University of Science and Technology of Hanoi
Information Communication Technology Department

Introduction to Deep Learning

Midterm Report

Study the ResNet-D for remote sensing image classification

Group 23

Ho Thanh Thuy Tien	22BI13419
Le Anh Quang	22BI13380
Do Minh Quang	22BI13379
Vu Ha Vy	22BI13485
Tran Bao Nguyen	22BI13342
Nguyen Thu Trang	22BI13426

Sep 27, 2024

Contents

1 Abstract

Traditional algorithms are becoming increasingly ineffective in managing the rapid growth of remote sensing image data and resolution. Therefore, much of the recent progress in remote sensing image classification research has been largely driven by refinements in training procedures, such as improvements in data augmentation and optimization methods. In this paper, we explore the application of ResNet-D, a modified version of the standard Residual Network (ResNet) [9] architecture, for remote sensing tasks. ResNet-D has shown promising results in this domain by utilizing dilated convolutions and optimized downsampling, enabling more accurate classification of land cover types. Using the EuroSat dataset, we demonstrate that ResNet-D achieves a classification accuracy of 95.58%, outperforming traditional convolutional networks in identifying distinct land cover types such as urban areas, forests, and water bodies. This report details the model architecture, training methodology, and experimental results of ResNet-D in classifying remote sensing data.

2 Introduction

Convolutional neural networks (CNNs) have archieved remarkable success in numerous challenging data processing tasks, particularly in image classification [1]. However, as network architectures become deeper, they tend to encounter the vanishing gradient [2] problem. This issue occurs when the gradients used to update the networks' parameters become to small, affect the model's ability to learn effectively. Residual Networks (ResNet) were introduced to address this problem by skipping connections, enabling the training of deeper networks.

In recent years, CNNs have also shown significant potential in remote sensing applications, particularly for Land Use and Land Cover (LULC) classification from satellite imagery. However, despite their success, our theoretical understanding of how deep neural networks (DNNs) learn spatial and spectral features from such data remains limited. In this study, we aim to evaluate the effectiveness of **ResNet-D** [10], a modified version of ResNet, specifically designed for better feature extraction in complex image data.

To achieve this, we adopt a transfer learning approach, using a pretrained ResNet-D architecture and fine-tuning it with the EuroSAT dataset, a collection of labeled satellite patch images extracted from the Copernicus Sentinel-2 satellites products. The classification task is to learn a function that outputs $y \in \mathbb{R}^n$, where each component y_j represents the probability that an image $x \in D$ belongs to Class j. In this study, we aim to evaluate how well the model performs in this classification task. [11]

3 Methodology

3.1 Dataset

For this project, we used the EuroSAT [8] dataset, which contains over 27,000 labeled satellite images acquired from the Sentinel-2 satellite, provides both RGB and multi-spectral imagery. The images are 64*64 pixels with a Ground Sampling Distance of 10m, covering 13 spectral bands and consisting out of 10 classes:

- Industrial Buildings
- Residential Buildings
- Annual Crop
- Permanent Crop
- River
- Sea and Lake
- Herbaceous Vegetation
- Highway
- Pasture
- Forest

This dataset is divided into 70% training, 20% validation and 10% test sets. [7]

