Deep Learning Project Report 2023

Md Moinul Islam
University of Oulu
Oulu, Finland
moinul.islam@student.oulu.f

Efta Khairul Bashar
University of Oulu
Oulu, Finland
ebashar23@student.oulu.fi

Taufiq Ahmed
University of Oul
Oulu, Finland
taufiq.ahmed@student.oulu.fi

Abstract

This study investigates the transferability and efficiency of various deep neural network architectures, specifically ResNet18, VGG, and Vision Transformer (ViT). The primary focus of the project is to train these models on the MiniImageNet dataset and then fine-tune them using the EuroSAT dataset. This approach is crucial for understanding how well the models can adapt and learn from different data environments, particularly in few-shot learning scenarios. The methodology involves a two-phased approach: initial training on MiniImageNet and subsequent fine-tuning on carefully selected samples from EuroSAT. The results indicate a diverse range of performances among the models, emphasizing the impact of architectural differences on the effectiveness of transfer learning. These findings have significant implications for image classification applications, especially in situations with limited data availability.

1. Introduction

A key component of machine learning, deep learning has enabled systems to independently decipher complicated patterns from enormous datasets, propelling the field to previously unheard-of heights. Image classification in this paradigm is a demonstration of the power of deep learning algorithms. But as these models' capacities increase, so do the difficulties they encounter, especially in situations where data is scarce. The importance of obtaining hierarchical representations from data in the vast field of deep learning cannot be emphasized. It allows systems to recognize complex features, making tasks like image classification much more accurate. However, the effectiveness of simple deep learning techniques faces a significant obstacle in situations where there are few labeled samples. In fields such as remote sensing applications, where it is frequently impractical or prohibitively expensive to acquire large datasets, conventional methods are unable to achieve the required performance. In order to address the issues raised by sparse data, the adaptive paradigm of transfer learning is explored in depth in this paper. The foundation of transfer learning is the idea of using pretrained models' fine-tuned knowledge from larger datasets to improve model performance in domains with smaller sample sizes. In this work, the EuroSAT dataset—selected for its applicability to remote sensing applications—serves as a concrete backdrop for a more general investigation of transfer learning principles. Transfer learning can be thought of as a tactical solution to the drawbacks of traditional deep learning when large amounts of labeled data are considered a luxury. Its fundamental tenet is the transferability of knowledge from one domain to another, which reduces the requirement for a large amount of data collection in the target domain. Within the present study, the transfer learning methodology entails pretraining models on the MiniImagenet dataset, which is widely recognized for its richness and diversity, and then refining them on the EuroSAT dataset. This work is novel not only because it applies transfer learning to the EuroSAT dataset but also because it compares and contrasts different pretrained models. The research goes beyond ResNet and includes models such as VGG16, ResNet18 and Vision Transformer. The research is made more complex and nuanced by the diversification of models, which enhances our knowledge of how various architectures react to transfer learning within the constraints of small sample sizes. The nuances of this methodology will be examined in the following sections, which will also highlight the painstaking steps that were taken in pre-training, fine-tuning, and comparative analysis. The results show promise for improving image classification accuracy on EuroSAT as well as for providing insights into the robustness and adaptability of transfer learning on a variety of deep learning architectures. The link of the code is as follows:

Deep Learning Project 2023.

1.1. Contributions

2. Approach

The approach section functions as an essential preface, giving readers a thorough grasp of the purpose and setting

Name	Code	Report
Md Moinul Islam	Transfer Learning for MiniImageNet	Drafting Overall Methodology
Taufiq Ahmed	Vision Transformer	Drafting Introduction, Approach
Efta Khairul Bashar	Transfer Learning for EuroSAT	Drafting Results

Table 1. Contributions

of the study project. To improve readability and understanding, a pictorial summary diagram is utilized to concisely summarize the methodology.

2.1. Overview Diagram

A visual overview diagram is created by condensing the research methodology into an understandable and succinct representation. This visual aid serves as a roadmap, explaining the important phases and how they relate to each other throughout the project. With the aid of a visual guide, readers are able to quickly understand the study's overall structure, which makes it easier for them to move into the in-depth investigation that follows.

2.2. Nature of the project

The focus of this research project is to address the difficulties associated with sparse data when classifying images using the EuroSAT dataset. The main focus is on the use of transfer learning, a calculated tactic that improves classification performance in situations with limited sample sizes by utilizing the knowledge obtained from pretrained models.

