



Faculty of Computer Science and Information Technology

Deep Learning Model with Attention Mechanism for Biometric Re-Identification of Green Sea Turtles in Long-Term Tracking Scenario

Khalif Amir Zakry

**Master of Science
2024**

Deep Learning Model with Attention Mechanism for Biometric Re-Identification of Green Sea Turtles in Long-Term Tracking Scenario

Khalif Amir Zakry

A thesis submitted

In fulfillment of the requirements for the degree of Master of Science

(Software Engineering)

Faculty of Computer Science and Information Technology

UNIVERSITI MALAYSIA SARAWAK

2024

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.



.....

Signature

Name: Khalif Amir Zakry

Matric No.: 21020463

Faculty of Computer Science and Information Technology

Universiti Malaysia Sarawak

Date: 16 January 2024

ACKNOWLEDGEMENT

Here I would like to take this opportunity to thank Dr. Irwandi Hipni for his mentorship and guidance throughout the entire course of this research journey. His academic advice and support have helped me tremendously in figuring out the steps and directions that directed me toward crafting this thesis. I am glad to have worked with him on this thesis and will be looking forward to working with him when future opportunities arise.

I would also like to thank my colleague Mohamad Syahiran for being a great working partner throughout this research project. Thank you for staying strong and committed during our field works, for providing great insights and responses regarding academic writing, and for being a dependable friend throughout this journey. It has been a great experience working with you.

I would like to say thanks as well to my family for being supportive and wonderfully understanding over the past 2 years while I was working on this thesis. It means so much and I am thankful to my mom and two sisters for the support given. The journey in completing this thesis has indeed been stressful but having a caring family has indeed helped to alleviate the troubles away.

Finally, I would like to thank the folks at Sarawak Forestry Corporation for providing us with permission and authorization to work with the endangered *Chelonia mydas* sea turtles. And to the park rangers and wardens, Mr. Syahrizal, Mr. Nulit, Mr. Alphonsus, Mr. Syahrizan, and the other park staff; I would like to say thank you to them for their physical, logistical, and technical assistance throughout our fieldwork at the Talang Satang Islands.

ABSTRACT

GPS tags, plastic tags, and Inconel tags that are difficult to apply and degrade over time are currently used for re-identification of *Chelonia mydas* sea turtles. A deep learning model with attention mechanism is proposed for biometric re-identification of individual sea turtles in a long-term scenario. The deep learning model with attention mechanisms re-identifies sea turtles on which it has been trained via its carapace. The carapace pattern is a stable feature that would change over time due to scratching or barnacle growth, making re-identification a non-trivial task. Results of testing with datasets obtained from sampling efforts at Talang Besar, Sarawak, using a deep learning model with attention mechanism achieves 97.23% accuracy on direct re-identification and 73.61% accuracy on long-term re-identification scenario. This is comparable to an AlexNet CNN model in direct re-identification and showed 1.39% accuracy improvement in long-term scenario tests. The attention-based model however is -1.38% and -2.77% less accurate than SqueezeNet and ResNet50 CNN models in direct re-identification and inaccurate by -6.95% and -18.06% respectively than SqueezeNet and ResNet50 CNN models in long-term scenario tests. The work shows that deep learning model with attention mechanisms can perform re-identification directly and in a long-term simulated scenario of barnacle growth with reliable accuracy above 70% but still underperforms when compared to some CNN models trained and tested on the same tasks. The findings of this work show that machine learning models can be used for the re-identification of *Chelonia mydas* sea turtles and that such models can be adopted for further use animal re-identification purposes.

Keywords: Deep learning, biometric re-identification, attention mechanism, sea turtles

Model Pembelajaran Mendalam dengan Mekanisme Perhatian Untuk Pengenalpastian Semula Berasaskan Biometric Penyu Laut Hijau dalam Senario Penjejakan Jangka Panjang

ABSTRAK

Tag artifisial bagi tujuan pengecaman identiti penyu laut (Chelonia mydas) sukar dipasang dan mudah rosak. Sebagai langkah penyelesaian, satu model pembelajaran mendalam dengan “attention mechanism” dilatih bagi tujuan pengecaman jangka panjang. Model ini dilatih menggunakan gambar cengkerang yang diambil dari sisi pandangan atas. Corak cengkerang boleh digunakan untuk pengecaman identiti individu, namun kesan calar atau pertumbuhan teritip boleh menghasilkan padanan yang salah. Dengan menggunakan data latihan yang dikumpul di Talang Besar Sarawak, model tersebut memperoleh kadar ketepatan sebanyak 97.23% untuk padanan terus, dan 73.61% untuk padanan menggunakan data jangka masa panjang. Jika dibandingkan dengan model lain dengan senibina CNN, model ini mencapai kadar ketepatan yang setara dengan AlexNet untuk padanan terus dan 1.39% lebih tepat untuk padanan jangka masa panjang. Untuk SqueezeNet dan ResNet50, model ini mencapai kadar ketepatan yang kurang bagi padanan terus (-1.38% dan -2.77%) dan jangka panjang (-6.95% dan -18.06%). Hasil kerja penyelidikan ini membuktikan model pembelajaran mendalam dengan “attention mechanism” boleh digunakan untuk pengecaman terus dan pengecaman senario jangka panjang. Kadar ketepatan yang diperolehi adalah baik iaitu di atas 70.00% namun tidak sebaik model-model dengan senibina CNN yang lain. Model ini mempunyai potensi untuk ditambah baik bagi tujuan pengecaman penyu Chelonia mydas dan haiwan lain secara amnya.