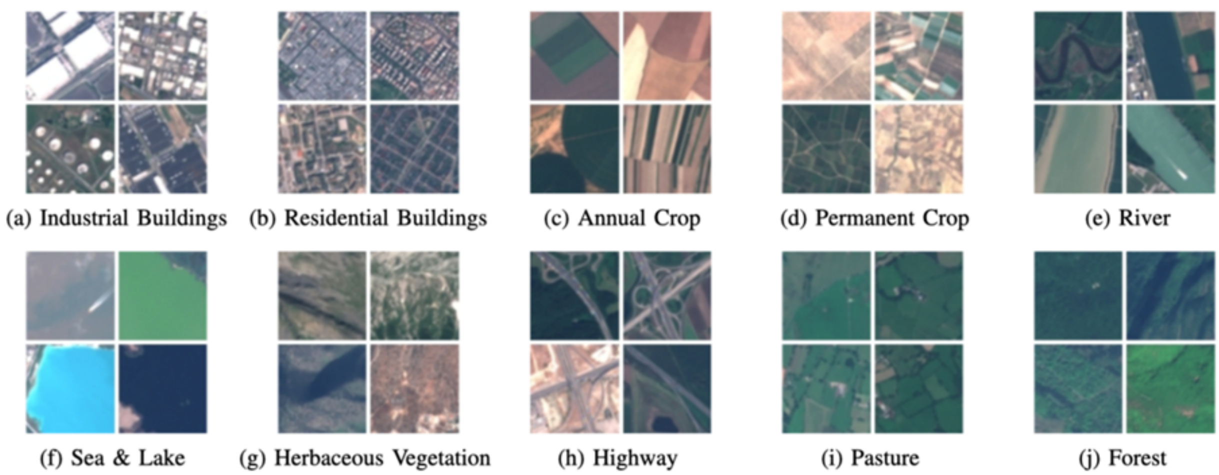


Figure 1: Sample image patches of all 10 classes covered in the proposed EuroSAT dataset

3.2 Data Preprocessing

Since ResNet-D has a fixed input size, all images must changed to 224*224 pixels. The images were then converted from Python Imaging Library format to PyTorch tensors, which are multi-dimensional arrays used for input into the neural network that allows efficiency data handling, and the pixel values of the images were normalized using pre-defined mean and standard deviation values for faster convergence and ensuring that each channel (red, green, blue) has similar distribution characteristics.

3.3 ResNet-D Architecture

3.3.1 ResNet

ResNet is a deep learning architecture that addresses the vanishing gradient problem commonly encountered in training very deep neural networks for tasks such as image classification. The key innovation of ResNet is the introduction of residual blocks, which using skip connections to pass by one or more layers. These skip connections enable the model to calculate gradients more effectively during backpropagation, and preventing them from diminishing as the network deepens. [9]

A typical ResNet block consists of several components: convolution layers to extract spatial features, batch normalization [3] to stabilize and accelerate training, and the ReLu Activation function. The skip connection adds the input of the block directly to its output, which effectively creates a shortcut, allowing the model to learn residual mappings rather than the full transformation. This approach enables ResNet models to be much deeper (with variants such as ResNet-50 [10], ResNet-101 [10], and ResNet-152 [10]) while maintaining good performance and solving the gradient vanishing issue.