2.3. Contextualization

The context of the project falls into the deep learning space, where traditional methods become unworkable when there is not enough labeled data. Specifically, applications related to remote sensing, which are typified by a lack of large-scale datasets, offer an appropriate context for the use of transfer learning methodologies. The objective is to modify pretrained models to fit the distinct characteristics of the EuroSAT dataset, which is designed for imagery from remote sensing systems.

2.4. Key Methodological Steps

The method includes pretraining the model, fine-tuning it, and preparing the data, among other important steps. The source domain used to pretrain a variety of models, including ResNet18, VGG16, and Vision Transformer, is MiniImagenet. After pretraining, a fine-tuning procedure applies the learned information to the EuroSAT dataset. The method is made more precise by carefully incorporating mathematical formulations

3. Experiments

Here, we undertake a thorough investigation of the trial stage with the goal of offering a detailed comprehension of the datasets, training procedures, outcomes, and the ensuing in-depth analysis. The main objective is to shed light on the empirical efforts made to determine whether transfer learning works well on the EuroSAT dataset.

3.1. Datasets

Careful dataset curation is essential to our experimentation. MiniImagenet and EuroSAT are our main sources of data. For initial model pretraining, MiniImagenet, a diverse image repository, serves as the source domain. The target domain for fine-tuning and further assessment is EuroSAT, which was developed specifically for remote sensing applications. A thorough investigation of the dynamics of transfer learning is made possible by the meticulous selection of these datasets. [1] [5] [4]

3.2. Working Machanism of Pre-trained model

3.2.1 ResNet18

ResNet18 transforms the training of deep neural networks with its novel architecture that makes use of residual blocks. Its capacity to pick up residual functions is fundamental to how it functions. Conventional deep networks encounter difficulties like the vanishing gradient problem, which makes learning in extremely deep structures difficult. ResNet18 addresses this problem by adding lingering connections between its blocks.

By acting as short cuts, these residual connections enable the model to learn the residual—that is, the difference between the input and output—instead of trying to learn the complete mapping. This new method allows training deeper networks without running into problems with diminishing gradients. ResNet18 uses these residual connections to extract complex features during pretraining on the MiniImagenet dataset. By effectively employing these short cuts, the model improves its comprehension of intricate patterns and subtleties present in the input data. Basically, ResNet18's architecture allows it to address depth-related problems, allowing for efficient transfer learning and modification to the EuroSAT dataset for reliable image classification in remote sensing applications. [3]

3.2.2 VGG16

The Visual Geometry Group model, or VGG16, sets itself apart with a deep network architecture that is simple and consistent. Small receptive fields in convolutional layers are used by VGG16 to capture complex and hierarchical features in images. This is how the algorithm operates. A fine-grained analysis of image content is ensured by stacking multiple convolutional layers with 3x3 receptive fields to create the architecture's depth.

VGG16 shows remarkable proficiency in identifying intricate patterns within images during pretraining on the MiniImagenet dataset. Convolutional layers distinguish intricate visual hierarchies by acting as feature extractors. The learned representations are gradually improved upon by each network layer, leading to a more thorough comprehension of the input data. With a focus on fine-grained feature learning, VGG16 is ready for later tasks, like transfer learning using the EuroSAT dataset.

The model can effectively capture both global and local features thanks to the small filters used in the design of VGG16's convolutional layers. The foundation of VGG16's flexibility is laid by this painstaking feature extraction, which also highlights the technology's suitability for image classification tasks in remote sensing applications, where complex and hierarchical features are essential for precise interpretation and classification. [6]

3.2.3 Vision Transformer

Due to its reliance on self-attention mechanisms, the Vision Transformer's (ViT) operating mechanism signifies a paradigm shift in image processing. In contrast to conventional convolutional neural networks (CNNs), ViT uses a transformer architecture that was first intended for sequential data instead of convolutional layers. ViT is able to extract global dependencies from images thanks to this unique methodology.

ViT uses self-attention mechanisms so that distinct areas of the image can be focused on by each position in the input sequence. The model gains the ability to dynamically attend to pertinent features during pretraining on the MiniImagenet dataset, which promotes a comprehensive comprehension of the image content. The extraction of local and global contextual information—both necessary for subtle pattern recognition—is made easier by this attention mechanism.

