sensors

MDPI

Article

Improved Agricultural Field Segmentation in Satellite Imagery Using TL-ResUNet Architecture

Furkat Safarov [1], Kuchkorov Temurbek [2], Djumanov Jamoljon [2], Ochilov Temur [2], Jean Chamberlain Chedjou [3], Akmalbek Bobomirzaevich Abdusalomov [1] and Young-Im Cho [1,*]

[1] Department of Computer Engineering, Gachon University, Sujeong-Gu, Seongnam-Si 461-701, Gyeonggi-Do, Republic of Korea
[2] Department of Computer Systems, Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi, Tashkent 100200, Uzbekistan
[3] Institute of Smart Systems Technologies, University of Klagenfurt, 9020 Klagenfurt, Austria
* Correspondence: yicho@gachon.ac.kr

Abstract: Currently, there is a growing population around the world, and this is particularly true in developing countries, where food security is becoming a major problem. Therefore, agricultural land monitoring, land use classification and analysis, and achieving high yields through efficient land use are important research topics in precision agriculture. Deep learning-based algorithms for the classification of satellite images provide more reliable and accurate results than traditional classification algorithms. In this study, we propose a transfer learning based residual UNet architecture (TL-ResUNet) model, which is a semantic segmentation deep neural network model of land cover classification and segmentation using satellite images. The proposed model combines the strengths of residual network, transfer learning, and UNet architecture. We tested the model on public datasets such as DeepGlobe, and the results showed that our proposed model outperforms the classic models initiated with random weights and pre-trained ImageNet coefficients. The TL-ResUNet model outperforms other models on several metrics commonly used as accuracy and performance measures for semantic segmentation tasks. Particularly, we obtained an IoU score of 0.81 on the validation subset of the DeepGlobe dataset for the TL-ResUNet model.

Keywords: image segmentation; agriculture; satellite imagery; deep learning; UNet architecture; transfer learning

Citation: Safarov, F.; Temurbek, K.; Jamoljon, D.; Temur, O.; Chedjou, J.C.; Abdusalomov, A.B.; Cho, Y.-I. Improved Agricultural Field Segmentation in Satellite Imagery Using TL-ResUNet Architecture. *Sensors* **2022**, *22*, 9784. https://doi.org/10.3390/s22249784

Academic Editor: Chiman Kwan

Received: 25 October 2022
Accepted: 10 December 2022
Published: 13 December 2022

1. Introduction

Most countries in the world, particularly European countries, have great agricultural potential. Some of the most important techniques that use machine and deep learning algorithms to achieve high productivity in precision agriculture include land cover classification and effective management of land resources. Numerous classifications of the physical coverage of the Earth's surface, such as croplands, forests, grasslands, lakes, and wetlands are depicted on land cover maps as spatial information. Dynamic land cover maps incorporate transitions of land cover classes through time, thereby capturing changes in land cover. Land use maps provide geospatial information on the structures, activities, and resources that humans use to establish, enhance, or sustain a particular type of land cover.

More objects can now be identified in satellite images because of the rise in spatial resolution, and studies have switched from spectral image classification, pixel-based image analysis, and object-based image analysis to pixel-level semantic segmentation. In this study, we analyze the development of semantic segmentation techniques based on deep learning and propose a TL-ResUNet segmentation model for land use/cover.

In deep learning, many algorithms for classifying satellite images provide more reliable and accurate results than traditional classification algorithms, and numerous researchers are conducting various scientific and practical studies in this field [1–3]. Land use/cover

maps are generated from different high-resolution satellite images, such as Sentinel [4], Landsat [5], and Worldview [6] satellite missions. These images can be used to classify different types of land cover, such as permanent water, built-up areas, residential areas, and agricultural fields. The Copernicus land monitoring service platform maintains general statistics on land use and cover across the world.

High-resolution satellite images have complex and deep features that require complex operations for image recognition. Creating land use maps is one of the most significant uses of satellite imagery, and this is possible through image segmentation and classification procedures. In recent years, different tasks and applications, such as producing regional and global land cover maps, creating advanced supervised and unsupervised classification algorithms, region-based image analysis, using numerous remote sensing features, and integrating map data into classification procedures, such as data on soil, roads, farmlands, crops, and other census data, have all seen significant advancements in the field of image classification. The main tasks of satellite image analysis are multi-object detection and classification that analyze numerical features or properties associated with an image, which can be divided into different classes. Comprehensive monitoring requires a highly productive evaluation of land cover via image segmentation and classification in different fields, particularly in agriculture.

Since 2012, CNN-based algorithms have been effectively used to solve classification tasks [7–10]. A set of convolution filters is used in each layer to identify image characteristics and features' structure [11]. The most popular CNN-based architectures, such as GoogleNet, VGGNet, AlexNet, and ZFNet, have recently been used for image classification. However, calculating land use from satellite imagery through the classical approach of classification is difficult. Thus, segmentation-based classification has become significantly efficient and smart. Deep learning is a group of machine learning methods used in image analysis to learn and display features, such as edges, curves, and patterns from an input image. CNN and FCN are well-known deep learning techniques for image analysis. CNN-based structures include convolution, pooling, dropout, batch normalization, and non-linearity operation layers.

Therefore, this study presents a CNN-based UNet architecture, residual network, and transfer learning for land use classification of satellite images through semantic segmentation. Additionally, we discuss an overview of the recent deep learning-based techniques for satellite image classification and the available training datasets.

The main contribution of this work is improving model performance and accuracy using a combination of residual network, transfer learning, and UNet architecture. Generally, UNet is a robust architecture for segmentation tasks. Since land cover and land use classification task is complex, UNet coupled with residual networks and transfer learning yields better results.

The rest of this study is divided as follows: Section 2 analyzes various recent and relevant research papers; Section 3 studies available common datasets for satellite image segmentation and classification; Section 4 proposes our encode-decoder-based deep learning architecture (TL-ResUNet); Section 5 presents the experimental settings as well as the qualitative and quantitative analysis of the semantic segmentation results; and Section 6 presents the final remarks and conclusion of the study.

2. Related Works

Land cover has been studied in several research papers ranging from machine learning to deep learning. Using neural networks a decade ago was unpreferable because of their high computational complexity. Histogram thresholding provided satisfactory results, but exhibited problems associated with the variations and challenges in satellite images. Similarly, classical machine learning algorithms, such as support vector machines and random forest methods, were used for LULC mapping. For example, in [12,13] they have applied these methods for land cover classification. In the land cover classification study that uses machine learning, a decision tree and an artificial neural network were applied

to Landsat ETM+ data to classify land cover. However, the drawback of these methods is that they require in-depth knowledge of the feature extraction process to improve model performance.