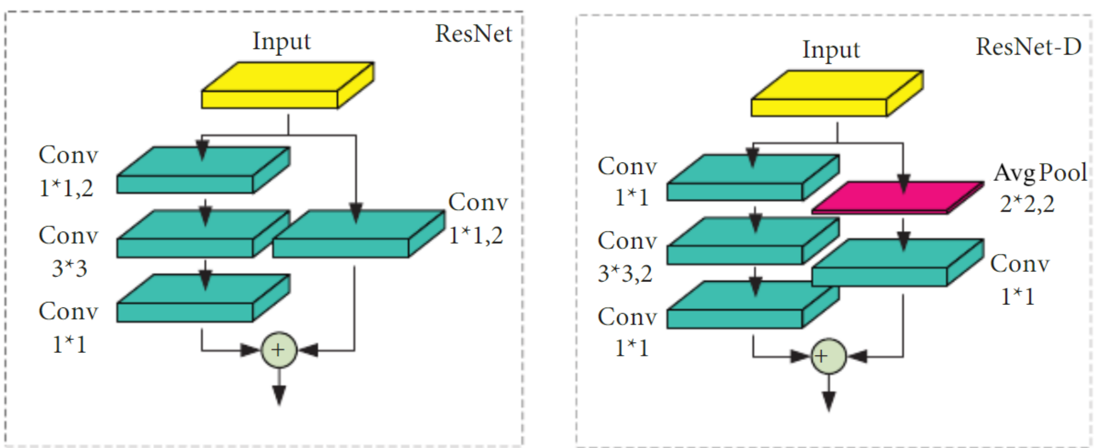


Figure 2: Comparison of ResNet and ResNet-D Architectures

3.3.2 ResNet-D

ResNet-D [4] builds on the ResNet architecture, improving its ability to preserve information during downsampling. While traditional ResNet only uses 1*1 convolutions for downsampling, ResNet-D adds an average pooling layer which helps retain more spatial information, crucial for high-resolution images like satellite imagery [6]. This adjustment helps prevent information loss during the reduction in feature map size.

Additionally, ResNet-D incorporates dilated convolutions to expand the network's receptive field without increasing parameters, allowing it to capture a wider range of spatial information. This makes ResNet-D particularly effective for large-scale images, such as those used in Land Use and Land Cover (LULC) classification, where both local and global context are important.

ResNet-D also uses bottleneck residual blocks. These consist of three layers: a 1*1 convolution for dimensionality reduction, a 3*3 convolution for feature extraction, and a final 1*1 convolution to restore dimensions. This design reduces the number of parameters while maintaining model depth, ensuring efficiency and accuracy in deep networks.

4 Model Training

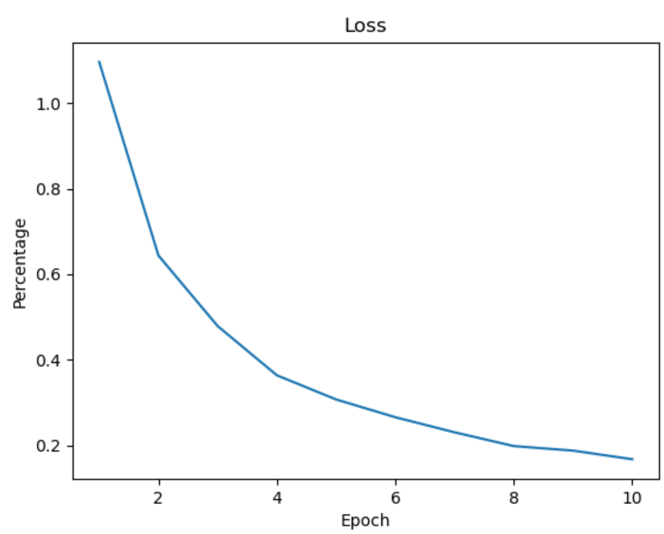
4.1 Evaluation Metrics

Accuracy was used to assess the outcomes of this experiment, where Accuracy is the proportion of correctly predicted outcomes across all samples.

$$\text{Accuracy} = \frac{\text{number of correctly classified samples}}{\text{total number of samples}}$$

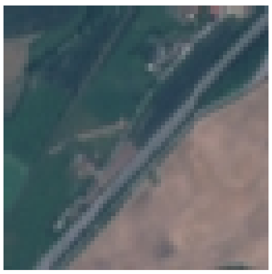
We are able to determine the accurate rate of image classification using this evaluation metric. [5]

In this work, we set the number of epochs to 10, which achieves a good balance between accuracy and computational efficiency. This approach resulted in an accuracy of nearly 96%. The exact value may vary slightly each time we fine-tune the model. The data is presented in the graph below.



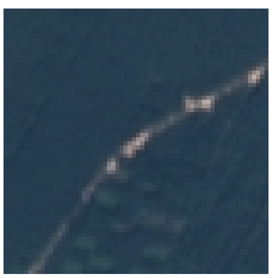
4.2 Results

Once we have achieved the desired accuracy, we can apply our fine-tuned model to new images, such as those extracted from Sentinel-2 products available through Sentinel-Hub. For this experiment, we extracted several patch images of the same size as those used for fine-tuning the model (224*224 pixels). To test how well the model performs on this task, we selected two sample images from two different labels, or landscapes. The first image shows a highway cutting through an agricultural area, and we will analyze how the model predicts it.



