CS228 Fall 2023 Project Report

NIKI NEZAKATI, 862464096, nneza001@ucr.edu
MUBARIZUDDIN MOHAMMED, 862466269, mmoha065@ucr.edu
ASHKAN VEDADI GARGARY, 862352576, aveda002@ucr.edu

The code is available on GitHub at CS228-Final-Project [1]

Additional Key Words and Phrases: Deep Learning, Classification, Land Cover Classification, Image Processing

1 INTRODUCTION & MOTIVATION

Classifying land use and land cover based on satellite images is a challenging study, and we are going to address this challenge. For our purpose, we will use the Sentinel-2 satellite dataset. These images are freely available, span 13 spectral bands, and have been compiled into a new dataset. The dataset that we will use is known as EuroSAT. The motivation for using this dataset is to utilize the data for domains such as agriculture, disaster recovery, climate change, urban development, or environmental monitoring. However, the satellite images must first be processed and transformed into structured semantics. One type of such fundamental semantics is Land Use and Land Cover Classification. The performance of classification systems strongly depends on the availability of high-quality datasets with a suitable set of classes. We will use advanced deep learning techniques to analyze these images. It is crucial to have large quantities of training data available.

We are motivated by the significant impact this technology can have on various real-world applications. Here are some key areas where accurate land cover classification plays a crucial role:

- Environmental Monitoring
- Urban Planning and Development
- Agriculture and Forestry Management
- Disaster Management and Response

2 DATASET DESCRIPTION

In this project, we aim to utilize the dataset introduced in the 'EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification' paper[1]. This dataset is based on Sentinel-2 RGB channels satellite images and is designed for the purpose of land use and land cover classification[2]. The dataset consists of 27,000 labeled images covering 10 distinct land use and land cover categories. The following are the features of the described dataset:

(1) **Filename:** Name of the image file in the format of text.
(2) **Image:** 64x64 RGB image of the land.
(3) **Label:** Label of the image from the class labels. The 10 classes in the EuroSAT Dataset are as follows:
    - Annual Crop
    - Forest
    - Herbaceous Vegetation
    - Highway
    - Industrial

