

Article

Land-Cover Classification Using Deep Learning with High-Resolution Remote-Sensing Imagery

Muhammad Fayaz ^{1,†} , Junyoung Nam ^{1,†}, L. Minh Dang ², Hyoung-Kyu Song ²  and Hyeonjoon Moon ^{1,*} 

¹ Department of Computer Science and Engineering, Sejong University, Seoul 05006, Republic of Korea; muhammadfayaz@sju.ac.kr (M.F.); nampang@sju.ac.kr (J.N.)

² Department of Information and Communication Engineering and Convergence Engineering for Intelligent Drone, Sejong University, Seoul 05006, Republic of Korea; minhdl@sejong.ac.kr (L.M.D.); songhk@sejong.ac.kr (H.-K.S.)

* Correspondence: hmoon@sejong.ac.kr

† These authors contributed equally to this work.

Abstract: Land-area classification (LAC) research offers a promising avenue to address the intricacies of urban planning, agricultural zoning, and environmental monitoring, with a specific focus on urban areas and their complex land usage patterns. The potential of LAC research is significantly propelled by advancements in high-resolution satellite imagery and machine learning strategies, particularly the use of convolutional neural networks (CNNs). Accurate LAC is paramount for informed urban development and effective land management. Traditional remote-sensing methods encounter limitations in precisely classifying dynamic and complex urban land areas. Therefore, in this study, we investigated the application of transfer learning with Inception-v3 and DenseNet121 architectures to establish a reliable LAC system for identifying urban land use classes. Leveraging transfer learning with these models provided distinct advantages, as it allows the LAC system to benefit from pre-trained features on large datasets, enhancing model generalization and performance compared to starting from scratch. Transfer learning also facilitates the effective utilization of limited labeled data for fine-tuning, making it a valuable strategy for optimizing model accuracy in complex urban land classification tasks. Moreover, we strategically employ fine-tuned versions of Inception-v3 and DenseNet121 networks, emphasizing the transformative impact of these architectures. The fine-tuning process enables the model to leverage pre-existing knowledge from extensive datasets, enhancing its adaptability to the intricacies of LC classification. By aligning with these advanced techniques, our research not only contributes to the evolution of remote-sensing methodologies but also underscores the paramount importance of incorporating cutting-edge methodologies, such as fine-tuning and the use of specific network architectures, in the continual enhancement of LC classification systems. Through experiments conducted on the UC-Merced_LandUse dataset, we demonstrate the effectiveness of our approach, achieving remarkable results, including 92% *accuracy*, 93% *recall*, 92% *precision*, and a 92% *F1-score*. Moreover, employing heatmap analysis further elucidates the decision-making process of the models, providing insights into the classification mechanism. The successful application of CNNs in LAC, coupled with heatmap analysis, opens promising avenues for enhanced urban planning, agricultural zoning, and environmental monitoring through more accurate and automated land-area classification.



Citation: Fayaz, M.; Nam, J.; Dang, L.M.; Song, H.-K.; Moon, H. Land-Cover Classification Using Deep Learning with High-Resolution Remote-Sensing Imagery. *Appl. Sci.* **2024**, *14*, 1844. <https://doi.org/10.3390/app14051844>

Academic Editor: João M. F. Rodrigues

Received: 18 January 2024

Revised: 11 February 2024

Accepted: 21 February 2024

Published: 23 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: deep learning; transfer learning; land-area classification; CNN

1. Introduction

The utilization of remote-sensing (RS) imagery for land-cover (LC) classification is of paramount importance across various domains, encompassing environmental protection, agriculture and urban planning, and land resource management [1]. Recent accessibility to high-resolution remote-sensing (HRRS) images and the ability to gather multi-temporal

and multi-source RS images from diverse geographic regions [2] present new opportunities for multiple-time-scale LC classification. Nevertheless, the complex features visible in HRRS images, such as geometrical and object structures, introduce new challenges for classification [3]. Variability in photographic distortions, scale variations, and illumination changes in RS images make it difficult to apply existing models to diverse HRRS images effectively [4]. Furthermore, spectral and spectral–spatial (SS) features have traditionally been employed in the literature to interpret RS images for land-cover classification [5]. However, these features struggle to capture the contextual information in HRRS images due to increased spatial resolution [6]. To address these challenges, deep CNNs have gained attention for their ability to comprehend HRRS images [7] and represent semantic and high-level image properties [8,9], such as scene classification [10], object detection [11], image retrieval [12], as well as LC classification [13]. However, two significant issues arise when applying deep models to LC classification with multi-source HRRS data: insufficient transferability and the lack of an extensive, well-annotated land-cover dataset. The challenges in applying land-cover classification with multi-source images include issues of insufficient transferability, where models struggle to adapt effectively to diverse dataset images, and the absence of extensive well-annotated land-cover dataset images.

To address these issues, therefore, this study introduces an efficient and robust LAC system that harnesses fine-tuned versions of Inception-v3 and DenseNet networks, leveraging transfer learning. The primary objectives include enhancing LC classification performance through the contextual understanding of deep CNNs and conducting a comparative analysis to assess the system's advancements and competitive edge over existing models in the field of satellite sensing imagery interpretation. This approach addresses the limitations posed by the diversity and complexity of HRRS images, ultimately contributing to improved land-cover classification in remote-sensing applications. Various CNN models, including Inception-v [14], DenseNet201 [8], and ResNet-50 [15], have been trained on a labeled dataset.

Recent research highlights innovative applications of deep learning, especially (CNNs) in the unmanned aerial vehicle (UAV) domain. For instance, Chen, Wang, and Zhang's work [16] combines CNNs and cooperative spectrum detection for unauthorized drone identification. Haq et al. [17] employ a stacked auto-encoder deep learning approach to achieve accurate forest area assessment using UAV-captured images applicable to forest management. Kawaguchi, Nakamura, and Hadama et al.'s research [18] leverages CNNs to identify diverse drone types, achieving over 90% accuracy in recognition and showcasing the prowess of CNNs in drone identification across various models and shapes, including radio-controlled flying objects [19].

Recent studies have showcased the diverse applications of deep learning in the fields of image classification and object detection. Chehreh, Moutinho, and Viegas [20] introduced a classification method for rotary-wing unmanned aerial vehicles and birds, enhancing image classification accuracy notably with the CDNTS layer. Youme and colleagues [21] presented an automated approach for detecting hidden waste dumps in Senegal, employing single-shot detector techniques for feature extraction, though facing challenges in regions with imprecise ground truths. Genze and team [22] explored deep learning's generalization capabilities for early weed detection in sorghum fields, creating expert-curated datasets and achieving strong model generalization with an F1-score exceeding 89% on testing data, even in challenging conditions, including degraded captures with motion blur and occluded plants, surpassing existing research.