However, recent studies show that deep learning algorithms are widely used in classification and segmentation tasks. Especially, due to greater number of features and complex structure of satellite images, deep learning yields better results in LULC tasks such as agricultural field monitoring, forest change detection, water resources monitoring, building detection, and urbanization. For example, automatic building recognition method have been implemented in [14] which collected a dataset using the MapBox API for OpenStreetMap to create a satellite image with building masks. Furthermore, the pixel-wise image segmentation methods for classifying different attributes of satellite images is explained in study [15]. Here the proposed method can achieve a high accuracy using the UNet model to detect a building in the INRIA dataset, which is composed of very high-resolution images. However, these studies only focused on segmenting one class. Developing a model which uses multi-class segmentation of satellite images is a more complex task. Deep learning architectures, such as UNet and DenseNet, are actively used for image segmentation, whereas architectures, such as ResNet, VGG, and EfficientNet, are used for classification tasks in computer vision. According to the results of recent research works, these deep learning models outperform classical feature extraction algorithms.

However, in terms of satellite image processing, more work must be done to achieve high performance. For example, results of modern semantic segmentation were not satisfactory [16–19] due to the complex shape of satellite images. Kuo et al. [20] proposed a method that delivers one of the top results in the DeepGlobe challenge, in which improving the performance of model depends on a variation of DeepLabV3+. Despite this, their model accuracy is not good because of the fixed value of the standard deviation gaussian filter. Renee Su et al. proposed a semantic segmentation model using DeepLab v3+ with an IoU score of 0.756, and as a dataset they used the DeepGlobe dataset [21]. However, their model requires a greater number of satellite images to train because the authors did not apply any augmentation techniques. SegNet is a deep convolutional encoder-decoder architecture which is a very effective model among the numerous image segmentation models. Lee et al. applied the SegNet model to an aerial image to categorize the land cover and then performed research to assess the accuracy of that classification [22].

In [23], authors proposed an architecture using DeepLab and ResNet18 as the backbone, accomplishing an IoU score of 0.433 s of the DeepGlobe land cover data. The authors of the transfer learning approach in this study used two neural network architectures. The ResNet50 model was used for classification. After classification, a pre-trained ResNet50 model was used as an encoder in the modified UNet model for segmentation [24]. The accuracy was not so high, and the authors claim that this is mainly because of the quality of the dataset. Also, the authors conclude that the CORINE dataset is not suitable for training machine learning algorithms.

One of the main components of LULC is agricultural field monitoring. Several studies were conducted for farmland segmentation using low resolution images [25,26]. However, in [27] researchers generated a new benchmark dataset from VHR Worldview-3 images for twelve distinct LULC classes of two different geographical locations. Segmentation using low resolution satellite images can be used to classify tasks of global or general changes in areas, whereas high-resolution images should be used for segmenting specific objects such as multi-class segmentation and small objects.

3. Datasets

We collected publicly available satellite images for training and testing. However, the training dataset is constrained using this approach for satellite image classification and segmentation. To address this, we used image augmentation and various computer vision techniques to enhance the number of satellite frames. The shortage of labeled training data in a dataset has been one of the greatest challenges in adopting deep convolutional network

pipelines in satellite image classification and segmentation. Datasets are created using middle or low-resolution satellite images. However, low and middle resolution satellite images may not produce the expected accuracy in satellite image segmentation. Pixel-based segmentation masks for image segmentation are considerably difficult to create. Applying a poorly supervised learning strategy, which is used in [28,29], is a method for tackling the lack of training data. The objective of weakly supervised methods is to reduce the need for complicated training datasets. Nivaggioli et al. [28] used a previously suggested method by producing pixel-level annotation from image-level annotation. They performed cropland segmentation using two types of labels commonly found in remote sensing datasets in [29]. To construct pixel-level maps of land cover, the study investigates weak labels in the form of a single-pixel label per image and class activation maps.

3.1. Labeled DeepGlobe Data

The DeepGlobe land cover classification challenge is the first publicly available dataset that focuses on rural regions using high-resolution submeter satellite images, as shown in Figure 1. The DeepGlobe dataset consists of approximately 1200 satellite images with a pixel size of 2448 × 2448, divided into training, validation, and test sets with a percentage of 70%, 15%, and 15%, respectively. Each image had RGB channels from the DigitalGlobe Vivid+ dataset with pixels at a resolution of 50 cm. Each satellite image was linked to a mask image to label the land cover. The mask is an RGB picture with seven classes, such as urban, agriculture, rangeland, forest, water, bare, and unknown (Table 1).

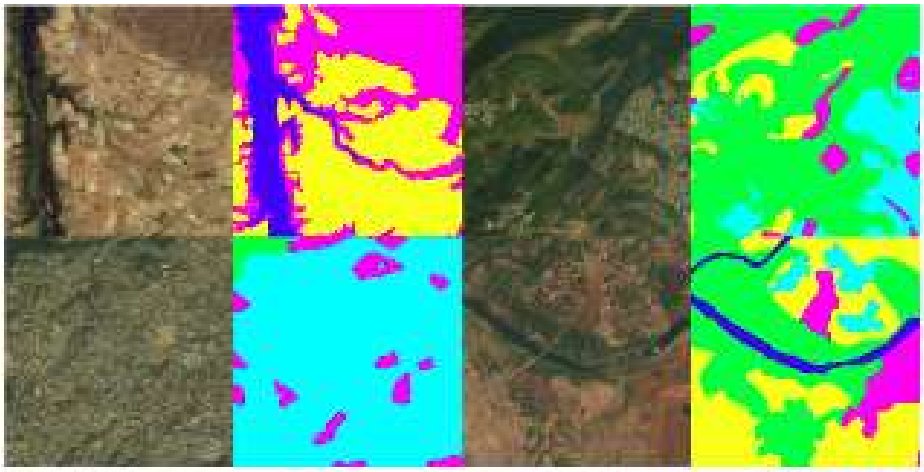

Figure 1. The land cover original image (**left**) and class label (**right**) pairs.

Table 1. Classes in the label data of the DeepGlobe dataset.

Class/Color	Pixel Count	Proportion
Urban	642.4 M	9.35%
Agriculture	3898.0 M	56.76%
Rangeland	701.1 M	10.21%
Forest	944.4 M	13.75%
Water	256.9 M	3.74%
Barren	421.8 M	6.14%
Unknown	3.0 M	0.04%

3.2. Defence Science and Technology Laboratory (Dstl) Dataset

The Dstl Kaggle dataset [30] is the second dataset, which provides 57 satellite images in a region of 1 sq. km. in both three-band RGB and 16-band multispectral formats. Here, we use three-band images with a spatial resolution of 1.24 m. In this dataset, 10 different classes, such as roads, buildings, vehicles, farms, trees, waterways, and others, have been

labeled within particular images. The panchromatic waveband ranges from 450 to 800 nm, whereas 8 multispectral (red, red edge, coastal, blue, green, yellow, near-IR1, and near-IR2) wavebands are between 400 and 1040 nm. According to the sensor resolution at Nadir, panchromatic, multispectral, and SWIR bands are equivalent to 0.31, 1.24, and 7.5 m, respectively [31].