[1] https://github.com/mubarizzz/CS228-Final-Project

- Pasture
- Permanent Crop
- Residential
- River
- SeaLake

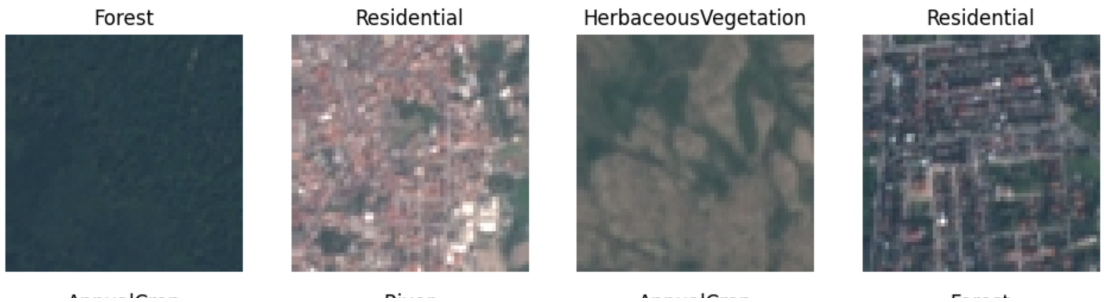

Fig. 1. Sample labeled images from the EuroSAT Dataset

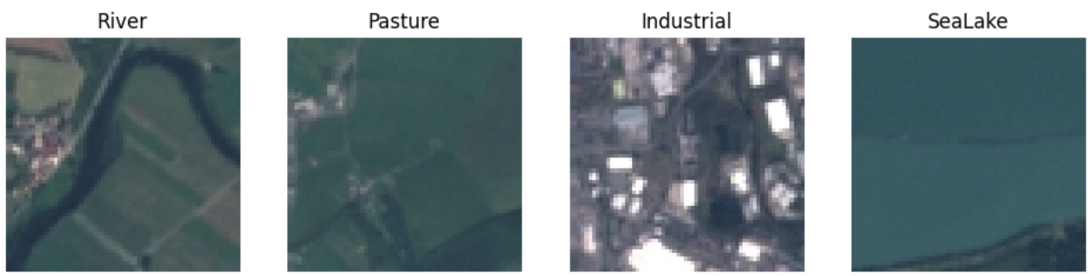

Fig. 2. Sample labeled images from the EuroSAT Dataset

3 DATA PRE-PROCESSING

Doing data pre-processing on this particular dataset was a crucial step in improving the accuracy of our models[3]. Some of the reasons that motivated us to do pre-processing are as follows:

- To prevent overfitting
- To improve robustness
- To address the class imbalance problem
- To enhance the generalizability of the model

First, we resize each image in the dataset to the desired input shape. Then, with a probability of 50%, we randomly apply a set of transformations to the image: left-right flip, up-down flip, brightness adjustment, saturation adjustment, and contrast adjustment[4]. These transformations help to augment the data and make the model more robust to variations in the input. Finally, we pre-process the input image using a specific method (preprocess_input) before feeding it to the model[5].

For the validation data, we simply resize the image to the desired input shape and pre-process it. We do not apply any random transformations to the validation data as we want to evaluate the model's performance on unseen data. Once the pre-processig is done and the data is augmented it should look something like this:

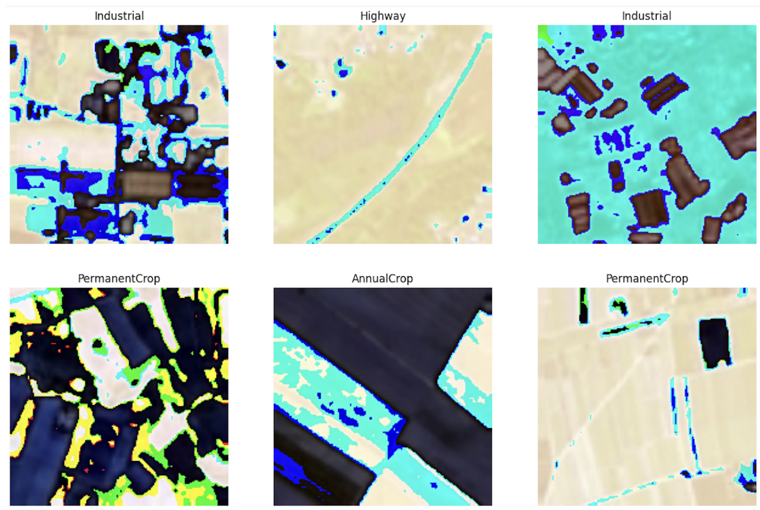

Fig. 3. Images after pre-processing (unit-8)

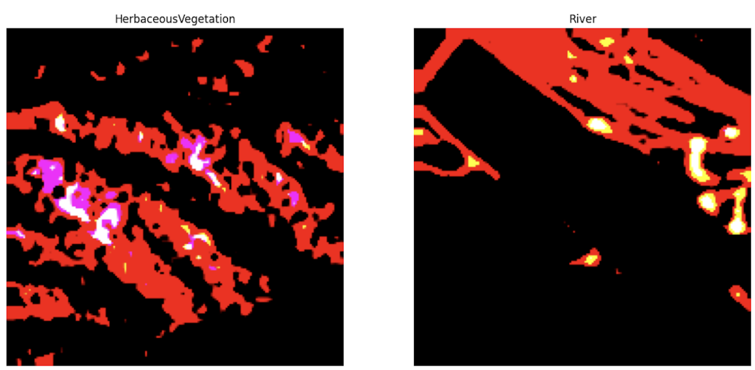

Fig. 4. Images after pre-processing (float-point number)

4 PROPOSED APPROACH

Currently, there are many existing methods for land cover classification tasks. Pre-trained CNNs like VGG16 and ResNet-50 [6] are among the most popular ones. In our case, the euroSTAT images have complex textures, which usually have high spatial variability and can be difficult to learn. For this, the solution is to use a patch-based approach. What this means is, dividing the image into small patches and then classifying each patch independently.

