**YekYiWei(p2107631)**

**Land Cover Classification**

**Using CNN ResNet and VGG19**

**ABSTRCT**

Fast development in Deep Learning and its hybrid methodologies has led diverse applications in different domains. For image classification tasks, Convolutional Neural Network (CNN) is mostly chosen for recent usages. This study employ two CNN architecture model Resnet and VGG19 to train the EuroSAT dataset which  consisits of 64x64 images captured by Sentinel-2A satellite and it has over **27000 images** spread across **10 classes**. Results showed VGG19 perform better than ResNet in Land Cover classification with the average accuarcy points 97.67%, the classifier was able to classify the majority of the images with high recognition percentage ranging from 93.53% to 99.84%.

Keywords

Convolutional Neural Network, ResNet, VGG19, Land Cover Classification

**1 INTRODUCTION**

Many existing land cover mapping products are manually created through visual interpretation, which takes advantage of human expertise in the labeling process. However, manual labeling may result in both false positives and false negatives due to observational mistakes. In addition, this approach usually requires multiple researchers to delineate land covers and substantial human resources make it infeasible for large regions or for a long period. The land cover prediction process also facing challenges because of the heterogeneity of data. For instance, the spectral features of land covers are different in different regions and can change over time. Such temporal variation is mainly caused by changes in temperature, sunlight and precipitation in different years, and potentially leads to misclassification. Therefore, to have more accurate and high-quality land images classification, the deep learning Convutional Neural Network (CNN) can perform the challenging task.

With the fast development and wide interest in Artificial Intelligence (AI), Machine Learning (ML), which is the sub-field of AI training computers to have the ability to learn without being explicitly programmed, has also been spotlighted with Deep Learning (DL). DL is part of ML where learning is more sophisticated to create human level intelligence. Deep Learning has highly influenced the field of computer vision when Convolutional Neural Networks (CNN) models were used in tasks like image classification, object detection, facial recognition etc. Convolutional Neural Network (CNN) extracts features using convolutional computation and is widely applied for image and video data. Compared to a regular neural network, CNN layers are organized in 3 dimensions: width, height, and depth. Convolutional neural network (CNN) is a type of artificial neural network commonly used and proven to be effective in the field of image recognition and classification [1]. The convolutional layer, pooling layer, and fully-connected layer, are the three main types of layers stacked together to form a CNN architecture.

This study aims to train and compare two classification models to learn the land cover patterns of given remote sensing image based on some well defined target class labels. For training the model we will use the EuroSAT dataset which  consisits of 64x64 images captured by Sentinel-2A satellite and it has over **27000 images** spread across **10 classes**.

2 RELATED WORK

2.1 Convolutional Neural Network (CNN)

The convolutional layer is the first layer in a convolutional neural network and is the core building block that does the most computations [2]. In this layer, array of weights known as filters convolves through the input image multiplying each value in the filter to all the original pixel values of the image. This process is repeated for every location on the input volume and acts as the feature extractor from the input image. The pooling layer then functions to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in order to control overfitting. The fully connected layer then gathers the output from the convolutional and pooling layer representing high-level features of the input image. LeCun created one of the earliest architecture of convolutional neural networks called the LeNet5 architecture. LeCun’s work was essential in using convolutions with learnable parameters as an effective way to extract similar features at multiple location with few parameters. His work was the basis for recent architectures to date [3]. Various convolutional neural network models have been developed such as AlexNet, Inception, Resnet, Desnet, MobileNet, GoogLeNet, VGGNet and etc to reduce the test error rates or increasing the accuracy in image classifications. ImageNet Large Scale Visual Recognition Challenge (ILSVRC) has contributed a lot of deeper CNN architectures with better results [4].

Among the CNN architectures, ResNet has achieved 3.6% top-5 error rate with 152 layers using residual learning to solve degrading problem of deep network [5]. Residual Network (ResNet) architecture is a type of artificial neural network that allows the model to skip layers without affecting performance. It’s become one of the most popular architectures for various computer vision tasks since ResNets demonstrated with experiments that they can now train a 1001-layer to outperform counterparts [6]. There are multiple versions of ResNetXX architectures where ‘XX’ denotes the number of layers. The most commonly used ones are ResNet50 and ResNet101 [7]. VGG19 is a variant of VGG model which in short consists of 19 layers (16 convolution layers, 3 Fully connected layer, 5 MaxPool layers and 1 SoftMax layer). VGG-19 is a convolutional neural network that is trained on more than a million images from the ImageNet database. The network is 19 layers deep and can classify images into 1000 object categories, such as a keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images [8]. Compared to VGGNets, ResNets are less complex since they have fewer filters.