3.3. LandCoverNet

The multispectral satellite imagery from the Sentinel-2 mission in 2018 is labeled using the worldwide yearly LandCoverNet training dataset, as shown in Table 2. This dataset contains data across Africa, and each pixel of the image is identified as one of the seven land cover classes, such as water, woody vegetation, cultivated vegetation, semi-natural vegetation, permanent snow/ice, natural bare ground, and artificial bare ground, based on its annual time series.

Table 2. Comparison of main the characteristics of the abovementioned datasets such as DeepGLobe, Dstl and LandCoverNet.

Datasets	Number of Classes	Spatial Resolution	Number of Images
DeepGlobe	7	1.24 m	1146
Dstl	10	1.24 m	57
LandCoverNet	7	10 m	1980
Augmented Images	7	-	6517
Total	-	-	9700

The first version of this dataset contains 1980 images with a size of 256 × 256 pixels, which contains 66 tiles from the Sentinel-2. Each image chip includes an annual class label and temporal data from the Sentinel-2 surface reflectance product (L2A) at a 10-m spatial resolution, which is stored as a GeoTIFF data format. The resolution and an annual class label of each image are stored in a raster format, precisely as GeoTIFF files [32].

Table 2 compares datasets in terms of number of classes, spatial resolution, and number of images. While both Dstl and DeepGlobe are high resolution images, the latter was chosen for the proposed model because of the greater number of images. As mentioned earlier, during the experiments we found that image data augmentation approaches, such as geometric transformations, brightness/contrast enhancement, and data normalization, proved to be the most effective way to improve the final accuracy rate. The effectiveness of deep learning models depends on the size and resolution of the training image datasets. Therefore, we rotated each original image and then flipped each rotated image horizontally to increase the number of images in the satellite segmentation dataset. By applying the data augmentation methods to the original 3183 fire images, we increased the total number of images to 9700.

4. Proposed Architecture

Two different neural network designs are suggested in this study. The first neural network architecture used for the segmentation task was the modified UNet model [33,34]. The second was the ResNet-50 model [35], which served both as the classification model and as an encoder for the modified UNet model, as shown in Figure 2. The UNet model was trained using different methods of ResNet backbone weight initialization models, that is, with random weights and ResNet pre-trained on the ImageNet dataset. With the help of this transfer learning strategy, we may apply the knowledge obtained from the first task to a new one, which is a more challenging task because obtaining training data is extremely difficult.

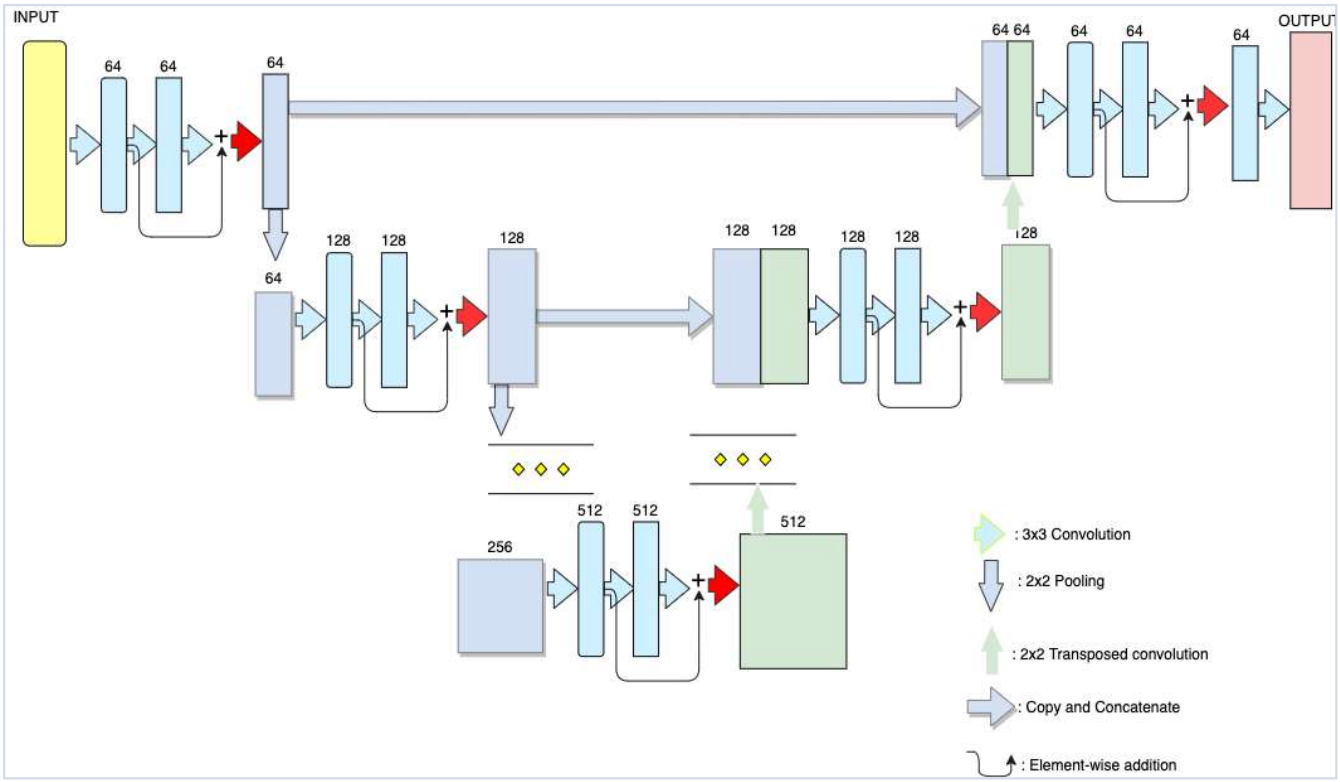

Figure 2. Modified UNet with a ResNet-50 encoder.

Additionally, the DeepGlobe dataset was used to train the satellite image segmentation model, which allowed for the use of ResNet weights that had already been learned, except for modifying and training the final layers of the network.

With regards to DeepGlobe dataset, it includes high resolution images with 1.24 m spatial resolution. The minimum requirements for the dataset is around 1000 high resolution satellite images, since deep learning models require greater number of images for training effectively. Using data augmentation techniques, the number of images in dataset increases during training model. The proposed model was trained using 9700 images.