These research studies highlight the application of deep learning and machine learning techniques for various tasks in the context of unmanned aerial vehicles (UAVs) and remote sensing. Shahi and colleagues [23] focus on the identification of crop diseases using UAV-based remote sensing, emphasizing the role of image processing methods, assessing the effectiveness of ML and DL techniques, and exploring future research directions in UAV-based crop disease detection and classification [24,25]. Behera, Bakshi, and Sa [26,27] present lightweight CNN architectures for real-time segmentation and object extraction

on IoT edge devices, achieving high performance on datasets suitable for urban and agricultural mapping, as well as road damage detection using YOLO algorithms. Shanthy and team [28] discuss the use of face identification algorithms in drones for security, identity verification, disaster relief, and more, aiming to aid technologists in developing hybrid algorithms for real-time face recognition across diverse scenarios. These studies demonstrate the versatility of deep learning and UAV-based applications in diverse domains.

Aydin and Singha [29] presented YOLOv5, a one-shot detector trained with augmented data and pre-trained weights, achieving a 90.40% mean average precision. This represents a significant 21.57% improvement over the previous YOLOv4 model when tested on the same dataset. Yao et al. [30] investigated split learning in IoD networks via simulations. Findings reveal that separation levels have minimal impact on accuracy, increasing client-side layers extends training time, communication overhead is a major bottleneck, client numbers insignificantly affect accuracy, and training time slightly rises with more clients.

The methods described in the literature offer advancements in deep learning for drone image classification, but they are subject to various limitations. These include challenges related to data diversity, labeling accuracy, real-time processing, adverse conditions, privacy, benchmarking against existing methods, real-world deployment, and model interpretability. To mitigate these limitations, we reduced bias and domain gap through fine-tuning and domain adaptation techniques, considering a more diverse and representative target dataset, and developing model interpretability tools specific to the transfer learning context. The main findings of this study are as follows:

- The primary objective of this study is the development of a highly efficient and reliable land-cover (LC) classification system. To achieve this, we strategically employ fine-tuned versions of Inception-v3 and DenseNet networks, emphasizing the transformative impact of transfer learning. The fine-tuning process enables the model to leverage pre-existing knowledge from extensive datasets, enhancing its adaptability to the intricacies of LC classification. This approach increases both the efficiency and reliability of the model, underscoring the vital role of transfer learning methodologies in optimizing the performance of our LC classification system. By aligning with these advanced techniques, our research not only contributes to the evolution of remote-sensing methodologies but also underscores the paramount importance of incorporating cutting-edge methodologies, such as fine-tuning, in the continual enhancement of LC classification systems.
- By enhancing classification accuracy through the precise characterization of contextual information in high-resolution remote-sensing (HRRS) images, we demonstrate the significance of leveraging the powerful capabilities inherent in deep convolutional neural networks (CNNs). The fine-tuned networks play a crucial role in capturing intricate features, contributing to the system's heightened accuracy and reliability in LC classification.
- A detailed comparative analysis is conducted to evaluate our land-cover classification system against state-of-the-art models within the field of satellite sensing imagery interpretation. This comprehensive assessment scrutinizes the system's performance, accuracy, and efficiency in comparison to the most advanced solutions, thereby highlighting its significant advancements and competitive edge within the domain of land-cover classification. Notably, the fine-tuned models achieved baseline accuracy comparatively.

The structure of the remaining sections of this work is as follows. The land-area classification system's comprehensive methodology is explained in Section 3. Section 4 reports experimental results and comparative analysis with baselines, and finally, Section 5 concludes the article.

2. Proposed Methodology

The proposed methodology for land-cover classification is discussed in this section, which mainly includes data preprocessing and model training and evaluations, as shown

in Figure 1. In this study, we use different deep learning models for fine-tuning and transfer learning, optimizing critical hyper-parameters like batch size, activation function, and learning rate. The utilization of optimizers assists in adjusting the learning rate within the neural models to enhance performance. All these steps are briefly explained in the subsequent sections.

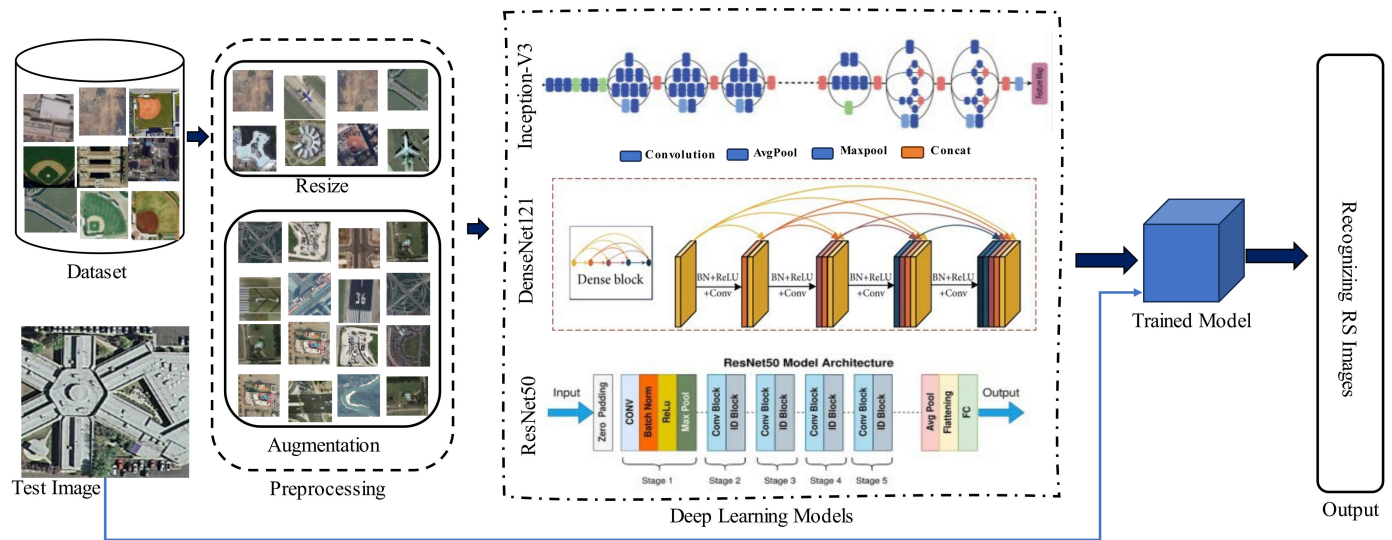


Figure 1. The high-level diagram of the proposed work for LAC.