The capacity of the model to focus on various areas of an image improves its flexibility in capturing complex details and relationships, which are important for tasks such as image classification. Pretraining on MiniImagenet teaches attention patterns that lay the groundwork for efficient knowledge transfer to the EuroSAT dataset. ViT's operation thus demonstrates its potential in remote sensing applications, where precise and context-aware classification relies on a thorough understanding of global dependencies within images. [2]

3.3. Training Setup

Our experiments have two main steps, firstly we trained our models with the MiniImageNet dataset. We used pretrained models of ResNet18, VGG16 and Vision Transformer, which were trained on Imagenet. Then we used transfer learning for MiniImagenet. We used only the train.tar file for training and testing purposes. We split the dataset into 80% for training and 20% for test and validation. All images were resized to 224*224 pixels.

We used different combination of learning rates and optimizers to find the best combination. On second stage we used transfer learning using the EuroSAT-RGB dataset. We separated 5 classes from the original dataset then we split the selected dataset as instructed.

3.4. Results and Analysis

In the first stage, we used several combinations of learning rates and optimizers for each training.

We found the best results (in terms of accuracy and loss) for each model as shown in Table 2.

In the second stage, we used these combinations for corresponding models to perform transfer learning using the EuroSat-RGB dataset. Results from those trainings are shown in Table 3.

Form the results it is clear that the two dataset we used are from different domains. So even the models are performing good on the first dataset (MiniImagenet), they are not performing well on the second data-set (EuroSAT). The training and testing results with respect to training and testing epochs are shown in Figures 1, 2 and 3.

4. Conclusion

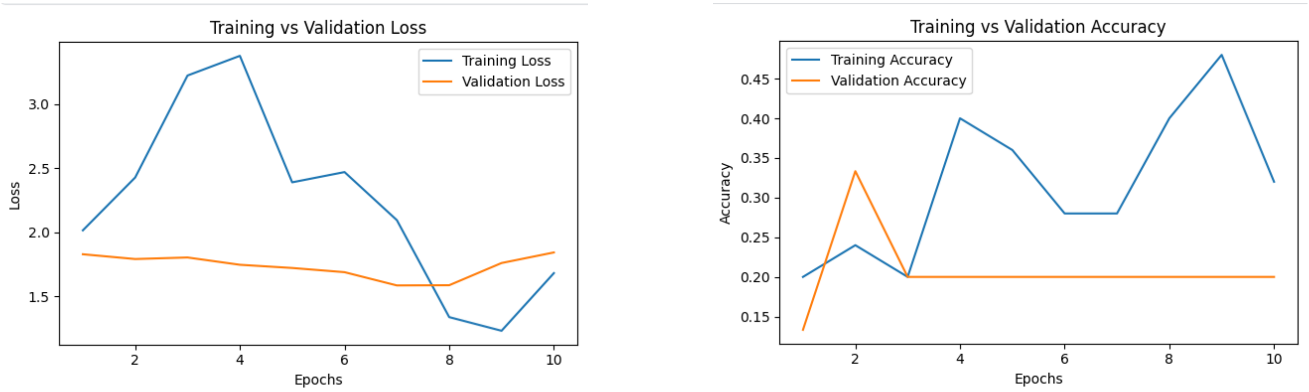
This report offers a concise examination of the application of deep learning models in image classification, utilizing the Mini-ImageNet and EuroSAT datasets. It outlines the setup of the environment, data preparation, and model initialization using architectures such as ResNet, VGG, and Vision Transformer. Hyperparameter tuning was systematically performed using Optuna to find optimal learning rates and optimizers. The core of the study involved an iterative process of fine-tuning and validation, where models pretrained on Mini-ImageNet were adapted to EuroSAT, guided by validation accuracy for model refinement. A comparative analysis of different architectures provided insights into their generalization capabilities. In the final phase, the best models were trained with optimal parameters, achieving notable accuracy and highlighting the iterative nature of machine learning workflows. The project

Methods	Optimizer	Best Parameters	Validation Loss	Validation Accuracy	Testing Loss	Testing Accuracy
ResNet18	Adam	LR = 10^{-4}, weight_decay = 0.1	0.7701	0.7875	0.8122	0.7765
VGG16	RMSprop	LR = 10^{-4}	1.5487	0.6049	1.5689	0.5948
Vision Transformer	RMSprop	LR = 10^{-5}	**0.3482**	**0.9380**	**0.3811**	**0.9315**