4.1. Modified ResUNet Architecture

UNet is the most easily scalable and sizable fully convolutional network architecture for semantic segmentation. Generally, UNet architecture consists of two paths: a path that contracts to record context and another that expands symmetrically to enable exact localization. The contracting path follows a similar architecture to the ResNet architecture, where there is a long skip connection on every level; moreover, there are local skip connections between convolutions at each step. Feature maps are downsampled during the convolution processes, which also increases the number of feature maps per layer. However, they are upsampled before each step in the expanded route by a transposed convolution and this expanding branch boosts the resolution of the feature map. The expanding path uses skip connections to mix high-resolution features from the contracting path with upsampled features to localize them [35]. The output of the UNet model is a pixel-wise mask that shows the class of each pixel.

We applied transposed convolution layers to build a matching decoder, which doubles the size of a feature map while cutting the number of channels in half. Then, the output of a transposed convolution is concatenated with an output of the corresponding part of the decoder. To maintain the same number of channels as in a symmetric encoder term, the resulting feature map is applied to a convolution process. Figure 3 shows that this upsampling process can be repeated several times to couple with max pooling layers.

Technically, fully connected layers can accept inputs of any size, but because our max pooling layer downsamples each image twice, the present network implementation can only accept inputs with sides divisible by two.

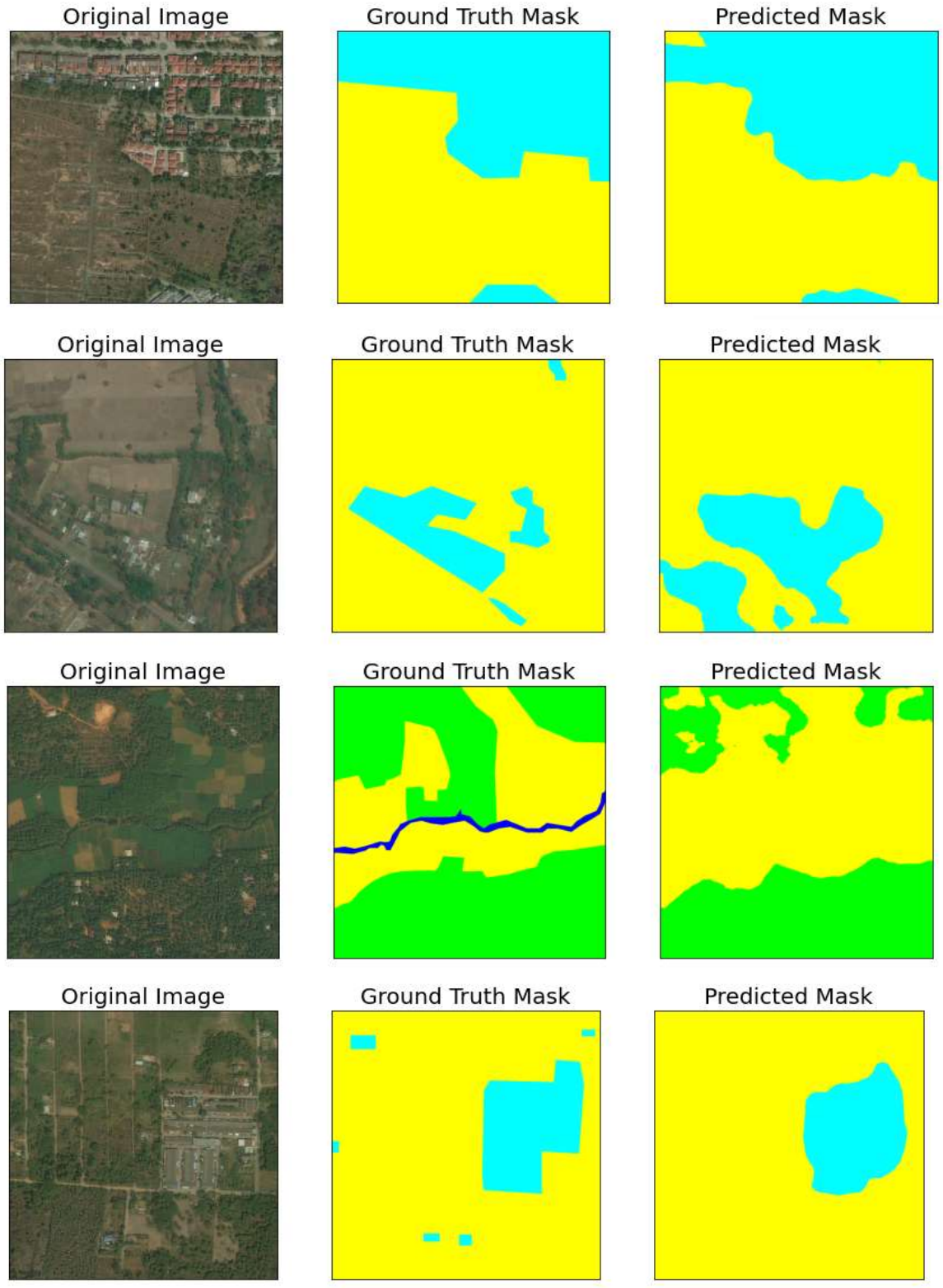

Figure 3. *Cont*.

Figure 3. Trained model results.

4.2. ResNet Architecture

As an encoder of UNet, we used the pre-trained ResNet architecture, which consists of 48 convolution layers and 1 MaxPool layer, known as ReNnet-50. The advantage of ResNet over the sequential convolutional networks is that it can avoid the vanishing gradient problem and mitigate the degradation problem, where adding more layers to the model causes higher training errors. The ResNet architecture uses the repetitive layers of ResBlocks, that is, the blocks with skip connections, which make the network deeper while avoiding model degradation. After winning the ImageNet large-scale visual recognition contest for image classification in 2015, the ResNet architecture became recognized as the most sophisticated model architecture for image classification [35–37].

4.3. Evaluation Metrics

Both architecture models that modified UNet and ResNet-50 were assessed using the validation set, which consisted of 20% data. Key metrics, such as precision, recall, F1 score [38–41], and Jaccard index [42–45], were used to evaluate the model results. The precision metric was used to calculate the percentage of correctly labeled predictions across all predicted labels. It is the ratio of the true positive (TP) and false positive (FP) results (1):

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

The *recall* metric was also considered to measure the proportion of correct labels in all predicted labels. It is the ratio of the TP and false negative (FN) results (2):

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

The F_1 score was used as a result of training the ResNet classifier, which combines precision and recall with the same weights (3):

$$F_1 = \frac{2 \times precision \times recall}{(precision) + recall} \tag{3}$$

The results of the UNet segmentation model were evaluated using the Jaccard index metric (4):

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{4}$$

4.4. Loss Functions

We can modify the evaluation metric of the Jaccard index for discrete pixel-wise picture objects, where y_i is the binary value (label) of the associated pixel and y^i is the expected

probability for the pixel. We used binary cross-entropy for the segmentation task since it can be viewed as a pixel-by-pixel classification problem. The area of intersection (J) between predicted masks and the related ground truth data is maximized by minimizing the loss function, which also optimizes the probability of correctly predicted pixels [46–48].

