CHAPTER I
INTRODUCTION

This chapter outlines research on the classification of emergency situations based on aerial photos using the RetinaNet architecture. The discussion in this chapter covers the background, problem statement, research objectives, research benefits, and problem limitations. This description is intended to provide an overview of the underlying issues of the research, the scope of the discussion, and the expected results from applying RetinaNet with various *backbones* and data balancing techniques to the AIDER dataset.

1.1. Background

An emergency is a situation that occurs suddenly, urgently, and seriously, or an unexpected change that requires immediate action to mitigate risks, particularly those threatening life, health, and property. From the beginning of the year through March 2023, there have been 721 disasters in Indonesia so far. During an emergency, identifying the area and environment where the incident occurred, as well as its impact on the surrounding area, is crucial because it significantly influences how rescue teams establish safety protocols [1]. To obtain this information, several methods are employed, one of which is the use of *unmanned aerial vehicles* (UAVs), more commonly known as drones [2].

A drone is a type of unmanned aircraft that can be controlled via a remote control or operated automatically using software. Equipped with a sensor camera, *drones* can enhance situational awareness during emergency situations. According to a news report on the National Search and Rescue Agency (BASARNAS) website, in 2017, the Indonesian SAR team purchased several *marine drones* and conducted training for emergency response in aquatic areas. *Drones* are used to capture aerial photos that are later analyzed to assess emergency situations that have occurred; this is done to facilitate mitigation and minimize risks during emergency response operations.

Deep learning is a method frequently used for automatic aerial photo analysis, and one such method is RetinaNet. RetinaNet is an object detection model that combines feature extraction capabilities via *a convolutional neural network* (CNN) *backbone* with a *one-stage* object *detection* mechanism. Several studies have been conducted on classifying emergency situations using aerial photographs, including the use *of deep learning* in their training. As we know, technological advancements have been rapid in recent years. Deep learning is one of the trending fields in the world of technology. A key component of *deep learning* is *the Convolutional Neural Network (CNN),* an architecture focused on object classification in images [3]. A CNN is a system that mimics how the human neural network recognizes images through the visual system. In this process, the CNN transforms each pixel of the input image into layers of neurons arranged in a feature map, where each neuron indicates whether a feature (defined by a convolution filter) is present at a specific location in the image. There are several CNN models available, one of which is RetinaNet

RetinaNet is one such CNN model. First introduced in 2017, this model is a single-model architecture comprising a backbone network and two subnets, each with its own task: classification and regression [4]. RetinaNet excels at classifying small, dense objects, which is why it is widely used as an object detection model for aerial or satellite imagery. RetinaNet is *a one-stage object detector*; models of this type tend to suffer from class imbalance. However, RetinaNet employs a modified version of *the cross-entropy loss* function*,* known as *focal loss.* This function works by reshaping the loss function to reduce the weight of " " data, thereby focusing training more on *"hard negative"* data. This increases the importance of correction values for misclassified data. The model is designed for detection tasks, balancing speed and accuracy, making it suitable for systems requiring rapid responses, such as emergency classification of aerial imagery.

Several studies have previously been conducted using RetinaNet. One of them was conducted by Jasman Pardede et al. in 2020. This study focused on the implementation of RetinaNet in malaria diagnosis by calculating the malaria index value based on the number of infected and normal red blood cells. In this study, two backbones were used: ResNet50 and ResNet101. ResNet50 achieved an accuracy rate of 71%, while ResNet101 achieved an accuracy rate of 73%. Furthermore, a study conducted by Irma Amelia Dewi et al. used RetinaNet with ResNet101 to detect vehicles as part of a smart transportation network. This study utilized 1,600 training images and 400 validation images, with 100 epochs, resulting in a mean average precision (mAP) of 65%.

Based on these studies, RetinaNet demonstrates strong capabilities in object detection. The use of *the* ResNet50 and ResNet101 *backbones* has also proven to help RetinaNet improve its performance in detection models. Additionally, the inclusion of *a Feature Pyramid Network* (FPN) enables this model to detect various object shapes, even against backgrounds of varying complexity. Therefore, in this study, RetinaNet was implemented for emergency classification using the *Aerial Image Dataset for Emergency Response* (AIDER). The AIDER dataset consists of aerial photographs capturing various emergency conditions captured by drones, including fires, floods, traffic accidents, and building collapses. This dataset is widely used in research on object detection and emergency classification because it provides data that represents emergency situations from an aerial perspective.