2.1. Preprocessing

In the initial phase of the research study, data is collected, typically captured with mobile device cameras or other vision sensors that record information crucial to influencing the output of subsequent deep learning (DL) and machine learning (ML) models. Following this data collection, the preprocessing stage becomes paramount, involving various techniques such as resizing, noise reduction, and data augmentation to enhance the quality of images or videos. As a foundational step in our investigation, we employ data normalization, a process illustrated mathematically in Equation (1), where each pixel value is standardized to the range of $[-1, 1]$.

$$R_i = \frac{I - I_{min}}{I_{max} - I_{min}} \quad (1)$$

In this context, the representation of the initial data is denoted as I , where I_{min} and I_{max} represent the minimum and maximum values within the input data, respectively. To standardize the data input, " R_i " is introduced. Before the training, all images underwent resizing to dimensions of $224 \times 224 \times 3$. Additionally, various data transformations were applied to the input images, encompassing operations such as flipping, 15° rotation, and 0.2 zooming. Introducing data augmentation, as depicted in Figure 2, played a pivotal role in enhancing dataset variability, effectively mitigating overfitting and bolstering the model's accuracy and generalization capabilities.

2.2. Convolutional Neural Networks

The convolutional neural network (CNN) was introduced by [31] in the late 1980s and contains three main key layers: convolution, pooling, and fully connected. Convolutional layers are used to extract features from input data by placing the kernel and the input image through a convolutional process to extract features. The feature map is created by sliding a small $5 \times 5 \times 3$ rectangular matrices over the image, which is known as the kernel. The pooling layer is used to condense the feature size, which is divided into average, maximum, and minimum pooling. In the field of computer vision (CV), CNN has recently achieved

higher performance in comparison to conventional machine learning methods. Several CNN-based architectures are developed for image classification, and each has its own types of pros and cons. For instance, some of these models are effective but computationally expensive, while some are lightweight and resource friendly. After a thorough analysis of recent models, we choose Inception-v3 and DenseNet121 for LAC. These models have a balance between effectiveness and computational efficiency, which makes them suitable for accurate and concise land-cover classification. The proposed diagram of fully connected layers is given in Figure 3.

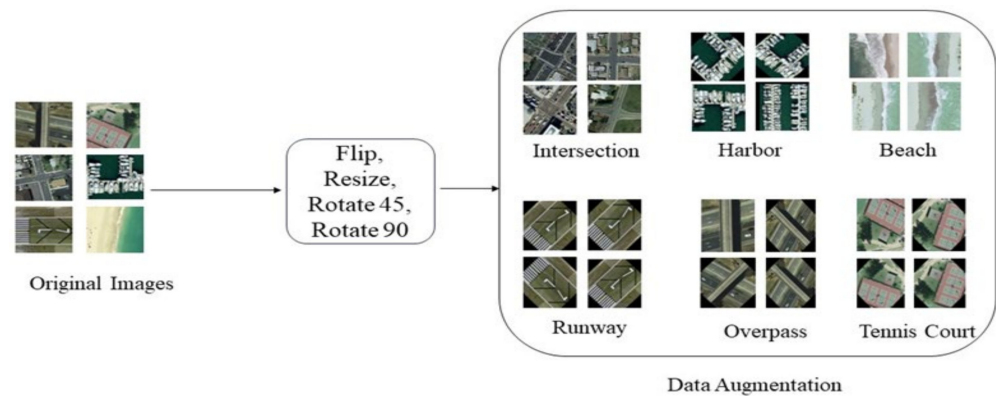


Figure 2. Data augmentation techniques, such as flipping, rotation, zooming, and cropping, employed to augment the training dataset and improve the model performance.

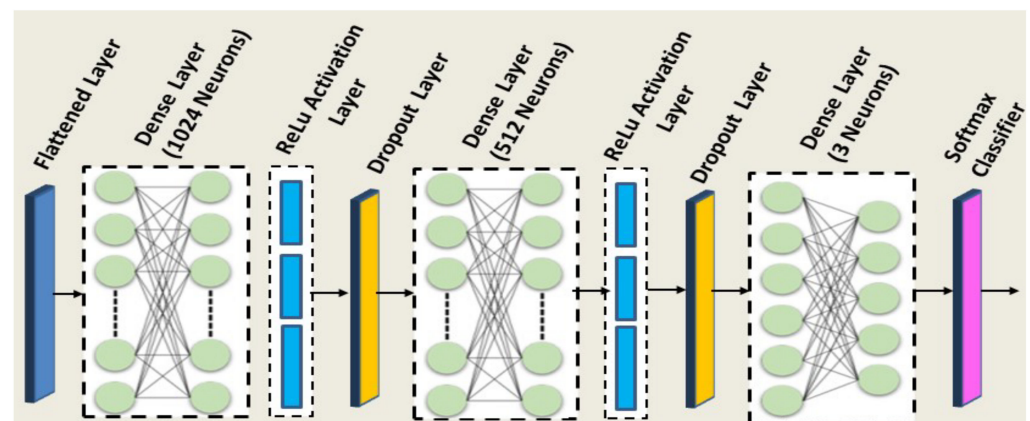


Figure 3. Proposed diagram of fully connected layers.

2.3. Inception-V3 Model

An upgraded variety of the Inception-v1 [31] model is called the Inception-v3 model [31]. For greater model adaptability, the network is optimized in several ways by the Inception-v3 model. Its network is larger than that of the Inception-v1 and v2 models. It is a low-configuration computer that was used to train a deep CNN model. Training can be difficult and time-consuming; it may even take several days at times. This issue is resolved by transfer learning, which preserves the model's last layer for application to new categories. The 48 layers of the Inception-V3 model comprise newly added layers based on our dataset classes, and all the model's upper layers are frozen, as shown in Figure 4.

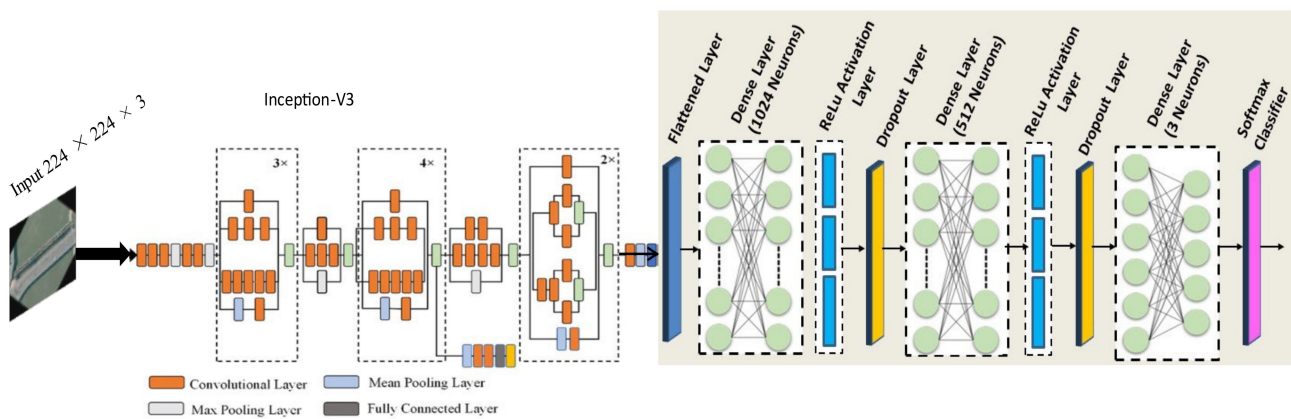


Figure 4. Architecture of Inception-V3 model with proposed fully connected layers.