5. Experimental Results

The modified UNet model was used for semantic segmentation of the DeepGlobe dataset. The final modified TL-ResUNet model with initialized and tuned ResNet-50 encoder was trained and evaluated on the DeepGlobe dataset.

5.1. Model Training

The Pytorch framework is mostly recommended to train the machine and deep learning models. The modified UNet architecture is prototyped using the PyTorch framework by combining the building blocks of the ResNet-50 as an encoder. GPU servers with Nvidia Tesla V100 graphics cards and 43 GB of RAM were used during the training.

The ResNet-50 model is trained on the DeepGlobe dataset and initialized using weights of the ResNet-50 model pre-trained on the ImageNet dataset. The model is trained in two stages because a pre-trained model was used from the beginning: the first stage involves training just the last layers, whereas the second involves unfreezing all the layers. The model was trained with 20 epochs in total, i.e., 13 and 7 epochs for the first and second stages, respectively.

Three methods of weight initialization were considered during the training. First, the weights were initialized using a LeCun uniform initializer, which has a random uniform distribution within [−L, L], where L = sqrt ($1/f_{in}$) and f_{in} is the number of input units in the weight tensor. Second, we reused the same architecture with the ResNet-50 encoder pre-trained on ImageNet, and all layers in the decoder were initialized using the LeCun uniform initializer. Third, we also used the latest trained segmentation model by initializing the encoder with the ResNet-50 but pre-trained on the DeepGlobe dataset, as shown in Figure 3.

5.2. Results

For the validation subset of the model using the DeepGlobe dataset, we achieved the following results after 30 training epochs:

(1) Best score on randomly initialized weights: IoU = 0.68;
(2) Best score on the encoder pre-trained weights on ImageNet: IoU = 0.81.

Although the model performs well in the majority of cases (Figure 3), it may fail to detect some classes such as narrow water bodies. An example of this case is given below. Likewise, small, forested areas are misclassified in some cases (Figure 4). However, dense forested areas are classified correctly though they are located near agricultural fields. Distinguishing forested areas from farming lands is a challenging task. Furthermore, the model performs extremely good for some classes such as urban and farming lands.

The learning curves for validation in Figure 5 below show the results of each approach. A steady value is attained faster in pre-trained networks than in the randomly initialized network, and the steady value is visually higher in the pre-trained models.

The visualization of overlaying the masks on the original image demonstrates the advantage of training with the pre-trained models.

Note that the hyperparameter optimization techniques or the dataset preprocessing can be applied to further improve the performance of the models. Table 3 specifies the detailed scores.

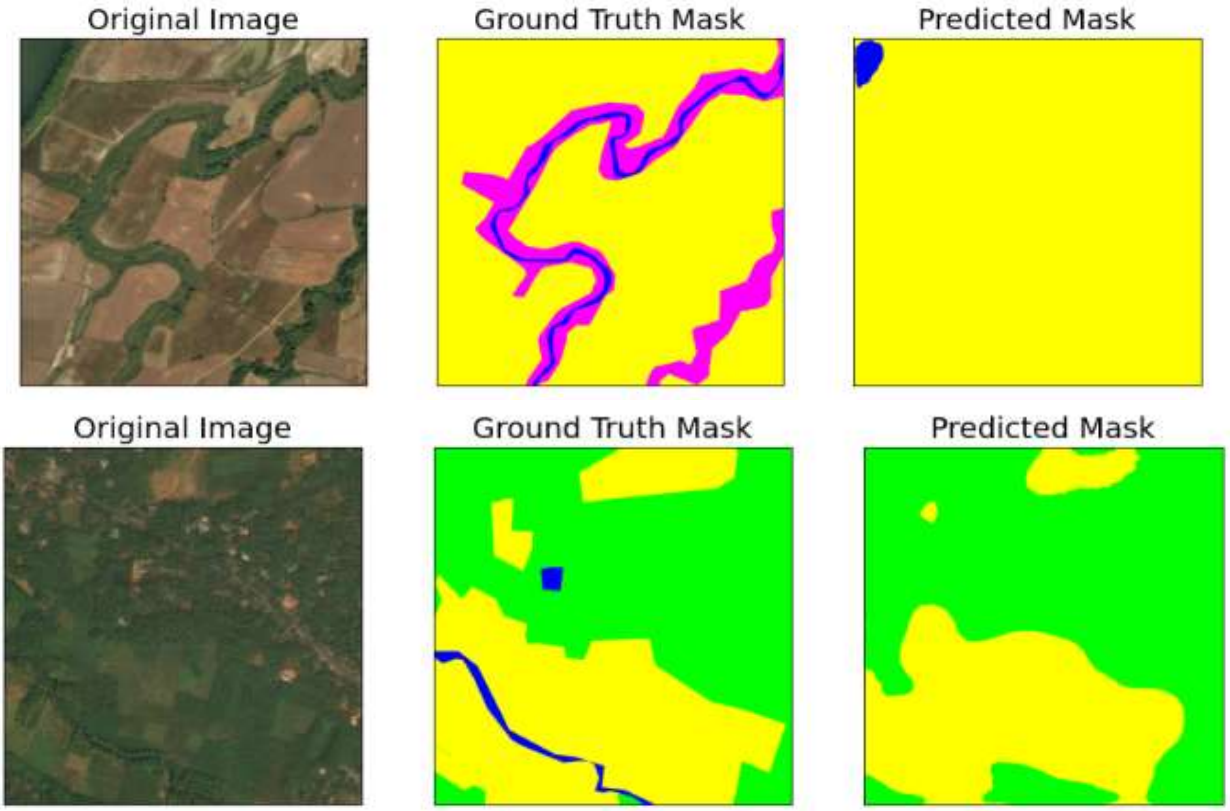

Figure 4. Examples of misclassification by the model.