2.2 Land Cover Classification

A group of researchers have performed some initial investigation into using ground-level photo collections for land cover [9] and land use classification [10], but there remains significant opportunity to expand upon this initial work. Land use classification using high-level image features extracted using CNNs from geolocated ground-level images. A significant result of the work is showing high-level, semantic image features extracted using pre-trained CNN models generalize well and provide practical implications for researchers wanting to extract geographic information from georeferenced ground-level images [11]. The study has achieved 76% accuracy on a challenging eight-class land use classification problem. There is lack of land cover classification related studies in literature.

**3. METHODOLOGY**

This project aims to carry out the land cover classification by applying two CNN architectures which are RestNet50 and VGG19

3. 1 Dataset and Features

The EuroSAT dataset is obtained from the webpage <https://github.com/phelber/eurosat>. It consists of Sentinel-2 satellite images covering 13 spectral bands and 10 classes with 27000 labeled and geo-referenced samples. Each image contains RGB values for each pixel, as well as Near Infrared channels. The RGB information is used to classify the images. The dataset is split into 10 classes of land cover. Each class varies in size. Figure 1 shows class distribution of the difference labels and example images. All the images are quite uniformly distributed among the class labels.



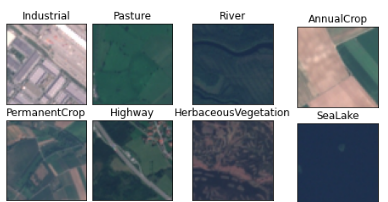


Figure 1. Class label distributions and example images

3.2 Data Preprocessing

A stratified shuffle-split using Scikit-learn is performed to maintain class proportions. 30% of the dataset will be held for evaluation purposes. Data is loaded into the Keras model using the ImageDataGenerator class. The images needed to be in their own respective land cover directories. After splitting the dataset, Some image augmentations are created using the generator and also denote a subset of the training data to be used as validation data during training.

3.3 Model Training

In this section, two CNN models which are RestNet50 and VGG19 are chosen to do the model training. The initial part of the CNN is freezed with imagenet weights and dense layers are trained with a high learning rate of 0.01 and in later part, the whole model end-to-end i.e. fine tune by keeping a small learning rate between 0.001 to 0.0001 are to be trainned. it must be noted that the different dense layers are setup for different CNN models.

**4. RESULTS and DISCUSSION**

4.1 ResNet50

Refer to Figure 2, plot of Validation loss is fluctuated at the first 5 epoch indicating the training dataset does not provide sufficient information to learn the problem, relative to the validation dataset used to evaluate it. However, closer gap between training accuracy and validation accuracy before 10 epoch.

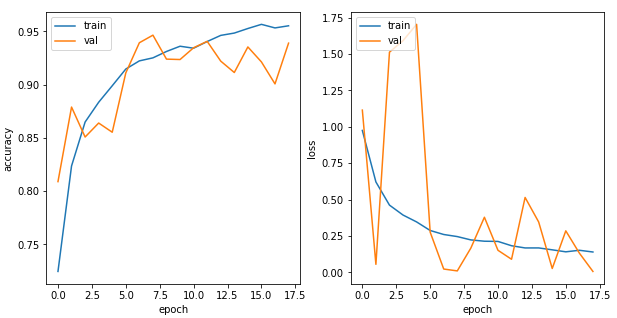


Figure 2. Plots of accuracy and loss for traning and validation of ResNet50 model

Table 1 displays that overall, the classifier was able to classify the majority of the images with high recognition percentage ranging from 90.94% (Highway) to 98.46%(SeaLake).

Table 1. ResNet50 classification performance result

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F-Score | Support |
| AnnualCrop | 0.846029 | 0.933810 | 0.887755 | 559.0 |
| Forest | 0.983471 | 0.983471 | 0.983471 | 605.0 |
| HerbaceousVegetation | 0.964419 | 0.895652 | 0.928765 | 575.0 |
| Highway | 0.909465 | 0.894737 | 0.902041 | 494.0 |
| Industrial | 0.976496 | 0.960084 | 0.968220 | 476.0 |
| Pasture | 0.956221 | 0.936795 | 0.946408 | 443.0 |
| PermanentCrop | 0.927152 | 0.878661 | 0.902256 | 478.0 |
| Residential | 0.951442 | 0.992089 | 0.971340 | 632.0 |
| River | 0.925403 | 0.925403 | 0.925403 | 496.0 |
| SeaLake | 0.984568 | 0.993769 | 0.989147 | 642.0 |

4.2 VGG19

The plots in Figure 3 shows that the training accuracy is closes to validation accuracy indicating the model is fiitting. Validation loss stays lower than training loss confirm the model is a good fit.