Table 2. Performace using **MiniImageNet**

Methods	Validation Loss	Validation Accuracy	Testing Loss	Testing Accuracy
ResNet18	1.8434	0.2000	1.8065	0.2000
VGG16	1.6359	0.2000	1.6494	0.1500
Vision Transformer	**0.6259**	**0.8000**	**0.6483**	**0.7833**

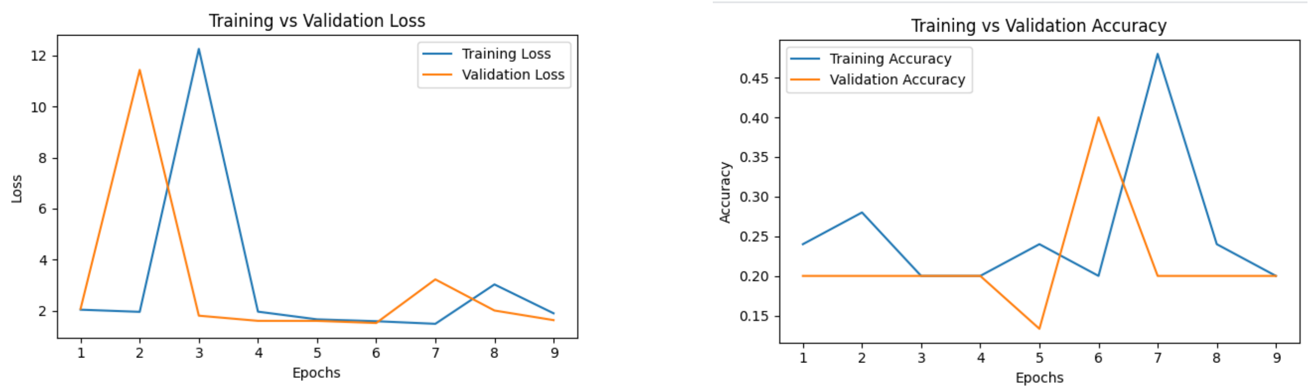
Table 3. Performace using **EuroSAT-RGB**



(a) ResNet18 Training

(b) ResNet18 Testing

Figure 1. ResNet18 Accuracy and Loss Curve



(a) VGG16 Training

(b) VGG16 Testing

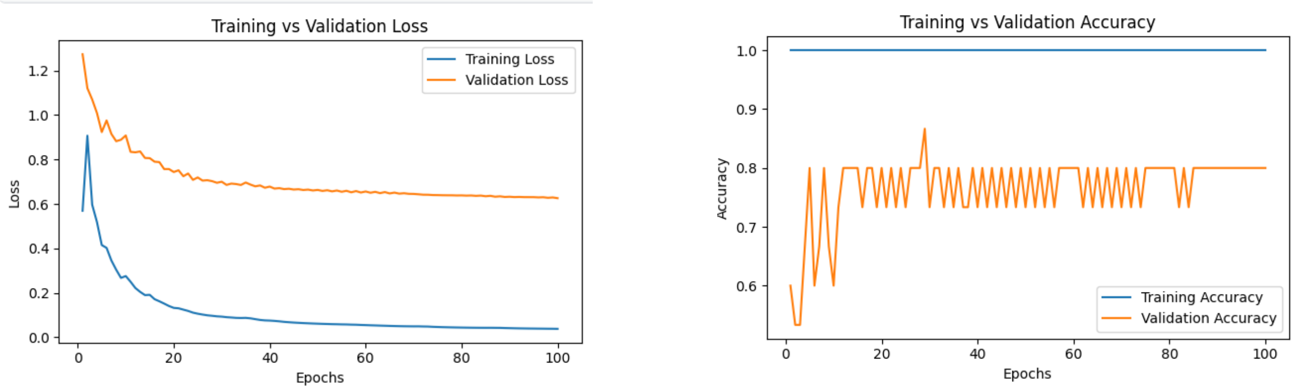
Figure 2. VGG16 Accuracy and Loss Curve

concludes by preserving the state of the best model, representing a convergence of theory and practicality in image classification research.

References

[1] Wendy Kan Addison Howard, Eunbyung Park. Imagenet object localization challenge, 2018. 2

[2] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov,



(a) ViT Training

(b) ViT Testing

Figure 3. Vision Transformer Accuracy and Loss Curve

Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 3

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[4] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 204–207. IEEE, 2018. 2

[5] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019. 2

[6] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 3