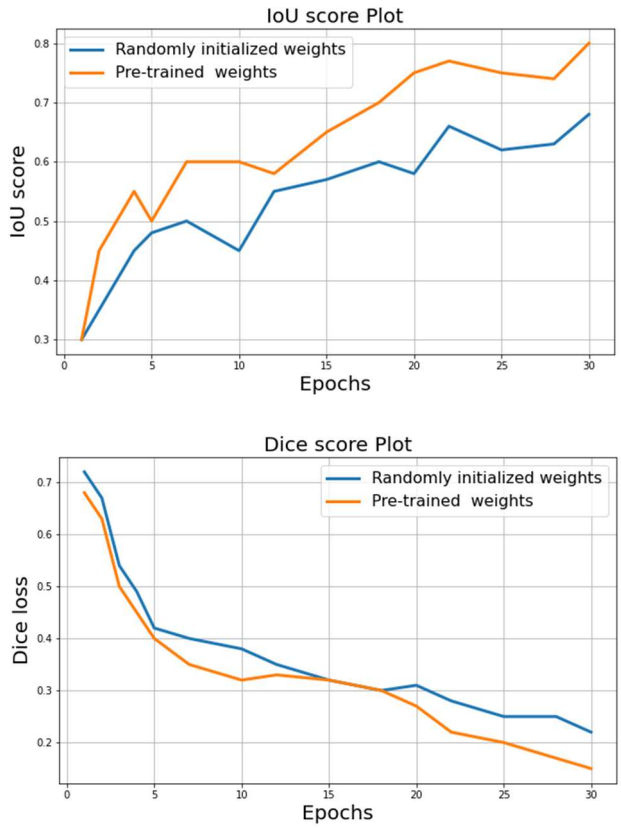

Figure 5. Learning curves of trained model at the training and validation stages.

Table 3. Comparison of the UNet training results on IoU metric throughout epochs.

Epoch	Train/Validation	UNet Trained with Randomly Initialized Weights	UNet Trained with Weights Trained on ImageNet without Residual Layers	UNet Trained with Weights Trained on ImageNet
10	train	0.51	0.52	0.54
	validation	0.49	0.50	0.51
20	train	0.59	0.62	0.69
	validation	0.58	0.61	0.64
30	train	0.68	0.75	0.84
	validation	0.68	0.74	0.81

The hyperparameter tuning techniques and results are shown in Table 3. Overall, it can be seen that UNet trained with transfer learning and residual layers can learn features faster and in an effective way. While UNet with random weights reaches a 0.68 IoU score in 30 epochs, UNet with ImageNet weights achieves a 0.74 IoU score. Finally, UNet with ResNet50 and with ImageNet weights achieved a 0.81 IoU score.

To further understand the advantage of our model against the others, we show some comparative results in Table 4. ClassmateNet produces fair segmentation output for larger areas but fails to segment short details such as smaller areas and field boundaries. DeepLabv3 and DeepLabv3+ improve performance on these details; however, they also produce artefacts and fail to keep producing stable results at larger areas in some cases. However, our model combines multi-level features effectively and produces more accurate segmentation results at both larger and detail areas.

Table 4. Comparison of the TL-ResUNet with other models.

Algorithms	IoU
Baseline	55.19
ClassmateNet	69.87
DFCNet	71.31
DeepLabv3	74.52
DeepLabv3+	75.6
TL-ResUNet	81.0

We compared the robustness and weaknesses of previous methods with the proposed method in different categories using quantitative and qualitative performance results, as shown in Table 5. Based on the evaluated scores, the performance of the proposed approach did not suffer with densely forested areas and classified them correctly though they are located near agricultural fields. In addition, the model performs extremely well for some classes such as urban and farming lands.

The outcomes of segmentation methods can be divided into three categories: robust, standard, and powerless. Robust measures show that the method is applicable to segment all types of land/field segmentation. The algorithm may fail in some circumstances, such as narrow water bodies or small forested areas, according to normal standards. Powerless evidence suggests that algorithms are unreliable in the presence of noise or color, and the land classification procedure frequently modifies the initial geometry of moving objects.

Table 5. Evaluation of the robustness and weaknesses of segmentation methods using different characteristics.

Criterion	DFCNet	DeepLabv3	DeepLabv3+	Proposed Method
Scene Independence	standard	robust	standard	robust
Object Independence	standard	robust	robust	standard
Robust to Noise	powerless	robust	standard	robust
Robust to Color	standard	standard	powerless	standard
Small Land Segmentation	robust	standard	robust	robust
Multiple Land Segmentation	standard	powerless	powerless	powerless
Processing Time	powerless	standard	robust	robust

6. Conclusions

In this study, we proposed a modified semantic segmentation deep neural network model called the TL-ResUNet for land use/cover classification and segmentation of satellite images. This developed model includes residual learning, UNet architecture, and a transfer learning approach. The proposed architecture section discussed the implementation of efficient training of the UNet model using pre-trained weights. The ResNet-50 model with pre-trained weights was chosen as a backbone of the UNet for experimental purposes. For the ease of building, training, and using the neural network, the library of the segmentation model, which is based on the PyTorch deep learning framework, was chosen. Finally, the environment and results of experimental training were analyzed using the commonly used IoU metric to determine the score of similarity of the predicted map and expected ground truth map. In the experiment, we verified the effectiveness of our proposed model and demonstrated that our model performs satisfactorily against the state-of-the-art models on the land use and cover task.

Future tasks include solving misclassification problems under similar color conditions and increasing the accuracy of the approach. We plan to develop a small real-time "land use land cover" model with YOLOv networks [49–51] using feature analyzing and extraction approach [52–56].

Author Contributions: This manuscript was designed and written by K.T. and F.S.; conceptualization, D.J. and J.C.C.; methodology, K.T.; software, O.T.; visualization, A.B.A.; and supervision, Y.-I.C. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by Korea Agency for Technology and Standards in 2022, project numbers are K_G012002073401, K_G012002234001 and by the Gachon University research fund of 2020 (GCU-202008460006).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Acknowledgments: The authors would like to express their sincere gratitude and appreciation to their supervisor, Young Im Cho (Gachon University), for her support, comments, remarks, and engagement over the period in which this manuscript was written. Moreover, the authors would like to thank the editor and anonymous referees for their constructive comments on improving the content and presentation of this paper.