- **Predicted Label:** Highway
- **Actual Label:** Highway

The second image depicts a segment of a forest, taking from Forest euroSAT dataset.



- **Predicted Label:** Forest
- **Actual Label:** Forest

4.3 Comparison

We evaluate ResNet-50 with its three variations, using the ImageNet and EuroSat datasets, with the batch size of 32. The results are presented in Table 1.

Model	Dataset	Accuracy (%)
ResNet-50	ImageNet	76.21
ResNet-50-B		76.66
ResNet-50-C		76.87
ResNet-50-D		77.16
ResNet-50-D	EuroSAT	95.58

Table 1: Comparison of ResNet-50 model variations on ImageNet and EuroSAT datasets.

The table highlights several improvements in performance on the ImageNet dataset, with ResNet-50-B and ResNet-50-C achieving increases of 0.45% and 0.21%, respectively. Notably, ResNet-50-D achieves the highest accuracy on ImageNet at 77.16%. However, its most significant performance is observed on the EuroSat dataset, where it reaches an accuracy of 95.58%, surpassing the other models in this evaluation. These results indicate that ResNet-50-D demonstrates strong generalization capabilities, particularly in the context of remote sensing applications. Overall, its performance positions ResNet-50-D as a highly effective model for Land Use and Land Cover classification tasks.

5 Conclusion

This project demonstrated the effectiveness of the ResNet-D architecture for Land Use and Land Cover (LULC) classification using satellite imagery from the EuroSAT dataset. We have fine-tuned a pretrained ResNet-D model, achieving strong accuracy in identifying various land cover types, such as forests, water bodies, and urban areas. ResNet-D's use of dilated convolutions and optimized downsampling enhanced its ability to capture essential spatial and spectral features from the satellite images.

The model's performance shows its capability in handling complex remote sensing data, giving better performance than traditional convolutional networks. Transfer learning further reduced training time while maintaining high accuracy.

Although the results are promising, future work could explore more diverse datasets and additional fine-tuning methods to improve performance. Overall, ResNet-D has proven to be a valuable model for remote sensing image classification and holds significant potential for broader applications in environmental and land-use analysis.

References

[1] I. Sutskever A. Krizhevsky and G. Hinton. Imagenet classification with deep convolutional neural networks, 2012. pages 3

[2] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks., 2010. pages 3

[3] S. Ren K. He, X. Zhang and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, 2015. pages 5

[4] Tong He Zhi Zhang Hang Zhang Zhongyue Zhang Junyuan Xie Mu Li. Bag of tricks for image classification with convolutional neural networks. `https://arxiv.org/pdf/1812.01187v2`, 2018. pages 5

[5] Xiaoye Wang Ruixing Wang Hao Chen Yinhai Zhu Liyuan Wang, Yulong Chen. Research on remote sensing image classification based on transfer learning and data augmentation. `https://link.springer.com/chapter/10.1007/978-3-031-40292-0_9`, 2023. pages 6

[6] Bushra Zafar Amsa Shabbir Nouman Ali Maryam Mehmood, Ahsan Shahzad. Remote sensing image classification: A comprehensive review and applications, 2022. pages 5

[7] Andreas.Dengel Damian.Borth Patrick.Helber, Benjamin.Bischke. Eurosat. `https://github.com/phelber/EuroSAT`, 2019. pages 4

[8] Andreas.Dengel Damian.Borth Patrick.Helber, Benjamin.Bischke. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. `https://arxiv.org/pdf/1709.00029`, 2019. pages 4

[9] Pablo Ruiz. Understanding and visualizing resnets. `https://towardsdatascience.com/understanding-and-visualizing-resnets-442284831be8`, 2018. pages 3, 5

[10] Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun. Deep residual learning for image recognition, 2015. pages 3, 5

[11] Linan Zhang and Hayden Schaeffer. Forward stability of resnet and its variants. `https://arxiv.org/pdf/1811.09885`, 2018. pages 3