Kata kunci: *Pembelajaran mendalam, pengenalpastian biometrik, attention mechanism, penyu*

TABLE OF CONTENTS

	Page
DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
<i>ABSTRAK</i>	iv
TABLE OF CONTENTS	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	xii
CHAPTER 1 RESEARCH INTRODUCTION	1
1.1 Introduction	1
1.2 Problem Statement	2
1.3 Research Questions	4
1.4 Research Objectives	5
1.5 Motivations	5
1.6 Research Scope	6
1.7 Thesis Structure	7
CHAPTER 2 LITERATURE REVIEW	10
2.1 Introduction	10
2.2 Brief on Chelonia Mydas Turtles	12
2.3 Current Techniques and Methods of Tracking and Re-Identification of Sea Turtles	14

2.4	Deep Learning and Object Detection Technology	18
2.5	Deep Learning Tools in Wildlife Re-Identification	22
2.6	Existing Usage of Deep Learning Tools Used for Chelonia Mydas Re- Identification	27
2.7	Summary	31
CHAPTER 3 METHODOLOGY		33
3.1	Introduction	33
3.2	Research Flowchart	34
3.3	Sampling and Data Collection Techniques Innovated	36
3.4	AI-Based Image Filtration, Analysis, and Extraction with YOLOv7	46
3.5	Creating Datasets with Python SplitFolders	50
3.6	CNN and Attention-Based Deep Learning Models for Re-Identification	52
3.7	Testing Developed CNN Models and Attention-Based Model with Artificially Aged Chelonia Mydas Test Images for Long-Term Scenario Simulation	61
3.8	Summary	63
CHAPTER 4 FINDINGS AND DISCUSSIONS		65
4.1	Setup of Experiments	65
4.2	Comparison of Training Loss Metric Among the Models	66
4.3	Comparison of Validation Loss Metric Among the Models	69
4.4	Comparison of Training Accuracy Metric Among the Models	71
4.5	Comparison of Validation Accuracy Metric Among the Models	73
4.6	Comparison of Training Time After Every Epoch Metric Among the Models	75
4.7	Re-Identification Capabilities of All Models with Direct Re-Identification on Test Set	76

4.8	Re-Identification Capabilities of All Models with Simulated Long-Term Scenario of Barnacle Growth	80
4.9	Summary	84
CHAPTER 5 CONCLUSION		86
5.1	Summary	86
5.2	Conclusion	88
5.3	Future Work	88
REFERENCES		90
APPENDICES		104

LIST OF TABLES

	Page
Table 2.1: Table of summary of deep learning tools in wildlife re-identification	27
Table 2.2: Table of summary on works utilizing images for turtle re-identification	30
Table 4.1: Table of Total Training Time Taken to Reach Every Epoch	75
Table 4.2: Accuracy results of confusion matrixes on the test set	79
Table 4.3: Accuracy results of confusion matrixes on the test set	82

LIST OF FIGURES

	Page
Figure 1.1: Map of Talang-Satang in Sarawak (Sarawak Forestry Corporation, 2023).	6
Figure 1.2: Captured image of a female <i>Chelonia mydas</i> individual treading along the sandy beach of Talang Besar Island in daylight (September 2022)	7
Figure 2.1: Cheetah equipped with GPS tracking collar (South African Tourism, 2009).	11
Figure 2.2: Adult <i>Chelonia mydas</i> swimming among the coral reef (Dupont, 2010).	13
Figure 2.3: A GPS tag mounted on a sea turtle individual (Sea Turtle Conservancy, 2023).	14
Figure 2.4: Inconel tag with its ID shown (National Band & Tag Company, 2018).	16
Figure 2.5: Inconel tag close-up on a flipper (National Band & Tag Company, 2018).	17
Figure 2.6: A pole-mounted trap camera positioned on a beach habitat (August 2022).	24
Figure 2.7: A Green Sea Turtle with distinct facial patterns (Witherington et al., 2006).	28
Figure 3.1: Research Flowchart of the study	34
Figure 3.2: The red square showcases the restricted zone where visitors are not permitted to be within 2.8km of the national park without special arrangements with SFC (October 2023).	37
Figure 3.3: High-resolution satellite image of Talang Besar Island's beachfront (October 2023).	39
Figure 3.4: High-resolution satellite image of Satang Island's beachfront (October 2023).	40
Figure 3.5: Three different sample images of a turtle's carapace taken from 3 different female <i>Chelonia mydas</i> individuals with Bushnell NatureView HD (August 2022).	41
Figure 3.6: The two different hardware and camera setups used in the sampling methods (August 2022).	43