In object detection, a common problem faced is class imbalance. When there are different amounts of data for each class, the model may concentrate more on understanding the features of the class with more data instead of the one with less. This can lead to decreased effectiveness of the model in detecting objects. RetinaNet includes a function known as focal loss. Focal loss is a variant of cross-entropy loss that has been modified to pay more attention to hard-to-identify data, which encourages the model to focus on objects that are often misclassified, thus enhancing the model's effectiveness in object detection.

The AIDER dataset displays an uneven distribution of data across classes, which could impact the model's ability to detect objects effectively. Therefore, besides using focal loss, data balancing methods such as upsampling and downsampling are also implemented. Upsampling involves increasing the amount of data in the minority class, while downsampling means decreasing the amount of data in the majority class. These methods help in evaluating how the distribution of data influences RetinaNet's success in classifying emergency situations from aerial images.

1.2. Problem Statement

The background section explains the need for further analysis regarding the application of the RetinaNet architecture in classifying emergency objects based on aerial photos. Therefore, the research problem in this study is formulated to examine the model's performance based on variations in *the backbone* and the data balancing techniques used, as follows.

1. How does the RetinaNet architecture perform in classifying emergency objects from aerial photos on the AIDER dataset?
2. How does the use of different ResNet50 and ResNet101 *backbones* affect the performance of the RetinaNet model in classifying emergency situations?
3. How does the application of data balancing techniques (*upsampling* and *downsampling*) compared to the original data affect the performance and classification results of the RetinaNet model?
4. Which combination of *backbone* architecture and data balancing techniques yields the most optimal emergency classification results for the RetinaNet model?

1.3. Research Objectives

The research objectives are formulated to guide the study in addressing each of the research questions. Therefore, based on the research questions, the objectives of this study are as follows.

1. To determine and analyze the performance of the RetinaNet architecture in classifying emergency objects based on aerial photos in the AIDER dataset.
2. To analyze the effect of using *the* ResNet50 and ResNet101 *backbones* on the RetinaNet model's performance in classifying emergency objects.
3. To analyze the impact of applying data balancing techniques (*upsampling* and *downsampling*) compared to using raw data on the performance and classification results of the RetinaNet model.
4. To determine the combination of *backbone* and data balancing techniques that yields the most optimal emergency classification performance in the RetinaNet model based on evaluation results.

1.4. Research Benefits

This research is expected to contribute to the development and application of *deep learning* for aerial image processing, particularly in the classification of emergency objects. The expected benefits of this research are as follows:

1. To provide information on the RetinaNet architecture's ability to classify emergency objects based on aerial photos in the AIDER dataset.
2. To provide an overview of the impact of using *the* ResNet50 and ResNet101 *backbones* on the RetinaNet model's performance, thereby serving as a consideration in selecting the appropriate *backbone*.
3. To provide an understanding of the impact of data balancing techniques—specifically *upsampling* and *downsampling*—compared to the use of raw data on the model's classification results.
4. To provide recommendations for combinations of *backbone networks* and data balancing techniques that yield the most optimal performance in emergency object classification.
5. To serve as a reference for future research on the classification or detection of emergency situations using aerial photographs, RetinaNet, or data balancing methods.
6. To serve as a foundation for the development of artificial intelligence-based systems that can help identify emergency situations from aerial photos more quickly and efficiently.

1.5. Problem Scope

Problem boundaries were established to ensure that the research discussion remained focused and aligned with the stated objectives. The problem boundaries in this study are as follows:

1. The study uses aerial images from the AIDER dataset as training, validation, and testing data.
2. The objects analyzed are limited to the emergency categories available in the AIDER dataset.
3. The *deep learning* architecture used is limited to RetinaNet with two *backbone* variations: ResNet50 and ResNet101.

4. The data balancing techniques compared consist of using raw data, *upsampling*, and *downsampling.*
5. Data balancing was applied only to the training data so that the validation and test data distributions would continue to reflect the actual data conditions.
6. The research focused on comparing model performance based on variations in *the backbone* and data balancing techniques, without comparing RetinaNet to other architectures such as YOLO, Faster R-CNN, or SSD.
7. Each model was trained using the same parameter configuration, with the *backbone* type and data distribution conditions serving as the variables of interest.
8. Model performance was evaluated using *precision, recall*, and mAP, which correspond to the model's output.
9. The best model was determined based on the test results.
10. This study focuses solely on computational image processing and classification and does not cover the direct capture of aerial photographs using *drones* or the application of the models to real-time emergency response systems.