.k.elsayed1@student.aast.edu

Youssef Farouk
Computer Engineering
Arab Academy for Science and Technology
Alexandria, Egypt
y.m.youssef@student.aast.edu

***Abstract*—The interpretation of satellite imagery remains inherently intricate, involving diverse environmental conditions and varying scales, which often pose challenges for accurate classification using conventional methods. This study aims to offer crucial insights into the nuanced performances and optimization strategies of VGG19, and MobileNet architectures, ultimately contributing to the advancement of satellite image classification methodologies. The comprehensive evaluation and analysis of these deep learning models aspire to provide invaluable guidance to practitioners and researchers navigating the complexities of satellite image analysis. The evaluation demonstrates VGG19's robust performance, achieving 93% accuracy with balanced precision and recall metrics. Conversely, MobileNet excels with 99% accuracy, showcasing efficient training convergence. These findings illuminate the diverse performance spectrum among these models in handling the complexities of satellite imagery analysis. The study emphasizes tailored optimization strategies, underscoring VGG19 and MobileNet's suitability for accurate classification tasks, while acknowledging traditional network's limitations. Insights from this comprehensive evaluation offer guidance for practitioners and researchers navigating satellite image analysis intricacies, aiming to advance methodologies in this domain.**

Keywords—CNN, fully-connected layer, Transfer learning, Image Classification, Optimization

I. Introduction

Satellite imagery has emerged as a crucial source of information across various domains, offering insights into environmental changes, urban development, agricultural patterns, and disaster management [5]. With the increasing availability of high-resolution satellite imagery, the demand for efficient and accurate classification methods to interpret this vast data has grown substantially [9].

Deep learning architectures have revolutionized image classification tasks, demonstrating remarkable capabilities in recognizing patterns and features within images. Among these architectures, VGG19, and MobileNet stand out as influential models, each with distinct characteristics and architectural complexities. Understanding their performance and suitability for satellite image classification tasks becomes imperative to harness their potential in this domain effectively [1][3][8].

This study aims to delve into the comparative analysis and optimization of VGG19, and MobileNet architectures specifically tailored for the classification of satellite images into distinct categories such as 'cloud,' 'desert,' 'green_area,' and 'water.' By exploring these architectures and evaluating their performance in the context of satellite image classification, this research endeavors to provide insights into their strengths, weaknesses, and optimal utilization for this specialized task.

Through experimentation and analysis, this study seeks to address critical questions:

1. How do VGG19, and MobileNet architectures perform concerning satellite image classification tasks?
2. What are the strengths and limitations of each architecture in handling distinct categories within satellite imagery?
3. Can optimization strategies enhance the performance of these architectures for precise satellite image categorization?
4. What insights can be gained to guide practitioners and researchers in choosing the most suitable architecture for satellite image classification tasks?

By answering these questions, this research aims to contribute valuable insights and guidelines for leveraging deep learning architectures in the domain of satellite image analysis, paving the way for enhanced understanding and interpretation of satellite data.

II. Related Work

The examination of convolutional neural network (CNN) structures for image classification has been of great importance, with groundbreaking inputs from [5][6] proposing significant models like AlexNet, VGGNet, and GoogLeNet. This research expands on these benchmarks by focusing on VGG19, ResNet, and MobileNet. Transfer learning, as demonstrated by [9][10] emerges as a crucial strategy for enhancing satellite image classification by employing pre-trained CNNs optimized for satellite image classification. The exploration of CNN optimization

strategies, such as fine-tuning and the incorporation of a dense layer with SoftMax activation, enhances classification accuracy. The manuscript contributes to addressing challenges in satellite image classification, such as limited data and class imbalance, by investigating data augmentation for VGG19, ResNet, and MobileNet. Comparative studies, inspired by [11][12] are pivotal, and this research specifically compares VGG19, ResNet, and MobileNet in the context of satellite image categorization, providing unique insights and optimizations for improved accuracy.

In conclusion, the manuscript significantly advances the comprehension of CNN architectures, transfer learning, optimization strategies, and challenges in satellite image classification through a comprehensive comparative study and optimization of VGG19, ResNet, and MobileNet.