Figure 3.7: Image section A shows the IR illumination of the CCTV camera while image section B shows the live video feed of the CCTV camera of image section A (October 2022)	45
Figure 3.8: Pseudocode of the cropping function in the YOLO v7 Object Cropping program	48
Figure 3.9: Several example images of the annotated dataset by using Roboflow	49
Figure 3.10: Example folder structure of the dataset	51
Figure 3.11: Code snippet of SplitFolders v0.5.1	52
Figure 3.12: Example of an AlexNet CNN encoder	54
Figure 3.13: The implementation of the vision transformer architecture in this research	55
Figure 3.14: Example on the Design of the Multi Head Attention	57
Figure 3.15: The hyperparameters used in the vision transformer model	59
Figure 3.16: The model parameters that are shared for all 3 CNN models (AlexNet, SqueezeNet, and ResNet50)	60
Figure 3.17: Shows two images of simulated barnacle growth on 2 individual carapaces (September 2022).	62
Figure 4.1: Graph of Training Loss for All Models At All Epochs During Training	66
Figure 4.2: Graph of Validation Loss for All Models At All Epochs During Training	69
Figure 4.3: Graph of Training Accuracy for All Models At All Epochs During Training	71
Figure 4.4: Graph of Training Accuracy for All Models At All Epochs During Training	73
Figure 4.5: Confusion Matrix for Vision Transformer Model on Direct Re-Identification	76
Figure 4.6: Confusion Matrix for AlexNet CNN Model on Direct Re-Identification	77
Figure 4.7: Confusion Matrix for SqueezeNet CNN Model on Direct Re-Identification	77
Figure 4.8: Confusion Matrix for ResNet50 CNN Model on Direct Re-Identification	78

Figure 4.9: Confusion Matrix for Vision Transformer Model on Long-Term Set	80
Figure 4.10: Confusion Matrix for AlexNet CNN Model on Long-Term Set	80
Figure 4.11: Confusion Matrix for SqueezeNet CNN Model on Long-Term Set	81
Figure 4.12: Confusion Matrix for ResNet50 CNN Model on Long-Term Set	81
Figure 4.13: These are 2 images that were used to test the re-identification capabilities of the models for direct and long-term scenario (September 2022).	83

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AMD	Advanced Micro Devices
ANN	Artificial Neural Networks
CCTV	Closed-Circuit Television
CMOS	Complementary Metal Oxide Semiconductor
CMR	Capture-Mark-Recapture
CNN	Convolutional Neural Networks
CPU	Central Processing Units
DNR	Digital Noise Reduction
DSLR	Digital Single-Lens Reflex
FOV	Field of View
GIMP	GNU Image Manipulation Software
GNU	Gnu's Not Unix
GPS	Global Positioning Satellite
GPU	Graphic Processing Units
HD	High Definition
HOG	Histogram of Oriented Gradients
ID	Identification
IR	Infrared
IUCN	International Union for Conservation of Nature
JPG	Joint Photographic Experts Group
LAN	Local Area Network

NLP	Natural Language Processing
PIT	Passive Integrated Transponders
RCA	Radio Corporation America
RFID	Radio Frequency Identification
ROI	Region of Interest
SaaS	Software As a Service
SFC	Sarawak Forestry Corporation
SIFT	Scale-Invariant Feature Transform
SURF	Speeded Up Robust Features
SVM	Support Vector Machines
UNIMAS	Universiti Malaysia Sarawak
YOLOv7	You Only Look Once v7

CHAPTER 1

RESEARCH INTRODUCTION

1.1 Introduction

The word AI stands for artificial intelligence, and it is now the latest trendy technological phenomenon on the block since blockchain's rapid decline in popularity. Due to the advancement in neural network architectures and the increasingly numerous availabilities of data, AI is now starting to enter the mainstream user space when it began gaining traction in 2020 (Chow & Perrigo, 2023). For 2018 alone, AI research papers account for 3% of all peer-reviewed journals and 9% of all published conference papers (Perrault et al., 2019). Compared to the year 1999, where only 1% of articles and 3% of conference papers published were about AI; this shows that the amount of research being done on AI is certainly growing and advancing at a significant pace.