Moreover, complex DL problems benefit from the use of a model such as Inception-V3. Other Deep learning models employ a basic stack of convolutional layers, pooling layers, and fully connected layers, like Vgg16, Vgg19, and AlexNet, among others. The Inception-V3 models apply 1×1 convolutions, also referred to as point-wise convolutions, and then apply convolutional layers with varying kernel sizes simultaneously. This results in an increased number of hidden layers. This means that Inception-V3 models perform better for complex tasks or extracting features. For more complicated issues, Inception is utilized because it enables it to learn more complex features.

2.4. DenseNet121 Model

A deep learning architecture called DenseNet121 [32] was revealed in 2017 for use in computer vision and image categorization applications. It is distinguished by its deep connectivity, in which every layer has a direct connection to every layer that comes before it and after it, encouraging feature reuse and gradient flow across the network. It employs bottleneck and transition layers to regulate the number of parameters and geographical dimensions to manage model complexity. This architecture, shown in Figure 5, has gained popularity in the computer vision field due to its efficiency and leading-edge performance, which address important issues with deep neural networks and produce amazing results in image classification applications.

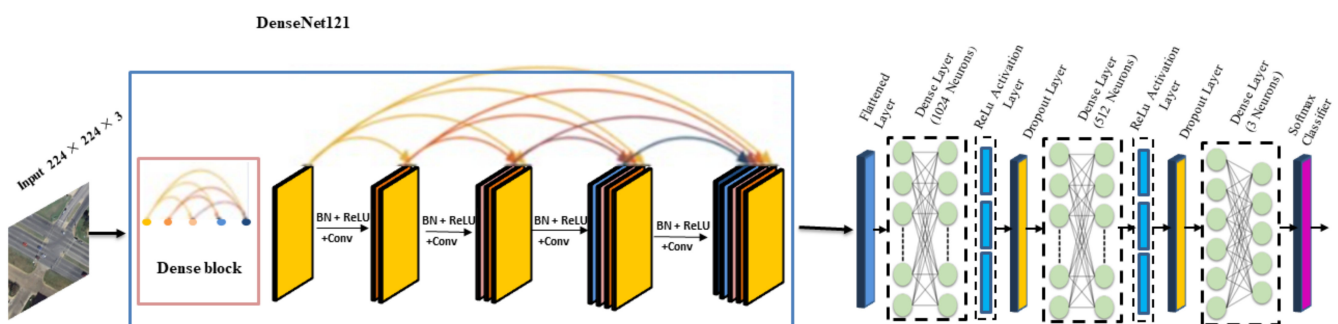


Figure 5. The architecture of DenseNet121 model with proposed fully connected layers.

2.5. ResNet-50 Model

A deep neural network architecture called ResNet-50 [33], or residual network, was created to overcome the difficulties involved in training extremely deep networks; the architecture is shown in Figure 6. Since its introduction in a 2015 study, it has developed into a key model in the domains of computer vision and deep learning. ResNet-50 uses residual blocks with skip connections to achieve both performance and depth. The vanishing gradient issue, which frequently impedes training in deep networks, is lessened by these connections, which allow the gradient to move through the network more efficiently.

ResNet-50 models are, therefore, capable of becoming extremely deep, with hundreds of layers, which enhances their accuracy and efficiency for many uses for computer vision, such as image classification, object identification, and segmentation.

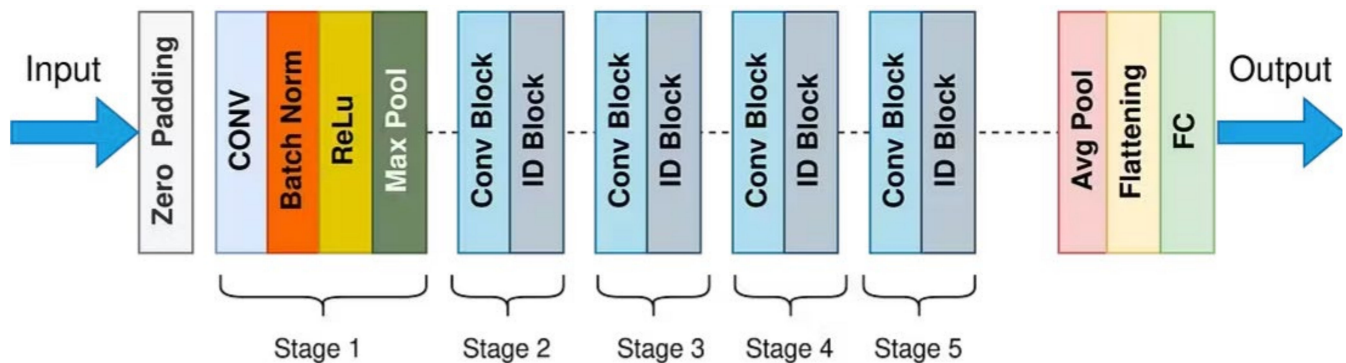


Figure 6. The architecture of ResNet-50 model with proposed fully connected layers.

ResNet-50 topologies are available at different depths, with the total number of layers indicated by numbers such as ResNet-18, ResNet-50, and so forth. Because of their novel approach to residual connections, which has been widely adopted and expanded to other domains outside image recognition, these models have had a significant influence on the evolution of deep neural networks. The success of deep learning has been greatly aided by the ResNet-50 concept, which has made it possible to create deep and highly accurate models that remain at the front of ML and AI research.

2.6. Fine-Tuning and Transfer Learning

This section outlines the steps involved in training and refining our models. Initially, pre-trained weights from the extensive ImageNet dataset, comprising 14 million images categorized into a thousand classes [34], are utilized. The Keras library facilitates the importation of these weights, accelerating convergence through the incorporation of previously learned features and improving image recognition performance. The transfer learning approach leverages the benefits of ImageNet weights, specifically tailored for image classification. This method expedites tasks by requiring less effort compared to the utilization of randomly initialized weights. Fine-tuning is then performed on the UCMerced_LandUse dataset images, focusing on adjusting the final layers of the base model while freezing all other layers to retain the initial UCMerced_LandUse dataset weights. This strategic approach optimizes training by preserving the valuable insights gained from the pre-trained ImageNet weights in the initial layers. During this process, all other layers are frozen to maintain the weights obtained from the initial training on the UCMerced_LandUse dataset [35]. Upon combining the proposed layers and classifier, the entire network is released once the final layers have been trained on the UCMerced_LandUse dataset. Evaluation of the model accuracy is then conducted using test data, incorporating weights from both the ImageNet dataset and the UCMerced_LandUse images. All DL models undergo fine-tuning with diverse hyper-parameters outlined in Table 1. Image input dimensions are $224 \times 224 \times 3$, utilizing a batch size of 32. We use an SGD optimizer and employ categorical cross-entropy (CC) as the loss function. The SoftMax activation function is applied to the output layer, ensuring the robustness and effectiveness of the land-cover classification model.