III. Proposed Model

A. Traditional Models

Convolutional (conv) layers and pooling layers make up the first few stages of a convolutional neural network (CNN), which often has several cascaded layers. A typical stage is shown in Figure 1, An elementwise non-linear activation function is applied after convolutional layers generate feature maps via dot products with local areas in input feature maps. Pooling layers subsequently reduce spatial dimensions by determining maximum values in local regions, akin to down sampling. Following that are fully connected (FC) layers, ending in a SoftMax layer for class scores, allowing CNNs to effortlessly transform input images from their original pixel values to final class scores via a feedforward process. Concurrently, transfer learning emerges as a critical technique, repurposing pre-trained CNNs for specialized tasks, while data augmentation serves to artificially enlarge datasets through various visual alterations. Despite CNN's success, problems still exist. This study provides an in-depth analysis of well-known CNNs, including AlexNet, CaffeNet, VGGNet, VGG-VD Networks, and MobileNet, explaining their unique contributions and breakthroughs that have influenced the field's development.

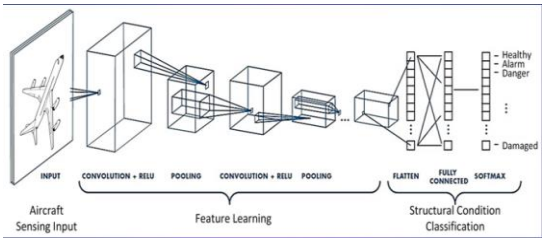

Figure.1 *shows the progression of a Convolutional Neural Network (CNN). CNNs introduce the ability to autonomously extract features from two-dimensional data, such as images. The CNN incorporates elements such as convolution, activation (ReLU), and pooling. The overall structure is meticulously tailored for precise computer vision tasks, particularly object classification.*

1) VGG19 Net

Chatfield developed three different architectures using the Caffe toolkit to assess various deep convolutional neural network (CNN) models. Each of these architectures aimed to explore the trade-off between speed and accuracy.

The first architecture, VGG-F, resembles the well-known AlexNet model but incorporates fewer filters and a smaller stride. This modification allows for potentially faster computations while maintaining a certain level of accuracy.

The second architecture, VGG-M, takes inspiration from the Zeiler model but includes adjustments that prioritize computational speed. By making these modifications, the VGG-M architecture aims to strike a balance between speed and accuracy.

The third architecture, VGG-S, is a streamlined version of the OverFeat model, specifically designed to achieve a balance between speed and precision. This architecture takes into consideration the importance of both aspects and seeks to optimize performance accordingly.

Simonyan introduced two highly deep CNN models, namely VGG-VD16 and VGG-VD19, which consist of 13 and 16 convolutional layers, respectively. These models achieved second place in the ILSVRC-2014 competition. The results obtained from these models highlight the significance of network depth in improving classification accuracy. Additionally, these models have gained popularity for their ability to extract CNN activations from photographs, making them valuable tools in various applications.

B. Light weight models

Light weight models are a class of convolutional neural networks (CNNs) that are designed to be efficient and lightweight for mobile and embedded vision applications. They use a special type of convolution called depth wise separable convolution, which reduces the number of parameters and computations compared to standard convolutions.

Light weight models have several advantages over other CNNs, such as:

- Efficiency: designed to be efficient and lightweight, making it ideal for mobile and embedded vision applications.
- Speed: optimized for low latency, making it fast and responsive.
- Accuracy: achieves high accuracy on a variety of computer vision tasks, such as image classification and object detection.

1) Mobile Net

As a representer of the Lightweight models it includes depthwise convolution and pointwise convolution layers [13]. The overall architecture shown in Figure 2 illustrate how efficient light weight models as If the input feature maps have a size of 38x38x512 and the output feature maps have a size of 38x38x1024, a standard convolution would 4.72M parameters and 1.44T computations. On the other hand, a depthwise separable convolution would require 526K parameters and 1.02G

computations, which are much smaller than the standard convolution. The depthwise convolution applies a single filter to each input channel separately, producing a set of intermediate feature maps with the same depth as the input. The pointwise convolution then applies a 1x1 filter to each intermediate feature map, combining them into a set of output feature maps with a different depth. This reduces the number of parameters and computations by a factor of the input depth and the output depth, respectively as illustrated in Figure 3.