Advancements in AI research, specifically in one of its branches, Machine Learning, have brought about many useful industrial, ecological, economic, and societal applications in today's rapidly modernizing society. The list of applications that can benefit and utilise the wonders of Machine Learning is certainly diverse and can range from self-driving vehicles to intelligent chat bots and tumour-detecting X-ray analysers. One of the major applications of Machine Learning is that it can be used for image recognition and computer vision. For these applications, Deep Learning (a subset of Machine Learning) which uses Convolutional Neural Networks (CNNs) to "learn" and "train" a model is used. The model that was trained can make predictions about the subject of its dataset input. This is especially useful in computer vision and image recognition, specifically for identifying characters,

animals, objects, and patterns within images as well as for biometric identification of people or the re-identifications of animals. This research proposal aims to use a deep learning model with attention mechanisms for re-identifying Green Sea Turtles (*Chelonia mydas*) for a long-term tracking scenario which could help in the conservation efforts of the species.

This chapter introduces the research's problem statement, research questions, objectives, background, methods, motivations, as well as its scope.

1.2 Problem Statement

Green Sea Turtles, also known by their scientific name, *Chelonia mydas*, can be found globally in tropical and subtropical waters. *Chelonia mydas* are a migratory species of sea turtles which can have a long lifespan of up to 100 years and only reaches sexual maturity at around 20 to 50 years of age (National Wildlife Federation, 2024). Females of the species would find a suitable beach on which they could lay their eggs in the sand, and after 60 days, the juveniles would hatch (National Oceanic and Atmospheric Administration, 2023). The juveniles then make their way into the sea, guided by the downward slope of the beach and the reflections of both the moon and stars reflected on the water's surface. Sea turtles are thought to return only to their respective natal beaches for nesting. Some sea turtles are recorded to nest on multiple beaches (Brothers & Lohmann, 2018). Green Sea Turtles, being a highly migratory species that it is, can be difficult to track and due to its long lifespan can be hard to re-identify as tags or markers placed on them can be lost due to the advance of time. Thus, perhaps some research can be done to see if deep learning with attention mechanisms or with convolution neural networks may be used for the re-identification purpose of *Chelonia mydas* individuals.

One of the current Re-ID techniques used for the re-identification and tracking of *Chelonia mydas* turtles is by using man-made artificial markers such as the externally applied Flipper Tags and GPS tag, or the internally applied PIT Tags (National Oceanic and Atmospheric Administration, 2023). Both external and internal tags when applied would introduce pain to the turtle (Balazs, 1999). Besides that, the procedure to tag a turtle can be complicated and requires a multitude of tools. For example, all tag applicators and piercing equipment must be disinfected before use and the separation of the tagging gears that was used on healthy and unhealthy turtles is recommended. The current disadvantage of using these tags, specifically for those applied externally to the turtle such as the flipper tags and GPS tags is that they may be lost. It is also noted that tag loss can occur over time due to tearing and corrosion. This significantly diminishes the value of its use, especially for long-term tracking scenarios. An alternative method for re-identification of Green Sea Turtles using biometric markings can be proposed as a solution to this problem. For *Chelonia mydas*, one of its unique biometric traits, its carapace, can be used for this purpose. A machine learning model based on deep learning with attention mechanisms or with the use of Convolutional Neural Networks may be used to determine an individual of the *Chelonia mydas* species, and thus be used in re-identification purposes for tracking the species. The aim is to prove that deep learning models can be used for this purpose if they can be able to correctly re-identify a significant number of Green Sea Turtle individuals via the methodologies of this work.

This research can help in the tracking and conservation efforts of *Chelonia mydas* in Sarawak. The Green Sea Turtles can be re-identified and kept track of without the need for physically intrusive tracking methods that can harm and jeopardize the physical health of the Green Sea Turtles. Besides that, unlike the man-made physical tags which can be lost or

degraded due to time; utilizing its carapace for tracking provides a possible advantage for long-term tracking scenarios due to the carapace being an actual feature stable physical attribute of the turtle instead of a fragile artificial attachment (i.e., flipper tags). The reliability of the carapace being useful in long-term tracking scenarios for re-identification purposes must also be tested against the possibility of natural changes that are commonly occurring on a sea turtle such as barnacle growth.

Without the need of human interactions for tracking; the Green Sea Turtles may no longer need to experience the trauma of human handling and interactions that can inhibit their preference to nest on the beaches of Sarawak. This can help protect and conserve the Green Sea Turtle population better.

1.3 Research Questions

There are several research questions that can help guide this work. These questions shall solidify the purpose of conducting this research and thus steer the work toward the right methodologies and procedures that shall be undertaken to achieve the research objectives.

In a gist, the questions are:

- How can deep learning models be used to filter, and extract cropped carapace images from field video recordings of sea turtles?
- What is the efficiency and effectiveness of the attention-based model and the CNN deep learning models for the re-identification of Green Sea turtles in short term and long-term barnacle growth scenario?
- What is the advantage of using an attention-based deep learning model compared to the standard CNN-based approach?