Table 1. Proposed model hyper-parameters.

Performance Measure	Inception-v3	DenseNet-121	ResNet-50
Image dimensions	224×224	224×224	224×224
Optimizers	SGD	SGD	SGD
Batch size	32	32	32
No. of epochs	50	50	50
Loss	CC	CC	CC
Activation function	SoftMax	SoftMax	SoftMax

3. Experimental Results and Discussion

3.1. Dataset

The “UCMerced_LandUse Dataset” is widely used in the field of computer vision and machine learning for land use classification tasks. This dataset was created by the University of California, Merced, and it consists of high-resolution aerial images of various land use and LC classes, as given in Table 2. It is commonly used for tasks such as image classification, object recognition, and segmentation. Researchers and practitioners often use the UCMerced_LandUse Dataset for training and evaluating ML algorithms, especially for image classification tasks related to land use and LC mapping. It provides a valuable resource for testing the effectiveness of different models and techniques in remote-sensing and computer vision applications.

Table 2. Detail distribution of dataset.

S. No	Class	No. of Images	Train Images	Test Images
1	Agricultural	1000	800	200
2	Airplane	1000	800	200
3	Baseball diamond	1000	800	200
4	Beach	1000	800	200
5	Buildings	1000	800	200
6	Chaparral	1000	800	200
7	Forest	1000	800	200
8	Freeway	1000	800	200
9	Golf course	1000	800	200
10	Harbor	1000	800	200
11	Intersection	1000	800	200
12	Mobile home park	1000	800	200
13	Overpass	1000	800	200
14	Parking lot	1000	800	200
15	River	1000	800	200
16	Runway	1000	800	200
17	Storage tanks	1000	800	200
18	Tennis court	1000	800	200
Total Images		18,000	14,400	3600

3.2. Evaluation Parameters

The evaluation parameters used in this study are *F1-score*, *accuracy*, *precision*, and *recall*. In comparison to all observations, the ratio of correctly predicted observations is known as accuracy. Accuracy is assessed using Equation (2), wherein the terms false positive (*FP*), false negative (*FN*), true positive (*TP*), and true negative (*TN*) are utilized.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Precision is the ratio of correctly predicted positive observations to all predicted positive observations, as shown in Equation (3).

$$Precision = \frac{TP}{TP + FN} \quad (3)$$

The ratio of all actual class observations to all correctly predicted positive observations is known as recall, as shown in Equation (4).

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

F1-score is the precision and recall weighted average. It is computed using Equation (5).

$$F1 - score = 2 \cdot \frac{(Precision \times Recall)}{Precision + Recall} \quad (5)$$

The hyper-parameters configuration is used to fine-tune both CNN networks. All the models are trained with a 0.0001 learning rate and 32-batch size using the Adam optimizer. Using the UCMerced_LandUse dataset, we fine-tune the inception-V3, DenseNet121, and ResNet-50 models by freezing all top layers and adding new layers according to the number of classes in the dataset classes. Each model is trained by using Python 3.8, TensorFlow 2.13.0, sklearn 1.0, matplotlib 3.7.3, and NumPy 1.24.3 libraries and an Intel(R) Core (TM) i7-8700 CPU with RAM of 16 GB, Seoul, South Korea.

3.3. Experimental Results

Evaluations and experiments are conducted on three pre-trained models: DenseNet121, Inception-v3, and ResNet-50. Assessment metrics encompass F1-score, recall, accuracy, and precision. The training process involves fifty epochs for each of the Inception-v3, DenseNet121, and ResNet-50 models. Upon completion of the training, the testing accuracies for the three models are as follows: 95% for Inception-v3, 94% for DenseNet121, and 93% for ResNet-50. Additionally, the validation accuracies are reported as 92%, 91%, and 91% for Inception-v3, DenseNet121, and ResNet-50, respectively. Table 3 shows the precision, recall and F1-Score for each model. These results provide insights into the model performance, indicating their ability to generalize well to unseen data. Furthermore, Figure 6 displays the model's static visual results. Actual labels are displayed in the first row, followed by the predicted labels of the Inception-v3 model in the second row, the DenseNet121 model predicted labels in the third row, and the ResNet-50 model predicted labels in the last row. The correct predicted labels are displayed as black, while the incorrect predicted labels are given as red as shown in Figure 7. We can see from Table 2 that Inception-V3 performs better than DenseNet121 and ResNet-50. Additionally, it demonstrates how the DenseNet121 and ResNet-50 models misunderstand certain labels, such as "river", "runway", and "airplane". Additionally, the confusion matrices and heatmaps for the models and classes are displayed in Figures 8 and 9, respectively.

Table 3. Average *precision*, *recall*, and *F1-score* of the three models.

S. No	Model	Precision %	Recall %	F1-Score
1	Inception-V3	93	92	92
2	DenseNet201	91	90	89
3	ResNet-50	91	90	88

Static visual results of all models

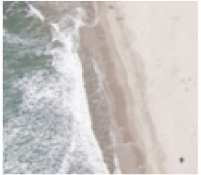
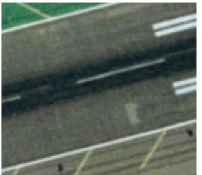
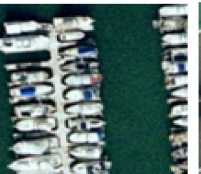
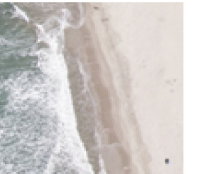
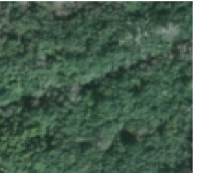
					
Actual Labels	Beach	Harbor	River	Runway	Airplane
Inception-V3	Beach	Harbor	River	Runway	Airplane
DenseNet121	Beach	Harbor	Beach	Runway	Airplane
ResNet-50	Beach	Harbor	River	Runway	Harbor
					
Actual Labels	Airplane	River	Harbor	Runway	Beach
Inception-V3	Airplane	River	Harbor	Runway	Beach
DenseNet121	Airplane	River	River	Runway	Beach
ResNet-50	Harbor	River	Harbor	Runway	River
					
Actual Labels	freeway	parking lot	Forest	Tennis Court	Overpass
Inception-V3	freeway	parking lot	Forest	Tennis Court	Overpass
DenseNet121	freeway	parking lot	Forest	Freeway	Overpass
ResNet-50	freeway	freeway	Forest	Tennis Court	parking lot

Figure 7. Visual results of the three models. Red color are used to represent the incorrect prediction where models are confused but inception model is not confused and give us correct prediction of all classes.