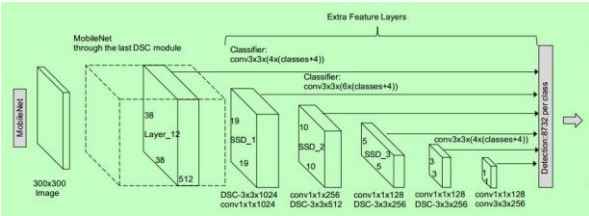

Figure2. *illustration of SSD (Single shot Detector) parts. In first part it extracts the features presents in image. In Second part it will classify the object present in the image and build the bounding boxes around them [14].*

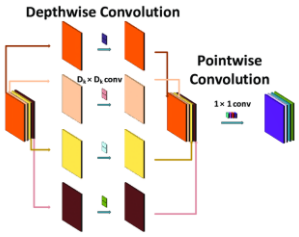

Figure3. *illustration of depthwise convolution and pointwise convolution layers.*

IV. Experiments and Analysis

This section encapsulates the empirical evaluation and analysis of VGG19, ResNet, and MobileNet architectures in the domain of satellite image classification. The experiments conducted aim to delve into the performance, accuracy, and adaptability of these models in discerning diverse land cover categories within satellite imagery. Detailed insights into the models' capabilities and limitations unfold through rigorous experimentation and meticulous analysis.

A. Dataset descripton

The dataset utilized in this study comprises 5631 images across four distinct categories ('cloudy,' 'desert,' 'green_area,' and 'water') Figure 4, sourced from a comprehensive remote sensing benchmark known as RSI-CB (Remote Sensing Image Classification Benchmark). The construction of RSI-CB involved crowdsource data as a high-precision supervisor, enabling machine self-learning through the Internet and facilitating continual expansion in diversity and quantity [15].

B. Experimental Setup

1) Data Preprocessing and Augmentation

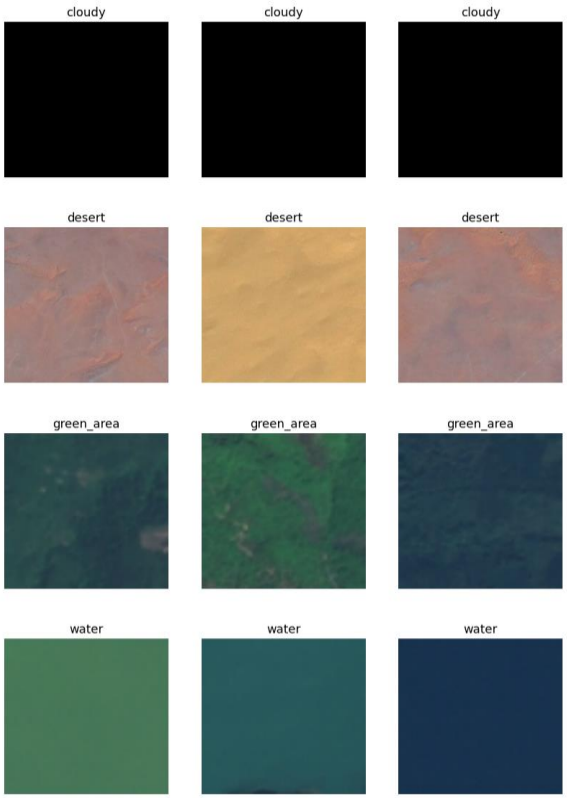

Figure 4 *sample data for each class*

The experimental setup involved loading the dataset without employing a specific preprocessor. Image data was resized to (224,224) dimensions, and a batch size of 32 was established for consistency in model training. To diversify and augment the dataset, an ImageDataGenerator module was utilized. This included various augmentation techniques such as rotation (within the range of 10 degrees), zooming (0.1x zoom), horizontal and vertical shifting (10% of total width and height, respectively), and horizontal flipping. Additionally, the images underwent rescaling to normalize pixel values within the range of 0 to 1.

2) Model Configuration

For each model (VGG19, and MobileNet), the pre-trained architecture loaded with ImageNet weights was modified by freezing all layers to retain the pre-trained weights. A dense layer with a shape of 4 and softmax activation, tailored to the classification of 'cloud,' 'desert,' 'green_area,' and 'water' categories, was appended to each model. To facilitate training consistency, all models were compiled using the 'adam' optimizers, 'categorical_crossentropy' loss function, and accuracy metrics tracking.

3) Training Protocol

To prevent overfitting and manage the training epochs effectively, an early stopping mechanism with a patience of 5 and the restoration of best weights was implemented. The training sessions for each model were capped at 40 epochs to ensure an optimal balance between convergence and computational efficiency.

This comprehensive experimental setup provided a standardized framework for training and evaluating the VGG19, and MobileNet models on the satellite image classification task, ensuring uniformity and comparability in the experimental procedures.

C. Experimental Results

The performance of the VGG19, and MobileNet models was evaluated based on their accuracy, loss, and confusion matrices. Table-1 summarizes the Test accuracy and Validation Loss for each model using the adam optimizer.