1.4 Research Objectives

The research objectives are:

- To Conduct a Preliminary Study on Deep Learning, Object Detection, *Chelonia mydas*, and Planning of Data Collection and Data Preprocessing Work on *Chelonia Mydas* sea turtles.
- To Build an Automated Pipeline for Cropping ROIs from Field Videos
- To Compare the Efficiency and Effectiveness of Standard CNN-based Models Versus

1.5 Motivations

Green Sea turtles are an endangered species as listed by the International Union for the Conservation of Nature. Innovating new technologies that can help in the conservation and protection efforts for the species is indeed beneficial. These new technologies can hopefully prevent the extinction and reduction of the population of this species. One of the United Nations Sustainable Development Goals is for life below the water. It aims to sustainably manage and protect marine and coastal ecosystems from pollution and its impacts. Contributing towards this goal for this marine species can also signify the possibility of using this technology for the conservation and tracking of other marine species. The work in this thesis provides a contribution from the field of computer science that can help to ensure that marine wildlife such as sea turtles can continue to exist for years and perhaps generations to come.

1.6 Research Scope

This research focuses only on one species of sea turtle which is the Green Sea Turtle, also known by its scientific name *Chelonia mydas*. The dataset for this research is obtained by conducting data collection fieldwork at the Talang-Satang Islands off the southwest coast of Sarawak. The four islands are designated as a national park with 3 of them being a restricted area that is only accessible to researchers, volunteers, or students with permits granted by the Sarawak Forestry Corporation (SFC). The three restricted areas are Pulau Satang Kecil, Pulau Talang Kecil, and Pulau Satang Besar. Pulau Talang Besar is where most of the data collection process takes place as it is the area where the largest concentration of *Chelonia mydas* turtles would come ashore for nesting among the four islands. Figure 1.1 shows the location of the Talang-Satang islands in Sarawak.



Figure 1.1: Map of Talang-Satang in Sarawak (Sarawak Forestry Corporation, 2023).

There are other known species of sea turtles besides the Green Sea Turtle, *Chelonia mydas*, that can also be found nesting among these islands such as the Hawksbill Sea Turtle, also known scientifically by its species name *Eretmochelys imbricata*. However, for this research, the focus will be only on the *Chelonia mydas* species of green sea turtles which

account for more than 90% of the sea turtle landings recorded on the islands. Figure 1.2 shows an example image of a *Chelonia mydas* individual taken at Talang Besar island.



Figure 1.2: Captured image of a female *Chelonia mydas* individual treading along the sandy beach of Talang Besar Island in daylight (September 2022)

1.7 Thesis Structure

For this thesis, it is formatted and written in the style following the UNIMAS Centre for Graduate Studies' November 2021 thesis template. The language of the thesis is English, and the length of this thesis is 5 chapters. The chapters' title and their short briefs are explained as the following:

- **Chapter 1 - Introduction:** This chapter provides the reader with an introductory brief on the topic of this research by giving them an introduction on artificial

intelligence and how it can be used to tackle problems such as animal re-identification for sea turtles. It provides context on the problem as well as the motivations behind this research as well as informing the reader regarding the research scope as well as what to expect from the following chapters.

- **Chapter 2 – Literature Review:** This chapter explores the related research and works that are currently in use or was used in the field of ecology and artificial intelligence. The current methods of animal tracking and re-identification in ecology are reviewed and the cutting-edge integration of deep learning models in this field is also looked into.
- **Chapter 3 – Methodology:** In this chapter, the methodology used in collecting the sample data, the devices used and the intricate details of how the data collection process occurs on the Talang-Satang Islands, especially Talang Besar island will be discussed. After that, the automated filtration technique used to aid in filtering out the collected noisy field video recordings that were obtained from the field videographic work that was done is explained. Next, explanations on how the dataset is created as well as more context on the CNN models and the attention-based deep learning model that was used for direct re-identification and long-term re-identification scenario by way of simulating various degrees of barnacle growth on each individual sea turtle is also discussed in this chapter.
- **Chapter 4 – Findings and Discussions:** In Chapter 4, the results of the experiments conducted in Chapter 3 are compiled. The results are analysed and discussions regarding the performance of the different deep learning models on several evaluation metrics are presented in this chapter. Comparisons are done in terms of metrics such as loss and accuracy, of both short-term and long-term test results of re-

identification on the two different test sets. Other metrics are also discussed in this chapter besides the two mentioned.

- **Chapter 5 – Conclusion:** The last chapter summarizes the thesis and provides a research conclusion to the overall research done on performing re-identification on *Chelonia mydas* turtles using CNN deep learning models and the attention-based model such as the Vision Transformer model used in this research. Future improvements and follow-up work are discussed in this chapter.