We have choose three different approaches: **ResNet**, **DenseNet**, and **VGG16**. These three differenet architectur are well-known for image processing problems (Figure 5).

To improve the existing model's performance we plan to use techniques like data augmentation [7]. This will help us improve the model's robustness and generalization ability. Taking a transfer learning approach would also lead to better overall performance in this case, since our dataset is relatively small.

In our proposed approach, we initiate the process by enhancing the dataset through data augmentation. As described in section 3, this involves various transformations such as flipping the image horizontally and vertically. Additionally, we adjust the image brightness, saturation, and contrast within specific ranges to diversify our dataset. Subsequently, we employ transfer learning using pre-trained models, namely ResNet50[8], DenseNet201[9], and VGG16[10], initially trained on ImageNet. We extend these models by incorporating a global average pooling layer, a dense layer with 512 units and ReLU activation, and a dense classification layer with softmax activation to tailor them to our dataset.

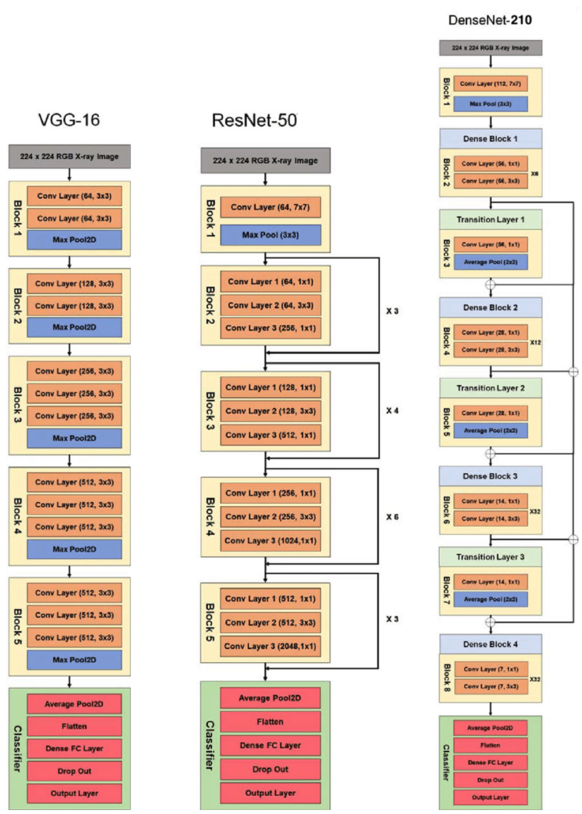

Fig. 5. Architecture of three different methods that we use for our project.

4.1 ResNet

ResNet, short for Residual Network, is a type of deep learning model that revolutionized the field of computer vision . ResNet addresses the problem of training very deep neural networks. As networks grow deeper, they often suffer from vanishing gradients and degradation problems, where adding more layers leads to higher training error. ResNet solves this by introducing “skip connections” or “residual connections” that allow gradients to flow through the network more effectively [8].

These connections essentially skip one or more layers and perform identity mapping, adding the input of the layer to its output. This innovative approach enables the training of networks that are much deeper than was previously possible, with architectures ranging from 18 to over a thousand layers. ResNet models have been highly successful, winning the ImageNet competition in 2015 and becoming a foundational architecture for many subsequent deep learning models. They are widely used in applications that require image recognition, object detection, and more, due to their ability to learn complex patterns and features from visual data.

4.2 DenseNet

DenseNet, which stands for Densely Connected Convolutional Network, is an innovative deep learning architecture that was introduced to address some of the challenges faced by traditional convolutional neural networks (CNNs). DenseNet’s key feature is its unique connectivity pattern: each layer is connected to every other layer in a feed-forward fashion1. This means that in a DenseNet, the output of each layer is used as input for all subsequent layers, creating a highly interconnected system [9].