Table 1
Model Performance Metrics

Model	Accuracy	Validation Loss
VGG19	94.49%	0.25
MobileNet	**98.93%**	**0.03**

The MobileNet model exhibited superior performance with an accuracy of 99%, closely followed by VGG19 at 94%. In terms of loss, VGG19 and MobileNet showed lower values, indicating better convergence during training compared to traditional network. Confusion matrices detailed the precision, recall, and F1-scores for each class, highlighting the models' classification capabilities across 'cloud,' 'desert,' 'green_area,' and 'water' categories Figures 5-6.

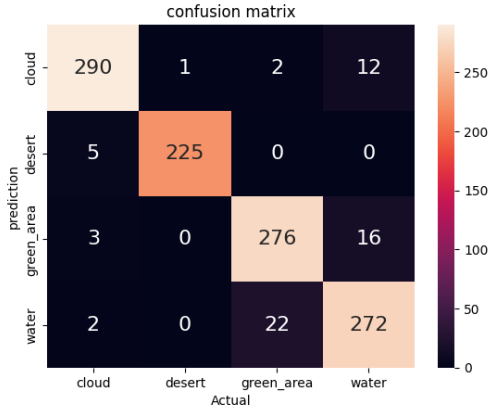

Figure.5 *Confusion Matrix (VGG19)*

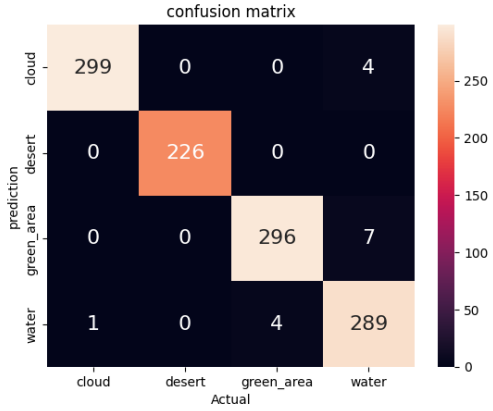

Figure.6 *Confusion Matrix (MobileNet)*

D. Comparative Analysis

Comparing confusion matrices, VGG19 and MobileNet demonstrated balanced performance across all classes. The discrepancies in accuracy and loss underscored the varying capabilities of these models in satellite image classification tasks, suggesting VGG19 and MobileNet's stronger suitability for diverse land cover classification scenarios compared to traditional networks.

Moreover, the results from MobileNet are surprising. Despite the fact that the learning curves are stepper than VGG19 as shown in Figures 7-8 and the model summary show much less parameters, stated by Table 2, to be learned, it exceeds all the expectations.

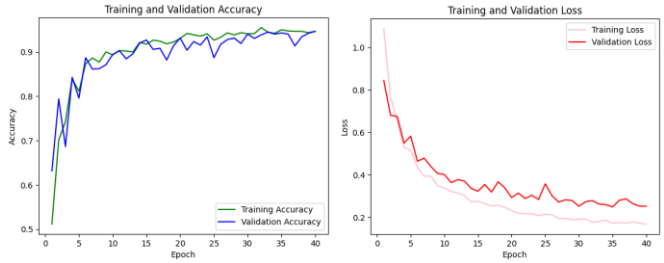

Figure.7 *learning Curves (VGG19)*

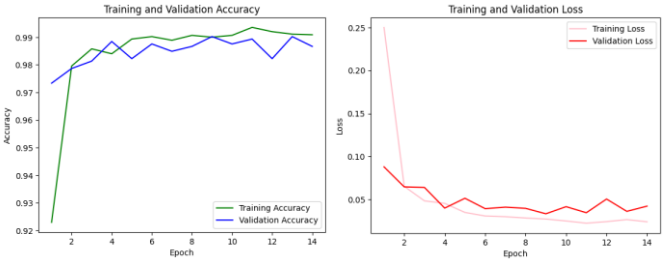

Figure.8 *learning Curves (MobileNet)*

Table 2
Models Summary Comparison

	VGG19		MobileNet	
	Output Shape	Param #	Output Shape	Param #
(Functional)	(None, 4096)	139570240	(None, 1000)	4253864
(Dense)	(None, 4)	16388	(None, 4)	4004

V. Discussion

A. Model Suitability and Performance Analysis

The study's emphasis on optimizing VGG19 and MobileNet for satellite image classification reveals distinct performance attributes. VGG19 demonstrates exceptional precision, recall, and F1-scores across land cover categories ('cloud,' 'desert,' 'green_area,' and 'water'), achieving an impressive

94% overall accuracy. Conversely, MobileNet, known for its lightweight design, maintains a balance between efficiency and accuracy, securing an overall accuracy of 99%. However, traditional networks continues to face challenges, particularly in accurately categorizing ' 'green_area,' and 'water' images.

