CHAPTER I

INTRODUCTION

1.1 Background

Deep Learning represents a sophisticated segment of Machine Learning that employs artificial neural networks featuring several hidden layers. This enables the models to acquire more intricate representations from the data (Dong et al., 2021). With the rapid progress of deep learning, Artificial intelligence has developed the ability to perform tasks that were previously challenging for standard algorithms, such as recognizing speech, processing natural language, and analyzing images. A key advancement in this field is the emergence of specialized architectures tailored to efficiently process different types of data.

In the realm of image data analysis, Convolutional Neural Networks (CNNs) are engineered to process visual data by mimicking how humans perceive images. These networks are composed of essential components, including convolutional layers, pooling layers, and fully connected layers, which function in a sequential manner to draw out features from visual data. The convolutional layers focus on recognizing local characteristics such as edges and textures, whereas pooling layers help in minimizing the data size while maintaining crucial information (Saleem et al., 2022). This approach allows CNNs to autonomously learn and recognize features in images, removing the need for the manual feature extraction typically necessary in conventional machine learning techniques.

A major challenge in applying CNNs as classification models is the need for large datasets to avoid overfitting and achieve strong generalization, which ultimately improves accuracy. Overfitting takes place when a model becomes overly familiar with the training data, picking up on particular patterns and extraneous noise, leading to a lack of generalization to fresh, unobserved data. Consequently, the model may show outstanding performance during the training phase but struggle to make precise predictions in practical situations. Therefore, an effective classification model requires a sufficiently large and diverse dataset. This is supported by the study conducted by Firmansyah (2021), which showed that a substantial dataset is necessary for flower classification tasks. Similarly, research by Praramadhan & Saputra (2022) noted that many classification algorithms require large datasets, presenting a challenge in managing and extracting optimal results from such models. However, in many real-world scenarios, the required datasets are often unavailable in sufficient quantities or may not exist online at all. This issue becomes more prominent in specialized domains such as medical imaging analysis, specific object recognition, and image processing in agriculture and industry. In such cases, researchers must collect data manually, a process that is resource-intensive in terms of time, effort, and cost.

In order to tackle this problem, this research utilizes Generative Adversarial Networks (GANs), which facilitate the enhancement of images through understanding the distribution of data (Goodfellow et al., 2014). However, GANs often face instability issues such as mode collapse and vanishing gradients (Ding et al., 2022). To mitigate these problems, Deep Convolutional GANs (DCGANs) incorporate convolutional layers and batch normalization to stabilize training and generate more realistic images (Radford et al., 2015). Using DCGANs, limited datasets can be expanded with synthetic data that resemble the original, increasing diversity, improving generalization, reducing overfitting, and enhancing classification accuracy.

This study will focus on evaluating the classification model's accuracy before and after using DCGAN, it also includes an additional experiment on the amount of synthetic data needed, inspired by previous research. Bird et al. (2022), for instance, experimented with varying amounts of synthetic data from 0 to 3,000 images and found that the use of around 400 synthetic data resulted in the best performance. This highlights the importance of determining the appropriate quantity of synthetic data to achieve optimal results. The dataset utilized in this research is the STL-10. STL-10 serves as a collection of images created particularly for training in unsupervised deep learning tasks. It is similar in nature to CIFAR-10, but each class in the training data contains fewer labeled examples compared to CIFAR-10 (Papastratis, 2021).

To ensure robust results, this study employs several CNN architectures, including EfficientNet-B0 and MobileNetV2. In research conducted by Azahro Choirunisa et al. (2021), 2,700 images across 9 different cat breeds were classified using the EfficientNet-B0 architecture. The most optimal model, tested on 180 cat images, achieved an accuracy of 95%. Additionally, Thaniket et al. (2025) demonstrated the strong performance of MobileNetV2, attaining a training accuracy of 98.49% and a validation accuracy of 97.50% in a classification task involving three species of birds-of-paradise.

The results of this study are expected to support the advancement of data augmentation techniques across various applications, particularly in animal classification, while also providing opportunities for further exploration in the field of computer vision.

1.2 Problem Identification

Drawing from the previously discussed context of the issue, this research will concentrate on the specific problems that have been identified as follows:

a. CNN models require a large amount of data to achieve satisfactory classification performance. However, in many cases especially in specific domains such as agriculture, available datasets are extremely limited or difficult to obtain.
b. Manually collecting data to meet the training requirements of the model is a time-consuming, labor-intensive, and costly process, making it an inefficient solution in many scenarios.
c. The effectiveness of synthetic data generated by DCGAN needs to be tested across different CNN architectures to determine the extent to which it can enhance classification performance in various scenarios.
d. Previous studies have shown that the proportion of synthetic data impacts model accuracy. However, no definitive standard currently exists regarding the optimal amount of synthetic data to be used in augmentation.

1.3 Research Questions

In light of the context of the dilemma and the challenges recognized, the inquiries for research are articulated in the next manner:

a. How does DCGAN work for data augmentation?
b. How do EfficientNet and MobileNet perform on the original dataset?

c. How do EfficientNet and MobileNet perform on the DCGAN-augmented dataset?

1.4 Scope of the Study

To maintain alignment with the research questions and prevent deviation, the scope of this study is defined as follows:

a. This study exclusively employs the Deep Convolutional Generative Adversarial Network (DCGAN) to augment the size of the image dataset.
b. The classification models used are based on deep learning, specifically EfficientNetB0 and MobileNetV2 architectures.
c. The dataset used in this research is the STL-10 dataset, with four selected classes: cat, deer, dog, and horse.
d. The classification models will be evaluated based on their accuracy before and after augmentation with various ratios of synthetic data generated by DCGAN.
e. The images used in the study are resized to 224×224 pixels.
f. The dataset is divided into 80% for training and 20% for testing.
g. The images processed are RGB (Red, Green, Blue) images in three-channel format to preserve color information that may influence classification outcomes.

1.5 Research Objectives

The objectives of this study are to provide validated insights into the following:

a. To analyze the effectiveness of DCGAN in generating synthetic data for data augmentation.

b. To evaluate the classification performance of EfficientNet-B0 and MobileNetV2 on the original STL-10 dataset.

c. To evaluate and compare the classification performance of EfficientNet-B0 and MobileNetV2 on the DCGAN-augmented dataset.

1.6 Research Significance

The findings from this research are anticipated to provide advantages to multiple parties, as detailed below:

1.6.1 Theoretical Significance

a. This research is expected to contribute to the enrichment of scientific knowledge, particularly in the field of generative models, and enhance understanding of their application in data processing and machine learning.

b. The findings of this study may also serve as a reference or foundation for future research that aims to explore similar topics, thereby supporting the ongoing advancement of scientific inquiry.

1.6.2 Practical Significance

a. This research provides a deeper insight into the impact of synthetic data on the effectiveness and precision of deep learning models, especially in tasks related to image classification.

b. It offers practical guidance regarding the optimal quantity or proportion of synthetic data to be used, particularly for the STL-10 dataset which can assist researchers and practitioners in optimizing the performance of the models they develop