This architecture has several advantages. It alleviates the vanishing gradient problem, strengthens feature propagation, encourages feature reuse, and significantly reduces the number of parameters compared to standard CNNs1. These benefits lead to improved efficiency and robustness in learning, making DenseNets particularly effective for tasks like image classification, object detection, and semantic segmentation2. The design of DenseNet ensures maximum information flow between layers in the network, which enables the training of deeper, more accurate models without a corresponding increase in computational complexity1. DenseNet has been a valuable contribution to the field of computer vision and continues to be a popular choice for various applications [9].

4.3 VGG16

VGG16 is a convolutional neural network (CNN) model that has become a classic in the field of computer vision due to its simplicity and deep architecture. Developed by Karen Simonyan and Andrew Zisserman from the University of Oxford, VGG16 stands out for its use of (3 X 3) convolutional filters with a stride of 1 and max pooling layers with a (2 X 2) filter of stride. This consistent use of small filters allows the network to have a depth of 16 layers with weights, which is where it gets its name from.

The architecture of VGG16 is designed to increase the depth while maintaining a relatively small number of hyper-parameters. With approximately 138 million parameters, VGG16 is a large network capable of capturing complex features from images1. It was one of the top performers in the ILSVRC-2014 competition, achieving 92.7% top-5 test accuracy on the ImageNet dataset, which contains over 14 million images across 1000 classes. VGG16’s structure has inspired many subsequent neural network designs and remains a popular choice for image recognition tasks and feature extraction in various applications.

Our training strategy involves 5 epochs, utilizing a batch size of 128. For optimization, we employ the Adam optimizer, Sparse Categorical Crossentropy as the loss function, and accuracy as the evaluation metric. To ensure a robust model, we split our dataset into three subsets: 70% for training, 20% for evaluation, and 10% for testing. This partitioning allows us to train, fine-tune, and assess the performance of our models effectively.

5 RESULTS & DISCUSSION

5.1 Previous Results

Models	Accuracy
Fully convolutional networks (FCNs)	87.4%
Long short-term memory (LSTM)	86.5%
Convolutional neural networks (CNNs)	91.9%

Table 1. Comparison of previous results for Land Cover Classification on EuroSat Dataset.

The authors in [11] utilized a patch-based approach for training their CNNs. This approach divides the images into smaller patches for classification, which can lead to loss of spatial context and hinder the model's ability to capture relationships between neighboring pixels[12]. They also failed to address various problems presented by the EuroSAT Dataset like, Limited Data Size and the class imbalance. It is also unclear if the authors employed sufficient data augmentation, which could have contributed to the low accuracy.

5.2 Our Results

Models	Train Accuracy	Test Accuracy	Validation Accuracy
ResNet50	93.0%	93.4%	92.8%
DenseNet201	92.5%	91.8%	90.7%
VGG16	90.4%	91.7%	90.5%

Table 2. Comparison of our results for Land Cover Classification on EuroSat Dataset.

We observed that in our project all three models achieved high overall accuracies, exceeding 90%. This demonstrates the effectiveness of deep learning models for land cover classification tasks. The ResNet50 model outperformed the results of the previous works on this dataset, and our other two models by a small margin, suggesting its superior ability to learn relevant features from the EuroSAT images[13]. ResNet50's residual connections might have helped it learn complex relationships between pixels and capture spatial context more effectively than VGG16 and DenseNet201.

We have also included the results of our experiments without employing data augmentation to demonstrate how this approach enhances our models' performance.

Models	Train Accuracy	Test Accuracy	Validation Accuracy
ResNet50	77.8%	79.3%	79.1%
DenseNet201	72.6%	76.8%	77.6%
VGG16	71.9%	73.7%	74.2%

Table 3. Comparison of our results for Land Cover Classification on EuroSat Dataset before Pre-Processing.