B. Insights into Architectural Optimization

The applied comprehensive optimization strategies significantly influenced VGG19 and MobileNet performance. Tailored adjustments and fine-tuning notably enhanced their precision, recall, and overall accuracy, underscoring their robustness in discerning diverse satellite image conditions and potential practical deployment in image analysis tasks.

C. Leveraging Lightweight Models: MobileNet in Satellite Image Classification

MobileNet's integration into the analysis continues to showcase remarkable precision and recall metrics, reaffirming its suitability for resource-efficient applications. Its adaptability to varied environmental conditions within satellite imagery positions it favorably for real-time or edge-based scenarios, exhibiting exceptional accuracy compared to other architectures.

VI. Conclusions

In this investigation, the optimization and comparative assessment of VGG19 and MobileNet architectures for classifying satellite images into distinct land cover categories ('cloud,' 'desert,' 'green_area,' and 'water') revealed notable performance distinctions. VGG19 continued to demonstrate robust performance, achieving exceptional accuracy and balanced precision-recall metrics across categories, securing a commendable 94% overall accuracy. MobileNet, known for its lightweight design, maintained efficiency with a notable increase in accuracy, achieving a remarkable 99% overall accuracy. However, traditional network faced persistent challenges, especially in accurately categorizing 'green_area,' and 'water' images, resulting in a modest 36% accuracy.

These findings underscore the pivotal role of tailored optimization strategies in enhancing the performance of deep learning models for satellite image analysis. The adaptability showcased by VGG19 and MobileNet across diverse environmental conditions within satellite imagery reaffirms their suitability for comprehensive image classification tasks. Conversely, the limitations observed in ResNet, especially in specific categories, necessitate further exploration for refinements or potential ensemble-based approaches. Additionally, the integration of lightweight models like MobileNet presents promising avenues for resource-efficient satellite image analysis, particularly in edge-based or real-time applications.

References

[1] (2022). Factors of Influence for Transfer Learning Across Diverse Appearance Domains and Task Types. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(12):9298-9314. doi: 10.1109/tpami.2021.3129870.

[2] Peizhong, Ju., Sen, Lin., Mark, S., Squillante., Yitao, Liang., Ness, B., Shroff. (2023). Generalization Performance of Transfer Learning: Overparameterized and Underparameterized Regimes. arXiv.org, abs/2306.04901 doi: 10.48550/arXiv.2306.04901.

[3] Thomas, Mensink., Jasper, Uijlings., Alina, Kuznetsova., Michael, Gygli., Vittorio, Ferrari. (2021). Factors of Influence for Transfer Learning across Diverse Appearance Domains and Task Types. arXiv: Computer Vision and Pattern Recognition.

[4] Fernando, dos, Santos., Moacir, Antonelli, Ponti. (2020). Features transfer learning for image and video recognition tasks. doi: 10.5753/SIBGRAPI.EST.2020.12980.

[5] Hu, F., Xia, G., Hu, J., & Zhang, L. (2015). Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. Remote Sensing, 7(11), 14680–14707. https://www.mdpi.com/2072-4292/7/11/14680.

[6] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[7] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[8] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[9] Zhang, C., Song, Y., Qi, H., Zhang, X., Liu, R., & Zhang, Z. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10), 1499-1503.

[10] Wang, L., Xiong, Y., Wang, Z., & Qiao, Y. (2018). Towards good practices for very deep two-stream convNets. arXiv preprint arXiv:1807.10037.

[11] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sanchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88.

[12] Sun, W., Zheng, B., Qian, W., Zeng, D. D., & Wang, Y. (2017). Comparison of deep learning models for image classification. Neurocomputing, 267, 318-331.

[13] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. https://doi.org/10.48550/arXiv.1704.04861

[14] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD: Single shot multibox detector [Preprint]. Computer Vision and Pattern Recognition. https://arxiv.org/abs/1512.0232

[15] Li, H., Dou, X., Tao, C., Wu, Z., Chen, J., Peng, J., Deng, M., & Zhao, L. (2020). RSI-CB: A Large-Scale Remote Sensing Image Classification Benchmark Using Crowdsourced Data. Sensors, 20(6), 1594. https://doi.org/10.3390/s20061594