The full tabulated results of the experiments and list of references cited in the literature review for this thesis shall be attached and listed on the following pages after Chapter 5 is concluded in the appendices and reference sections. They serve as a point of reference for the reader and examiner to investigate as supplementary materials to complement this thesis.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

With the current rise in availability and numerous options of cost-efficient hardware such as graphic processing units (GPU) which are much more capable of running intensive deep learning tasks such as deep learning compared to traditional core processing units (CPU) (Kayid et al., 2018); society has begun to see widespread adoption of deep learning models in today's world. Such examples include the adoption of artificial intelligence (AI) powered chatbots with robust natural language processing capabilities in the e-commerce industry that have provided companies with the ability to service their customers 24/7, thus giving customers a readily convenient point of interaction with the company (Misischia et al., 2022) or when automotive companies such as Tesla use Nvidia Drive PX 2 AI processors and neural networks to attain self-driving capabilities for their electric vehicles (Chen & Mao, 2022). The benefits of AI and deep learning models can certainly be worked on and developed more to be adapted to more sectors and fields such as the research field of this paper which is ecology. Ecology is currently defined as a diverse discipline that studies the interactions between living organisms and their ecosystems by way of understanding the distribution or abundance of those organisms and their relationship with the structure and processes of the natural world (Urban et al., 2021).

One of the common ways how ecologists understand the abundance or distribution of living organisms for their studies is by monitoring and tracking the respective species of living organisms that they are researching. Tracking and monitoring provide important

insights into how individuals in a species are dispersed and its impact on the population densities of the species' local community by providing spatial data of its movement and interactions (Schlägel et al., 2020, as cited by Costa-Pereira et al., 2022). Ecology as a field should be able to evolve and improve upon their traditional methods of tracking and monitoring by incorporating new technological inventions to automate and increase the efficiency of their workflow processes. Modern-day ecologists are certainly interested in applying these new methods such as ones that include global positioning satellite (GPS) based tracking or smart animal tags that can capture more data such as acceleration and accurate status locations for animal behavioural studies (Trappes, 2023). Such devices, including GPS collars and trackers, are beneficial in providing information regarding the population number of the species animal especially when the animals are wild animals populated over a large area (Uno, 2020). Figure 2.1 shows an example of a GPS collar fastened on a cheetah.



Figure 2.1: Cheetah equipped with GPS tracking collar (South African Tourism, 2009).

However, non-smart tagging methods such as utilizing physical ID plastic/Inconel tags or radio frequency identification (RFID) tags are still widely used to keep track and re-

identify individuals of an animal species in their natural communities. Both methods require the human researchers to be within radio range or nearby to read off the RFID or physical ID written on the tags. These tags are still used as they present a low-cost, applicable and effective method of animal tracking and does not require constant active power in the case of RFID tags that can operate without batteries (Azmi et al., 2019).

As such, this research aims to investigate the potential of utilizing AI-based deep learning models to be able to track an animal by being able to re-identify the individual of the animal species. The research animal for this thesis is the endangered and protected species of green sea turtle known scientifically as *Chelonia mydas*. This method of tracking the *Chelonia mydas* species is advantageous because it allows the possibility for a non-invasive and non-physical base for re-identification solely by using the unique carapace characteristics of each individual turtle from image data.

2.2 Brief on Chelonia Mydas Turtles

The green sea turtle, *Chelonia mydas*, is a sea turtle species that is recognized as globally endangered by the International Union for Conservation of Nature (IUCN) and is included in their IUCN red list. Such status assignment to the *Chelonia mydas* species is due to the recorded extensive decline of its population, the poaching of its eggs by hunters as well as by accidental or intentional death due to human fishing, and the destruction of its marine and nesting habitats (IUCN, 2004). In the country of Malaysia, Sarawak is blessed to be one of the only 3 states in the nation where the *Chelonia mydas* species of sea turtles can be found nesting on a select number of its beaches. Only two other countries that share the southeast asian seas can this species of sea turtle be found nesting on their beaches, and they are the Indonesian Berau Islands and the Philippine Turtle Islands. The *Chelonia mydas*

turtles represent an integral part of what makes Malaysia a biologically mega-diverse country that possesses coastal marine waters teeming with biodiversity (Mazlan et al., 2005). Figure 2.2 shows an image of an adult *Chelonia mydas* individual swimming in its sea habitat.



Figure 2.2: Adult *Chelonia mydas* swimming among the coral reef (Dupont, 2010).