The performance results of the models are presented in the following part:

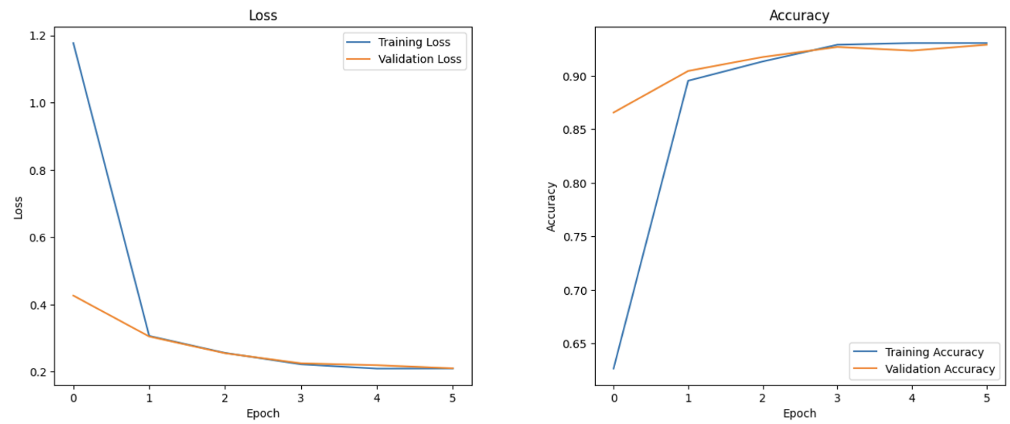

Fig. 6. Loss Vs. Epoch and Accuracy Vs. Epoch graphs for RESNET50

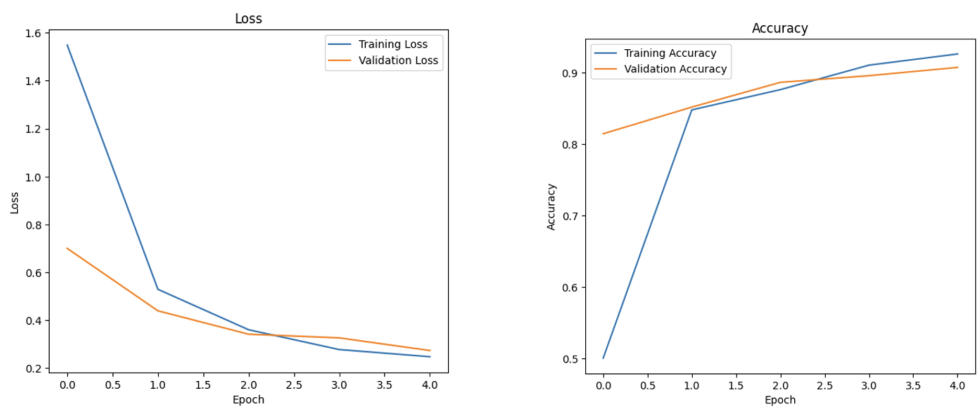

Fig. 7. Loss Vs. Epoch and Accuracy Vs. Epoch graphs for DENSENET201

5.3 Discussion

In the context where all parameters remained consistent for ResNet50, DenseNet201, and VGG16 on the EuroSAT dataset, the superior performance of ResNet50, with a test accuracy of 93.4%, can be due to its intricate architecture featuring residual connections. The residual connections facilitate the flow of information through the network, allowing for the effective capture of complex hierarchical features in images. Following closely, DenseNet201 achieves a test accuracy of 91.8%, leveraging its dense connections that promote the reuse of features across layers. Finally, VGG16, with a test accuracy of 91.7%, exhibits relatively lower performance, likely attributable to its simpler architecture and the absence of skip connections. This architectural simplicity renders VGG16 more prone to overfitting and less capable at capturing complex patterns compared to ResNet50 and DenseNet201 in this specific scenario.