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

References

1. Neupane, B.; Horanont, T.; Aryal, J. Deep Learning-Based Semantic Segmentation of Urban Features in Satellite Images: A Review and Meta-Analysis. *Remote Sens.* **2021**, *13*, 808. [CrossRef]
2. Shafaey, M.A.; Salem, M.A.M.; Ebied, H.M.; Al-Berry, M.N.; Tolba, M.F. *Deep Learning for Satellite Image Classification*; Springer: Berlin/Heidelberg, Germany, 2019; Volume 2019, pp. 383–391.
3. Alias, B.; Karthika, R.; Parameswaran, L. Classification of high resolution remote sensing images using deep learning techniques. In Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore Karnataka, India, 19–22 September 2018.
4. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]
5. Irons, J.R.; Dwyer, J.L.; Barsi, J.A. The next Landsat satellite: The Landsat Data Continuity Mission. *Remote Sens. Environ.* **2012**, *122*, 11–21. [CrossRef]
6. Johnson, K.; Koperski, K. WorldView-3 SWIR land use-land cover mineral classification: Cuprite, Nevada. *Remote Sens. GIS* **2017**. Available online: https://www.researchgate.net/project/Remote-Sensing-and-GIS-4 (accessed on 22 July 2022).
7. Scott, G.J.; England, M.R.; Starms, W.A.; Marcum, R.A.; Davis, C.H. Training Deep Convolutional Neural Networks for Land–Cover Classification of High-Resolution Imagery. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 549–553. [CrossRef]
8. Musaev, M.; Khujayorov, I.; Ochilov, M. Image Approach to Speech Recognition on CNN. In Proceedings of the 2019 3rd International Symposium on Computer Science and Intelligent Control (ISCSIC 2019), Amsterdam, The Netherlands, 25–27 September 2019; Article 57. pp. 1–6. [CrossRef]
9. Mukhamadiyev, A.; Khujayarov, I.; Djuraev, O.; Cho, J. Automatic Speech Recognition Method Based on Deep Learning Approaches for Uzbek Language. *Sensors* **2022**, *22*, 3683. [CrossRef] [PubMed]
10. Valikhujaev, Y.; Abdusalomov, A.; Cho, Y. Automatic Fire and Smoke Detection Method for Surveillance Systems Based on Dilated CNNs. *Atmosphere* **2020**, *11*, 1241. [CrossRef]
11. Kuchkorov, T.A.; Urmanov, S.N.; Nosirov, K.K.; Kyamakya, K. Perspectives of deep learning based satellite imagery analysis and efficient training of the U-Net architecture for land-use classification. In *World Scientific Proceedings Series on Computer Engineering and Information Science, Developments of Artificial Intelligence Technologies in Computation and Robotics*; World Scientific: Singapore, 2020; pp. 1041–1048.
12. Bengana, N.; Heikkilä, J. Improving land cover segmentation across satellites using domain adaptation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 1399–1410. [CrossRef]
13. Tian, C.; Li, C.; Shi, J. Dense fusion classmate network for land cover classification. In Proceedings of the IEEE/CVF Conference on Computing and Vision Pattern Recognition Workshops 2018, Salt Lake City, UT, USA, 18–22 June 2018; p. 262.
14. Chhor, G.; Aramburu, C.B.; Bougdal-Lambert, I. Satellite Image Segmentation for Building Detection using U-net. *Comput. Sci. Semant. Sch.* **2017**, *15*, 114–120.
15. Karwowska, K.; Wierzbicki, D. Improving Spatial Resolution of Satellite Imagery Using Generative Adversarial Networks and Window Functions. *Remote Sens.* **2022**, *14*, 6285. [CrossRef]
16. Wafa, R.; Khan, M.Q.; Malik, F.; Abdusalomov, A.B.; Cho, Y.I.; Odarchenko, R. The Impact of Agile Methodology on Project Success, with a Moderating Role of Person's Job Fit in the IT Industry of Pakistan. *Appl. Sci.* **2022**, *12*, 10698. [CrossRef]
17. Abdusalomov, A.; Mukhiddinov, M.; Djuraev, O.; Khamdamov, U.; Whangbo, T.K. Automatic Salient Object Extraction Based on Locally Adaptive Thresholding to Generate Tactile Graphics. *Appl. Sci.* **2020**, *10*, 3350. [CrossRef]
18. Sevak, J.S.; Kapadia, A.D.; Chavda, J.B.; Shah, A.; Rahevar, M. Survey on semantic image segmentation techniques. In Proceedings of the 2017 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, 7–8 December 2017; pp. 306–313. [CrossRef]
19. Huang, G.; Liu, Z.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computing and Vision Pattern Recognition 2017, Honolulu, HI, USA, 21–26 June 2017; pp. 2261–2269.
20. Kuo, T.S.; Tseng, K.S.; Yan, J.; Liu, Y.C.; Wang, Y.C.F. Deep aggregation net for land cover classification. In Proceedings of the IEEE/CVF Conference on Computing and Vision Pattern Recognition Workshops 2018, Salt Lake City, UT, USA, 18–22 June 2018; p. 247.
21. Su, R.; Chen, R. Land cover change detection via semantic segmentation. *arXiv* **2019**, arXiv:1911.12903.
22. Lee, S.; Park, S.; Son, S.; Han, J.; Kim, S.; Kim, J. Land cover segmentation of aerial imagery using SegNet. *Earth Resour. Environ. Remote Sens./GIS Appl. X. SPIE* **2019**, *11156*, 313–318.
23. Demir, I.; Koperski, K.; Lindenbaum, D.; Pang, G.; Huang, J.; Basu, S.; Hughes, F.; Tuia, D.; Raskar, R. DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018.
24. Ulmas, P.; Liiv, I. Segmentation of satellite imagery using U-net models for land cover classification. *arXiv* **2020**, arXiv:2003.02899.
25. Sharifzadeh, S.; Tata, J.; Sharifzadeh, H.; Tan, B. Farm Area Segmentation in Satellite Images Using DeepLabv3+ Neural Networks. In *Data Management Technologies and Applications, DATA 2019*; Hammoudi, S., Quix, C., Bernardino, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2020; Volume 1255. [CrossRef]