The green sea turtle, *Chelonia mydas*, is known to reach a long lifespan of up to 75 years of age (Mayne et al., 2020). The age to maturity for juvenile *Chelonia mydas* individuals is among the longest of any sea turtle, but the exact age of which they reach sexual maturity varies between 15 to 50 years of age. To protect the green sea turtles from poaching and exploitative activities, the state government of Sarawak have enacted laws that criminalizes these offences with prison sentences and heavy fines (Hassan & Yahya, 2022). As such, all sea turtle species of Sarawak, including the *Chelonia mydas* green sea turtles that can be found nesting at the Talang-Satang Islands of Sarawak, are listed as a Totally Protected Animal by the state. Besides that, the turtle custodians of Sarawak have been collecting and incubating the eggs nested at the Talang-Satang Islands as part of the conservation efforts to increase the population size of the endangered species. The effects of these measures can only be confirmed by monitoring and tracking the population data of the *Chelonia mydas* individuals over time. Thus, it is vital to be able to re-identify *Chelonia*

mydas individuals who come ashore to these beaches for each nesting season to determine the impact of the said measures.

2.3 Current Techniques and Methods of Tracking and Re-Identification of Sea Turtles

Animal re-identification is one of the key components in ecological research for understanding the population and impact of conservation efforts as discussed in the previous sections. In the case of the *Chelonia mydas* sea turtles however, it requires more effort in re-identifying this animal species compared to land-based terrestrial animals. Unlike a cheetah or a lion, where a collar can be affixed to the animal and remain on the animal reliably, affixing a GPS-based tracker on a *Chelonia mydas* sea turtle is more difficult. Figure 2.3 shows a GPS tag affixed on top of a sea turtle's shell.



Figure 2.3: A GPS tag mounted on a sea turtle individual (Sea Turtle Conservancy, 2023).

The current method of satellite tracking requires that the GPS tag be firmly attached to the carapace of the *Chelonia mydas* individuals via the application of strong hard glue.

The GPS tag has to be mounted as low-profile as possible and to be as unimpactful to the hydrodynamics of the turtle by only providing minimal extra drag (Hays & Hawkes, 2018). The *Chelonia mydas* sea turtle is a marine animal freely living in the wild, its habitat consists of sandy beaches, the open ocean, and underwater seagrass feeding grounds (Lamont et al., 2015). These environments are highly unfriendly to the GPS tags which are man-made physical objects that can be susceptible to damage while being on the *Chelonia mydas* individuals. For example, the GPS tags can detach prematurely due to an insufficient amount of glue used to attach the tags. Besides that, the GPS tags require a line of sight to the satellite to be able to send data for ecologists to study, as such the GPS tags are equipped with an antenna. This antenna is prone to damage and may break off prematurely (Piacenza et al., 2018). Internal antenna for the GPS-tags is possible but offers significant disadvantages over the external antenna which has better transmission quality. The GPS tags are also only usable while their batteries are still able to power the GPS tag to remain active. Satellite charges are also to be considered as satellite service providers charge a subscription rate for the use of their satellite tracking capabilities. Due to the costs and limitations of GPS tags for individual tracking and re-identification of sea turtles, more traditional simple tagging identifications (ID) such as physical ID tags are still commonly used.

For the Sarawak Forestry Corporation (SFC), their rangers and park wardens on Talang-Satang Islands utilise Inconel metal tags that are etched with a unique serial ID that consists of alphanumeric characters. The current practice for re-identification of *Chelonia mydas* individuals at the Talang-Satang Islands by using these Inconel metal tags and tagging these onto both the flippers of the female *Chelonia mydas* individuals that come up for nesting during nesting season. The function of double tagging according to the discussions made with the park warden and rangers is so that should one tag were to go missing due to

circumstances, the second can still be enough for re-identification as both left and right IDs are unique. The nesting females that swim ashore are tagged with these ID tags so that the park rangers and wardens can record and track how many female *Chelonia mydas* turtles come each season and re-identify if a particular turtle is a returning turtle (a regular nester on the beach). The wardens and park rangers would have to apply these tags manually.



Figure 2.4: Inconel tag with its ID shown (National Band & Tag Company, 2018).

This method of tracking is also known as the capture-mark-recapture (CMR) technique normally used in ecological research to understand the population, migration, growth, and returns of sea turtles (Pfaller et al., 2019). Since this type of physical tag is applied externally on the flippers of the sea turtle individual, they have several disadvantages. Firstly, these tags must be clipped onto the flippers utilizing special “tag clippers”. These clippers attach the Inconel tags onto the sea turtles’ flippers by piercing through the flipper scale and through the turtles’ bodies to clip the tag on both sides. As such, it induces short-term pain on the turtle as well as making them feel uncomfortable during tagging.



Figure 2.5: Inconel tag close-up on a flipper (National Band & Tag Company, 2018).

However, it is noted that the tags that are applied correctly onto the flipper of the sea turtle can heal completely in a relatively short amount of time (Balazs, 1999). Besides that, this type of tag is also susceptible to tag loss as coral or barnacle growth may accidentally tear the tags off the flipper making a scar on the turtle. Should these tags be lost or have their ID characters naturally scratched over time; it would be difficult to re-identify the sea turtles that are important to the ecological study (Rivalan et al., 2005). On another note, plastic-type external tags are also available but have even more disadvantages than Inconel tags such as being more easily eroded than Inconel tags.