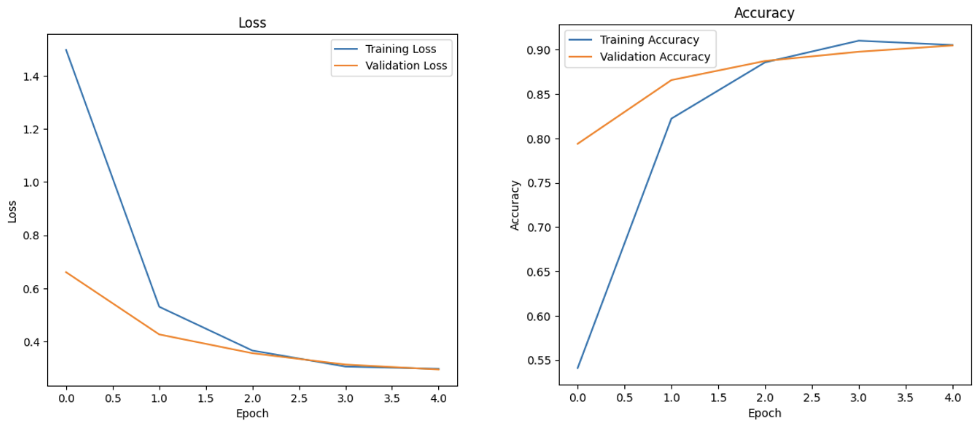

Fig. 8. Loss Vs. Epoch and Accuracy Vs. Epoch graphs for VGG16

It's also worth noting that our performance results are constrained by the computational limitations in our setting, and extending the training duration for more epochs holds the potential to further enhance model performance.

Fig. 9. ResNet50 evaluation results visualization

6 EVALUATION

We thoroughly assessed our most effective model by subjecting it to evaluation using the test set. To enhance our understanding and facilitate a more insightful analysis, we opted to visualize the model's outputs. The visual representations are provided in figure 9. This approach allows for a clearer and more comprehensive examination of the model's performance, aiding in the interpretation of results and insights derived from the evaluation process.

7 CONCLUSION

Working on land cover images derived from satellite imagery is a challenging yet valuable study. Obtaining results would enable us to classify different types of land and make plans for future purposes and industries. We have experimented with three different transfer learning approaches: ResNet, VGG16, and DenseNet for our models. To achieve better performance, we conducted some pre-processing, such as data augmentation, to make our data suitable for our classification problem. Ultimately, we demonstrate that our model outperforms the currently proposed models, even with fewer epochs. Furthermore, we show that the effect of data augmentation and pre-processing is significant for our model and greatly assists in achieving the best possible results. In summary, we introduce the best implementations and models for landscape classification problems on the EuroSat dataset.

REFERENCES

[1] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," 2018.
[2] L. D. Weng Q, Yang C, "A synthesis of land use/land cover studies: Definitions, classification systems, meta-studies, challenges and knowledge gaps on a global landscape," 2020.
[3] G. L. Li Y, Niu L, "Land cover classification of remote sensing images based on hierarchical convolutional recurrent neural network," 2022.
[4] L. C. Chen Y, Li Y, "Land cover classification in an era of big and open data: Optimizing localized implementation and training data selection to improve mapping outcomes," 2021.
[5] L. Y. Mou L, Muhammad S, "Land-use and land-cover classification in semi-arid areas from medium-resolution remote-sensing imagery: A deep learning approach," 2021.
[6] S. L. S. Y. Qikai Lu, Huanfeng Shen and L. Zhang, "Land-cover classification with high-resolution remote sensing images using transferable deep models," 2018.
[7] S. C. Ava Vali and M. Matteucci, "Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review," 2020.
[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
[9] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4700–4708, 2017.
[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[11] A. M. Ghayour B, Bahauddin K, "Classification of land use/land cover using artificial intelligence (ann-rf)," 2022.
[12] W. J. Zhang X, Li X, "Cost-effective land cover classification for remote sensing images," 2022.
[13] F. Y. Xu Y, Yang J, "Deep learning for land-use/land-cover classification: A meta-analysis and systematic review," 2022.