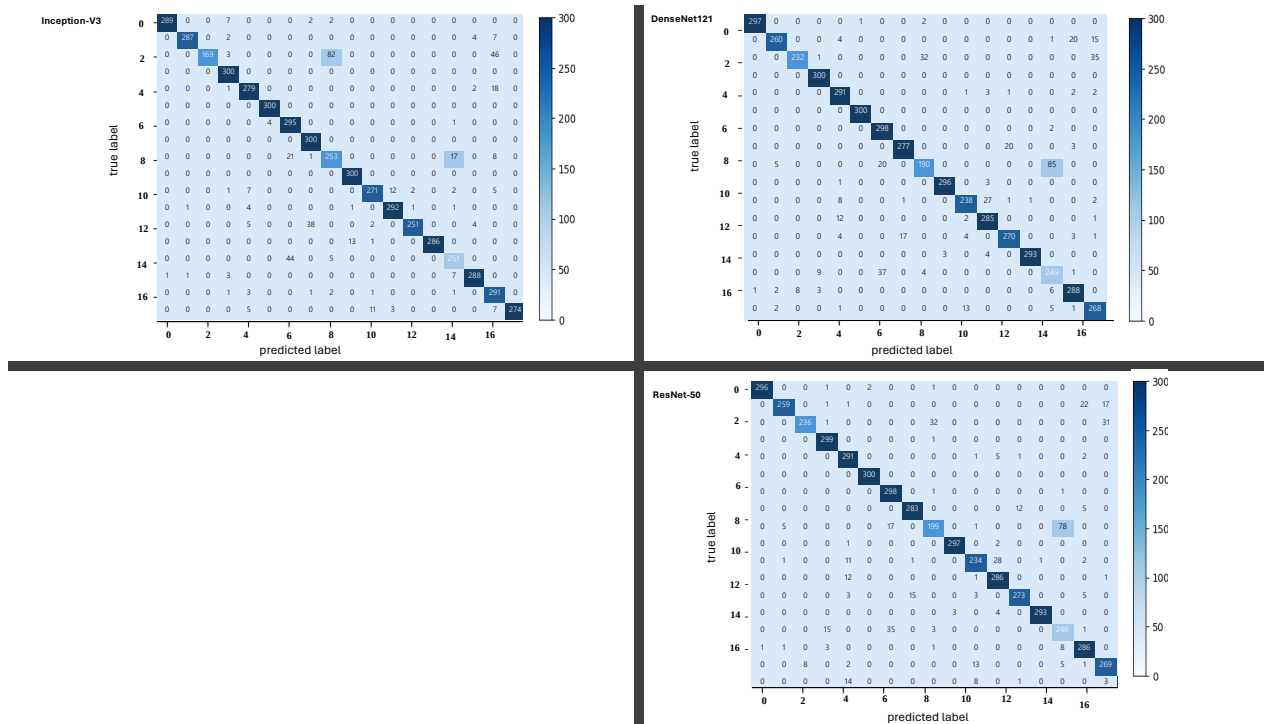


Figure 8. Confusion matrices of all three models: Inception-V3, DenseNet121, and ResNet-50.

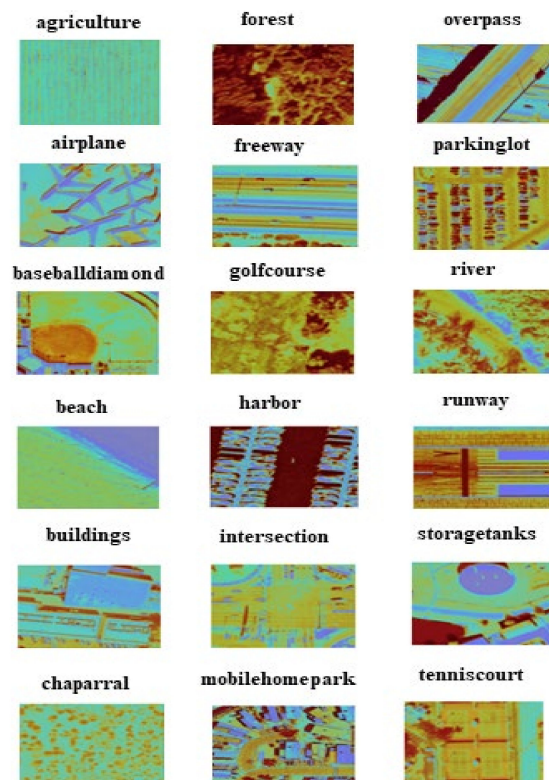


Figure 9. Heatmap analysis of all classes.

Finally, Table 2 displays the *F1-score*, *precision*, and *recall* for all Inception-v3, DenseNet121, and ResNet-50 models in the land areas. This illustrates that DenseNet121 and ResNet-50 are not as effective as Inception-V3 because some classes' accuracies are lower comparatively.

3.4. Comparison with Contemporary Techniques

The comparative analysis presented in Table 4 assesses the performance of the proposed model against modern techniques, employing various deep learning models. The first model, EfficientNet [36], demonstrates an accuracy of 83%, a precision of 83%, a recall of 71%, and an F1-score of 77%. Moving on to the second model, CNN [37] achieved an accuracy of 89%, a precision of 87%, a recall of 84%, and an F1-score of 85%. This model outperforms EfficientNet in terms of accuracy and F1-score. The third model, ResNet34 [38], displays an accuracy of 70%, a precision of 70%, a recall of 70%, and an F1-score of 71%. While this model achieves a balanced performance, it falls behind in accuracy compared to the other models. However, the Inception-V3 model surpasses all others with an accuracy of 92%, a precision of 93%, a recall of 92%, and an F1-score of 92%. This indicates superior performance compared to the aforementioned models. The DenseNet121 model in the proposed system achieves an accuracy of 91%, a precision of 91%, a recall of 90%, and an F1-score of 89%. Although slightly lower than Inception-V3, it outperforms EfficientNet, CNN, and ResNet34. Lastly, the ResNet-50 model in the proposed system yields an accuracy of 91%, a precision of 90% [31], a recall of 90%, and an F1-score of 88%. It maintains competitive performance, particularly in accuracy and F1-score, when compared to other models. In short, the proposed Inception-V3 model stands out as an effective model in terms of accuracy and F1-score, showcasing the efficacy of the suggested system in surpassing the state-of-the-art techniques.