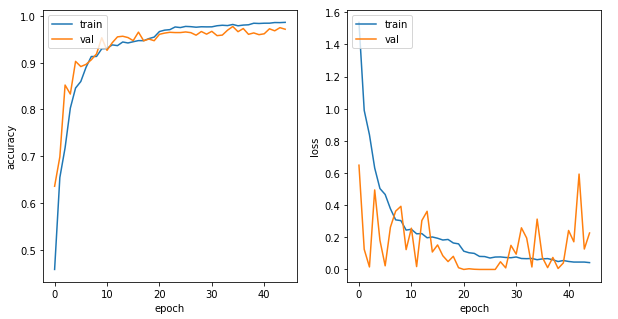


Figure 3. Plots of accuracy and loss for traning and validation VGG19 model

Table 2 shows that overall, the classifier was able to classify the majority of the images with high recognition percentage ranging from 93.53% (PermanentCrop) to 99.84%(SeaLake).

Table 2. VGG19 classification performance result

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F-Score | Support |
| AnnualCrop | 0.958559 | 0.951699 | 0.955117 | 559.0 |
| Forest | 0.990099 | 0.991736 | 0.990917 | 605.0 |
| HerbaceousVegetation | 0.985663 | 0.956522 | 0.970874 | 575.0 |
| Highway | 0.985801 | 0.983806 | 0.984802 | 494.0 |
| Industrial | 0.986842 | 0.945378 | 0.965665 | 476.0 |
| Pasture | 0.969027 | 0.988713 | 0.978771 | 443.0 |
| PermanentCrop | 0.935354 | 0.968619 | 0.951696 | 478.0 |
| Residential | 0.966309 | 0.998418 | 0.982101 | 632.0 |
| River | 0.985887 | 0.985887 | 0.985887 | 496.0 |
| SeaLake | 0.998428 | 0.989097 | 0.993740 | 642.0 |

Table 3 shows that between the two classifiers, VGG19 achieves the highee classification accuracy which 97.67% on test images.

Table 3. Global F2-scores and accuracy

|  |  |  |
| --- | --- | --- |
| Model | Global F2-scores | Accuracy |
| ResNet50 | 0.9426 | 0.9426 |
| VGG19 | 0.9767 | 0.9767 |

**5 CONCLUSION**

The deep learning methods have achieved great success in various images classification, three challenges remain. One, selecting appropriate network structures, parameters, and algorithms is critical to the success of deep learning approaches. Two, these methods often have millions of parameters and many layers, which are difficult to train, and sometimes several weeks are required to tune the parameters. Three, convolutional neural networks tend to cause overfitting with poor generalized capability because of the complicated mechanism. These challenges increase the difficulty of using deep learning algorithms in applications running on low computing power devices. The development of parallel computing and the popularization of graphics processing units (GPU) promoted the development of deep learning methods, especially convolutional neural networks, which have been amazingly successful in image recognition because of their outstanding capability to learn complex and robust feature representations. Future study can consider various convolutional neural network models, such as AlexNet, Inception, Resnet, Desnet, and MobileNet which have been proved in reducing the test error rates or increasing the accuracy in image classifications.

**REFERENCES**

[1] Karpathy, A. (n.d.). Convolutional Neural Networks for Visual Recognition. cs231n.github.io/convolutionalnetworks. Last accessed 11/11/2022

[2] Culurciello, E. (2017, April). The history of Neural Networks. Retrieved from Dataconomy: http://dataconomy.com/2017/04/history-neural-networks/. Last accessed 11/11/2021

[3] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998d). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 22782324.

[4] Sze, V., Chen, Y. H., Yang, T. J., and Emer, J. S. 2017. Efficient processing of deep neural networks: A tutorial and survey. IEEE. 105, 12 (2017), 2295–2329. DOI=http://dx.doi.org/10.1109/jproc.2017.2761740.

[5] Simonyan, K. and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. Unpublished results. (2014).

[6] Zijian Feng, Z. J. 2022. An Overview of ResNet Architecture and Its Variants. Retrrieved from <https://builtin.com/artificial-intelligence/resnet-architecture>. Last accessed 23/11/2022

[7] Aqeel Anwar. 2019. Difference between AlexNet, VGGNet, ResNet, and Inception. Retrrieved from <https://towardsdatascience.com/the-w3h-of-alexnet-vggnet-resnet-and-inception-7baaaecccc96>. Last accessed 24/11/2022

[8] Gaudenz, B. VGG-19 Convolutional Neural Network. Retrrieved from <https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks>. Last accessed 23/11/2022

[9] H. Oba et al. Towards Better Land Cover Classification using Geo-tagged Photographs. International Symposium on Multimedia, 2014.

[10] Leung, D and Newsam, S. Exploring Geotagged Images for Land-Use Classification. In ACM MM Workshop, 2012.

[11] Zhu. Y., and Newsam. S. 2015. Land Use Classification using Convolutional Neural Networks Applied to Ground-Level Images. DOI: http://dx.doi.org/10.1145/2820783.2820851