26. Kutlimuratov, A.; Abdusalomov, A.; Whangbo, T.K. Evolving Hierarchical and Tag Information via the Deeply Enhanced Weighted Non-Negative Matrix Factorization of Rating Predictions. *Symmetry* **2020**, *12*, 1930. [CrossRef]
27. Sertel, E.; Ekim, B.; Osgouei, P.E.; Kabadayi, M.E. Land Use and Land Cover Mapping Using Deep Learning Based Segmentation Approaches and VHR Worldview-3 Images. *Remote Sens.* **2022**, *14*, 4558. [CrossRef]
28. Nivaggioli, A.; Randrianarivo, H. Weakly Supervised Semantic Segmentation of Satellite Images. Clinical Orthopaedics and Related Research. 2019. Available online: http://arxiv.org/abs/1904.03983 (accessed on 22 July 2022).
29. Wang, S.; Chen, W.; Xie, S.M.; Azzari, G.; Lobell, D.B. Weakly Supervised Deep Learning for Segmentation of Remote Sensing Imagery. *Remote Sens.* **2020**, *12*, 207. [CrossRef]
30. Dstl Satellite Imagery Feature Detection. Available online: https://www.kaggle.com/competitions/dstl-satellite-imagery-feature-detection/data (accessed on 22 July 2022).
31. Li, Q.; Shi, Y.; Huang, X.; Zhu, X.X. Building Footprint Generation by Integrating Convolution Neural Network with Feature Pairwise Conditional Random Field (FPCRF). *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7502–7519. [CrossRef]
32. Alemohammad, H.; Booth, K. LandCoverNet: A global benchmark land cover classification training dataset. *arXiv* **2020**, arXiv:2012.03111.
33. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. *arXiv* **2015**, arXiv:1505.04597.
34. Nodirov, J.; Abdusalomov, A.B.; Whangbo, T.K. Attention 3D U-Net with Multiple Skip Connections for Segmentation of Brain Tumor Images. *Sensors* **2022**, *22*, 6501. [CrossRef] [PubMed]
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computing and Vision Pattern Recognition 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
36. Kuchkorov, T.; Ochilov, T.; Gaybulloev, E.; Sobitova, N.; Ruzibaev, O. Agro-field Boundary Detection using Mask R-CNN from Satellite and Aerial Images. In Proceedings of the 2021 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 3–5 November 2021; pp. 1–3. [CrossRef]
37. Kuchkorov, T.; Urmanov, S.; Kuvvatova, M.; Anvarov, I. Satellite image formation and preprocessing methods. In Proceedings of the 2020 International Conference on Information Science and Communications Technologies (ICISCT), Sanya, China, 4–6 December 2020; pp. 1–4. [CrossRef]
38. Hossin, M.; Sulaiman, M.N. A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process* **2015**, *5*, 1–11. [CrossRef]
39. Abdusalomov, A.; Whangbo, T.K. An improvement for the foreground recognition method using shadow removal technique for indoor environments. *Int. J. Wavelets Multiresolution Inf. Process.* **2017**, *15*, 1750039. [CrossRef]
40. Abdusalomov, A.; Whangbo, T.K. Detection and Removal of Moving Object Shadows Using Geometry and Color Information for Indoor Video Streams. *Appl. Sci.* **2019**, *9*, 5165. [CrossRef]
41. Farkhod, A.; Abdusalomov, A.; Makhmudov, F.; Cho, Y.I. LDA-Based Topic Modeling Sentiment Analysis Using Topic/Document/Sentence (TDS) Model. *Appl. Sci.* **2021**, *11*, 11091. [CrossRef]
42. Fletcher, S.; Islam, Z. Comparing sets of patterns with the Jaccard index. *Australas. J. Inf. Syst.* **2018**, *22*, 220. [CrossRef]
43. Jakhongir, N.; Abdusalomov, A.; Whangbo, T.K. 3D Volume Reconstruction from MRI Slices based on VTK. In Proceedings of the 2021 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Republic of Korea, 19–21 October 2021; pp. 689–692. [CrossRef]
44. Umirzakova, S.; Abdusalomov, A.; Whangbo, T.K. Fully Automatic Stroke Symptom Detection Method Based on Facial Features and Moving Hand Differences. In Proceedings of the 2019 International Symposium on Multimedia and Communication Technology (ISMAC), Quezon City, Philippines, 19–21 August 2019; pp. 1–5. [CrossRef]
45. Kutlimuratov, A.; Abdusalomov, A.B.; Oteniyazov, R.; Mirzakhalilov, S.; Whangbo, T.K. Modeling and Applying Implicit Dormant Features for Recommendation via Clustering and Deep Factorization. *Sensors* **2022**, *22*, 8224. [CrossRef]
46. Ayvaz, U.; Gürüler, H.; Khan, F.; Ahmed, N.; Whangbo, T.; Abdusalomov, A. Automatic Speaker Recognition Using Mel-Frequency Cepstral Coefficients Through Machine Learning. *CMC-Comput. Mater. Contin.* **2022**, *71*, 5511–5521. [CrossRef]
47. Makhmudov, F.; Mukhiddinov, M.; Abdusalomov, A.; Avazov, K.; Khamdamov, U.; Cho, Y.I. Improvement of the end-to-end scene text recognition method for "text-to-speech" conversion. *Int. J. Wavelets Multiresolution Inf. Process.* **2020**, *18*, 2050052. [CrossRef]
48. Khamdamov, R.; Saliev, E.; Rakhmanov, K. Classification of crops by multispectral satellite images of sentinel 2 based on the analysis of vegetation signatures. *J. Phys. Conf. Ser.* **2020**, *1441*, 012143. [CrossRef]
49. Abdusalomov, A.; Baratov, N.; Kutlimuratov, A.; Whangbo, T.K. An Improvement of the Fire Detection and Classification Method Using YOLOv3 for Surveillance Systems. *Sensors* **2021**, *21*, 6519. [CrossRef] [PubMed]
50. Mukhiddinov, M.; Abdusalomov, A.B.; Cho, J. Automatic Fire Detection and Notification System Based on Improved YOLOv4 for the Blind and Visually Impaired. *Sensors* **2022**, *22*, 3307. [CrossRef] [PubMed]
51. Abdusalomov, A.B.; Mukhiddinov, M.; Kutlimuratov, A.; Whangbo, T.K. Improved Real-Time Fire Warning System Based on Advanced Technologies for Visually Impaired People. *Sensors* **2022**, *22*, 7305. [CrossRef] [PubMed]
52. Abdusalomov, A.B.; Safarov, F.; Rakhimov, M.; Turaev, B.; Whangbo, T.K. Improved Feature Parameter Extraction from Speech Signals Using Machine Learning Algorithm. *Sensors* **2022**, *22*, 8122. [CrossRef]

53. Khan, F.; Tarimer, I.; Alwageed, H.S.; Karadağ, B.C.; Fayaz, M.; Abdusalomov, A.B.; Cho, Y.-I. Effect of Feature Selection on the Accuracy of Music Popularity Classification Using Machine Learning Algorithms. *Electronics* **2022**, *11*, 3518. [CrossRef]
54. Abdusalomov, A.; Whangbo, T.K.; Djuraev, O. A Review on various widely used shadow detection methods to identify a shadow from images. *Int. J. Sci. Res. Publ.* **2016**, *6*, 2250–3153.
55. Akmalbek, A.; Djurayev, A. Robust shadow removal technique for improving image enhancement based on segmentation method. *IOSR J. Electron. Commun. Eng.* **2016**, *11*, 17–21.
56. Farkhod, A.; Abdusalomov, A.B.; Mukhiddinov, M.; Cho, Y.-I. Development of Real-Time Landmark-Based Emotion Recognition CNN for Masked Faces. *Sensors* **2022**, *22*, 8704. [CrossRef]