Table 4. Comparative analysis of the proposed system with SOTA methods.

Test Models	Total No. of Classes	Model Accuracy %	Precision %	Recall %	F1-Score %
EfficientNet [34]	-	83	83	71	77
CNN [35]	04	89	87	84	85
ResNet34 [36]	10	70	70	70	71
Proposed Inception-V3	18	92	93	92	92
Proposed DenseNet121	18	91	91	90	89
Proposed ResNet-50	18	91	90	90	88

4. Ablation Study

In a comparison of the AID (Aerial Image Dataset) and UC Merced datasets, UC Merced comes out on top for certain applications. UC Merced provides a good balance of manageable dataset size and high-resolution aerial images, making it suitable for tasks like land-use classification. The dataset's moderate scale allows for efficient model training and testing without compromising scene diversity. UC Merced provides a comprehensive representation for training robust models by covering a wide range of land-use categories. The higher resolution of the dataset aids in the detailed analysis of aerial imagery, which is useful for tasks requiring precision in land-area classification. While both the AID and UC Merced datasets are useful in the field of aerial image analysis, the UC Merced dataset's balance of dataset size, diversity, and resolution makes it a better choice for applications that require both efficiency and detailed scene understanding. Herein, additional experiments are performed with different models to fully assess their generalization abilities. The detailed result of each model is given in Table 5.

Table 5. Comparative analysis of datasets.

Dataset	UC Merced	AID
No. of Classes	18	30
No. of images	18,000	9000
Inception V3 Acc	92%	86%
DenseNet121	91%	89%
Res-Net50	91%	85%

5. Conclusions

In this study, we introduced a robust approach to land-area scene classification (LASC) through the implementation of three distinct deep learning (DL) models—Inception-v3, DenseNet121, and ResNet-50—leveraging the power of transfer learning. Our LASC system demonstrates impressive accuracy rates of 92%, 91%, and 91% in classifying both static and real-time land-area scenes from images. While these results show promise, there remains potential for further enhancement. Our system has the capacity to evolve, extending its capabilities to classify multiple land-area scene categories in real time. Such advancements could significantly contribute to improved urban planning and land management across both urban and rural areas, benefiting a diverse range of stakeholders. Upon analyzing our models' performance metrics, including *recall*, *accuracy*, *precision*, and *F1-score*, Inception-v3 emerges as the top performer, outclassing DenseNet121 and ResNet-50. Inception-v3 excels in accurately classifying land-area scene poses in images, achieving remarkable scores of 92% *accuracy*, 93% *precision*, 92% *recall*, and 92% *F1-score*. DenseNet121 follows closely with scores of 91%, 91%, 90%, and 89%, while the ResNet-50 model achieves scores of 91%, 91%, 90%, and 88%, respectively.

Our future endeavors with the LASC system will focus on the development of robotics and smartphone applications. Additionally, we aim to enhance the system's accuracy by creating an improved CNN incorporating advanced data fusion techniques. This strategic approach is expected to result in even more precise land-area scene classification, ultimately facilitating more informed decision-making in the realms of land management and urban planning. The widespread applications of these advancements underscore the potential benefits for various stakeholders in diverse domains.

Author Contributions: Conceptualization, M.F. and J.N.; methodology, M.F. and J.N.; software, M.F. and J.N.; validation, M.F., J.N. and L.M.D.; formal analysis, L.M.D.; investigation, H.M.; resources, H.M.; data curation M.F. and J.N.; writing—original draft preparation, M.F. and J.N.; writing—review and editing, L.M.D.; visualization, M.F., J.N. and H.-K.S.; supervision, H.-K.S.; project administration, H.M.; funding acquisition, H.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A6A1A03038540); the Institute of Information and Communications Technology Planning and Evaluation (IITP) under the Metaverse Support Program to Nurture the Best Talents (IITP-2023-RS-2023-00254529) grant funded by the Korean government (MSIT); and the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) through the Digital Breeding Transformation Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (322063-03-1-SB010).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ansith, S.; Bini, A. Land use classification of high resolution remote sensing images using an encoder based modified GAN architecture. *Displays* **2022**, *74*, 102229.
2. Amin, S.U.; Hussain, A.; Kim, B.; Seo, S. Deep learning based active learning technique for data annotation and improve the overall performance of classification models. *Expert Syst. Appl.* **2023**, *228*, 120391. [\[CrossRef\]](#)
3. Amin, S.U.; Taj, S.; Hussain, A.; Seo, S. An automated chest X-ray analysis for COVID-19, tuberculosis, and pneumonia employing ensemble learning approach. *Biomed. Signal Process. Control* **2024**, *87*, 105408. [\[CrossRef\]](#)
4. Hussain, A.; Imad, M.; Khan, A.; Ullah, B. Multi-class classification for the identification of COVID-19 in X-ray images using customized efficient neural network. In *AI and IoT for Sustainable Development in Emerging Countries: Challenges and Opportunities*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 473–486.
5. Alem, A.; Kumar, S. Transfer learning models for land cover and land use classification in remote sensing image. *Appl. Artif. Intell.* **2022**, *36*, 2014192. [\[CrossRef\]](#)
6. Alrayes, F.S.; Alotaibi, S.S.; Alissa, K.A.; Maashi, M.; Alhogail, A.; Alotaibi, N.; Mohsen, H.; Motwakel, A. Artificial intelligence-based secure communication and classification for drone-enabled emergency monitoring systems. *Drones* **2022**, *6*, 222. [\[CrossRef\]](#)
7. Tong, X.-Y.; Xia, G.-S.; Lu, Q.; Shen, H.; Li, S.; You, S.; Zhang, L. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sens. Environ.* **2020**, *237*, 111322. [\[CrossRef\]](#)
8. Zhao, X.; Zhang, J.; Tian, J.; Zhuo, L.; Zhang, J. Residual dense network based on channel-spatial attention for the scene classification of a high-resolution remote sensing image. *Remote Sens.* **2020**, *12*, 1887. [\[CrossRef\]](#)
9. Xie, J.; Fang, L.; Zhang, B.; Chanussot, J.; Li, S. Super resolution guided deep network for land cover classification from remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5611812. [\[CrossRef\]](#)
10. Giang, T.L.; Dang, K.B.; Le, Q.T.; Nguyen, V.G.; Tong, S.S.; Pham, V.-M. U-Net convolutional networks for mining land cover classification based on high-resolution UAV imagery. *IEEE Access* **2020**, *8*, 186257–186273. [\[CrossRef\]](#)
11. Digra, M.; Dhir, R.; Sharma, N. Land use land cover classification of remote sensing images based on the deep learning approaches: A statistical analysis and review. *Arab. J. Geosci.* **2022**, *15*, 1003. [\[CrossRef\]](#)
12. Sravya, N.; Lal, S.; Nalini, J.; Reddy, C.S.; Dell’Acqua, F. DPPNet: An efficient and robust deep learning network for land cover segmentation from high-resolution satellite images. *IEEE Trans. Emerg. Top. Comput. Intell.* **2022**, *7*, 128–139.
13. Al-Najjar, H.A.; Kalantar, B.; Pradhan, B.; Saeidi, V.; Halin, A.A.; Ueda, N.; Mansor, S. Land cover classification from fused DSM and UAV images using convolutional neural networks. *Remote Sens.* **2019**, *11*, 1461. [\[CrossRef\]](#)
14. Minu, M.; Canessane, R.A. Deep learning-based aerial image classification model using inception with residual network and multilayer perceptron. *Microprocess. Microsyst.* **2022**, *95*, 104652. [\[CrossRef\]](#)
15. Shabbir, A.; Ali, N.; Ahmed, J.; Zafar, B.; Rasheed, A.; Sajid, M.; Ahmed, A.; Dar, S.H. Satellite and scene image classification based on transfer learning and fine tuning of ResNet-50. *Math. Probl. Eng.* **2021**, *2021*, 5843816. [\[CrossRef\]](#)
16. Chen, H.; Wang, Z.; Zhang, L. Collaborative spectrum sensing for illegal drone detection: A deep learning-based image classification perspective. *China Commun.* **2020**, *17*, 81–92. [\[CrossRef\]](#)
17. Haq, M.A.; Rahaman, G.; Baral, P.; Ghosh, A. Deep learning based supervised image classification using UAV images for forest areas classification. *J. Indian Soc. Remote Sens.* **2021**, *49*, 601–606. [\[CrossRef\]](#)
18. Kawaguchi, D.; Nakamura, R.; Hadama, H. Evaluation on a drone classification method using UWB radar image recognition with deep learning. In Proceedings of the 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), Helsinki, Finland, 25–28 April 2021; IEEE: New York, NY, USA, 2021.
19. Nguyen, T.N.; Lee, S.; Nguyen, P.-C.; Nguyen-Xuan, H.; Lee, J. Geometrically nonlinear postbuckling behavior of imperfect FG-CNTRC shells under axial compression using isogeometric analysis. *Eur. J. Mech. A/Solids* **2020**, *84*, 104066. [\[CrossRef\]](#)
20. Chehreh, B.; Moutinho, A.; Viegas, C. Latest Trends on Tree Classification and Segmentation Using UAV Data—A Review of Agroforestry Applications. *Remote Sens.* **2023**, *15*, 2263. [\[CrossRef\]](#)
21. Youme, O.; Bayet, T.; Dembele, J.M.; Cambier, C. Deep learning and remote sensing: Detection of dumping waste using UAV. *Procedia Comput. Sci.* **2021**, *185*, 361–369. [\[CrossRef\]](#)
22. Genze, N.; Ajekwe, R.; Güreli, Z.; Haselbeck, F.; Grieb, M.; Grimm, D.G. Deep learning-based early weed segmentation using motion blurred UAV images of sorghum fields. *Comput. Electron. Agric.* **2022**, *202*, 107388. [\[CrossRef\]](#)
23. Shahi, T.B.; Xu, C.-Y.; Neupane, A.; Guo, W. Recent Advances in Crop Disease Detection Using UAV and Deep Learning Techniques. *Remote Sens.* **2023**, *15*, 2450. [\[CrossRef\]](#)
24. Hussain, A.; Ul Amin, S.; Fayaz, M.; Seo, S. An Efficient and Robust Hand Gesture Recognition System of Sign Language Employing Finetuned Inception-V3 and Efficientnet-B0 Network. *Comput. Syst. Sci. Eng.* **2023**, *46*, 3509–3525. [\[CrossRef\]](#)
25. Danish, S.; Khan, A.; Dang, L.M.; Alonazi, M.; Alanazi, S.; Song, H.-K.; Moon, H. Metaverse Applications in Bioinformatics: A Machine Learning Framework for the Discrimination of Anti-Cancer Peptides. *Information* **2024**, *15*, 48. [\[CrossRef\]](#)
26. Behera, T.K.; Bakshi, S.; Sa, P.K. A lightweight deep learning architecture for vegetation segmentation using UAV-captured aerial images. *Sustain. Comput. Inform. Syst.* **2023**, *37*, 100841. [\[CrossRef\]](#)
27. Nguyen, T.N.; Nguyen-Xuan, H.; Lee, J. A novel data-driven nonlinear solver for solid mechanics using time series forecasting. *Finite Elem. Anal. Des.* **2020**, *171*, 103377. [\[CrossRef\]](#)
28. Shanthi, K.; Vidhya, S.S.; Vishakha, K.; Subiksha, S.; Srija, K.; Mamtha, R.S. Algorithms for face recognition drones. *Mater. Today Proc.* **2023**, *80*, 2224–2227. [\[CrossRef\]](#)

29. Aydin, B.; Singha, S. Drone Detection Using YOLOv5. *Eng* **2023**, *4*, 416–433. [[CrossRef](#)]
30. Yao, J. Split Learning for Image Classification in Internet of Drones Networks. In Proceedings of the 2023 IEEE 24th International Conference on High Performance Switching and Routing (HPSR), Albuquerque, NM, USA, 5–7 June 2023; IEEE: New York, NY, USA, 2023.
31. Mohapatra, R.K.; Shaswat, K.; Kedia, S. Offline handwritten signature verification using CNN inspired by inception V1 architecture. In Proceedings of the 2019 Fifth International Conference on Image Information Processing (ICIIP), Shimla, India, 15–17 November 2019; IEEE: New York, NY, USA, 2019.
32. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
34. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]
35. Yang, Y.; Newsam, S. Bag-of-visual-words and spatial extensions for land-use classification. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010.
36. Anuar, M.M.; Halin, A.A.; Perumal, T.; Kalantar, B. Aerial imagery paddy seedlings inspection using deep learning. *Remote Sens.* **2022**, *14*, 274. [[CrossRef](#)]
37. Kareem, R.S.A.; Ramanjineyulu, A.G.; Rajan, R.; Setiawan, R.; Sharma, D.K.; Gupta, M.K.; Joshi, H.; Kumar, A.; Harikrishnan, H.; Sengan, S. Multilabel land cover aerial image classification using convolutional neural networks. *Arab. J. Geosci.* **2021**, *14*, 1681. [[CrossRef](#)]
38. Puttagunta, R.S.; Li, Z.; Bhattacharyya, S.; York, G. Appearance Label Balanced Triplet Loss for Multi-Modal Aerial View Object Classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 20–22 June 2023.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.