There are also other forms of internal tags that are in use by researchers outside of Sarawak. These internal tags are known as Passive Integrated Transponders (PIT) tags. They are applied by syringe injection into the muscle tissues of the sea turtles and the ID of the tag once injected can only be read by using a Radio Frequency Identification (RFID) reading device. The main problem with this type of internal tag is that over time as the sea turtle grows, the position of the PIT tag shifts and moves away from the initial injection site onto

unknown locations away from the injection site in a phenomenon known as PIT migration (Wyneken et al., 2010). As such, this is disadvantageous for long-term study as the PIT tag would be difficult to read and scanned after some passage of time.

In conclusion, the potential of deep learning models to utilise non-physical re-identification capabilities for ecological research into *Chelonia mydas* populations is relevant and should be pursued. They would be able to provide an alternative to the previously mentioned methods that are currently in use. As such, literature review to look into the deep learning-based alternatives that are currently in use in ecology and review its capabilities in the following sections.

2.4 Deep Learning and Object Detection Technology

Deep learning is a subset of machine learning whereby machine learning helps in the implementation of artificial intelligence on a system while deep learning is a method that machine learning uses to achieve its machine learning goals more systematically (Chahal & Gulia, 2019). The advantage of using deep learning is that it is simple to configure, versatile, short to develop and is highly performant (Shetty et al., 2020). Deep learning attempts to mimic the human nervous system by having a multi-layer hierarchically connected neural network mimicking that of the human body (Shetty et al., 2020). Deep learning approaches usually uses multiple different layers of information-processing stages in hierarchical structures to learn and as such would contain multiple hidden layers including input and output layers (Sarker, 2021). Deep learning can be adapted to different learning tasks such as supervised learning, unsupervised learning, or even a hybrid of both forms of learning for various purposes. For classification applications, deep learning techniques that involve supervised learning which includes architectures such as the Multi-Layer Perceptron (MLP),

Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN) are typically used. These are just some examples of Deep Neural Network (DNN) architectures. Besides them, there are also other different types of DNN architectures such as Graph Neural Networks (GNN), Generative Adversarial Networks (GAN), and Transformers. The latter, Transformers, is unique in that it is a DNN architecture that utilizes self-attention mechanisms. The use of these DNN architectures for machine learning applications are currently an ongoing study and it is certainly possible that more innovations and adaptations of these current architectures will arrive in the future.

Object detection is essential in performing computer vision-related tasks. This task involves predicting the bounding boxes and class of the objects that are in a particular image (Song et al., 2021). CNN- and Transformer-based machine learning models with their deep neural network architectures are both capable of performing this task with varying degrees of efficiency, accuracy, and success, depending on their implementation. The applications of object detection currently in practice range from home security cameras to traffic detection, satellite imagery, and facial detection. To implement this, machine learning models such as CNN and Transformer models requires huge amounts of training data to train their respective models for object detection capabilities. CNN-based object detection models generally perform the following operations in sequence starting from informative region selection, feature extraction, and classification (Patel & Patel, 2021). On the other hand, Transformer-based object detection models like Vision Transformers (ViT), splits an image into patches by converting them into a sequence of linear embeddings as the input to the Transformer first, then utilizing a constant latent vector to flatten and map those patches to a vector size dimension with trainable projections in all the transformer layers, before finally using MLP layers for classification (Amjoud & Amrouch, 2023).

The two DNN architectures, CNN and Transformers, are both two machine learning methods that can be researched to be used with our problem statement as they can both learn and be trained on the features of a particular object to perform re-identification capabilities. However, if performance comparisons between various models with different architectures are to be made, it is important to identify and understand what kind of metrics can be used to evaluate the models. Based on the type of task that is being performed by the machine learning model, there are various different metrics that can be used to assess the performance of the model (Santafe et al., 2015). The research problem for this work is regarding the possibility of performing re-identification of a *Chelonia mydas* turtle based on the images of the carapace in short- and long-term scenarios. This task is a multi-class classification problem as the deep learning models are required to identify which image belongs to which turtle. For classification problems, accuracy is widely used as it is easy to compute and understand while also being a good choice of metric when class labels are relevant and the proportion of data samples for each class is similar (Santafe et al., 2015). Accuracy provides an overall measure of how well a model is at making correct predictions on an entire set of data (Grandini et al., 2020). This is important and useful for evaluating the models that is to be designed and developed later on as it helps in identifying which model is better overall in making correct predictions of which carapace image belongs to which turtle during the short- and long-term re-identification tests. Accuracy is also important in evaluating the performance of the model during the training phase. Several studies that involve comparing different deep learning models have utilized training and validation accuracy as a metric for evaluating model performance (Hindarto, 2023; Matapurkar, 2023; Pant et al., 2023).

Besides accuracy, other metrics such as training loss and validation loss can also be an important metric to consider when evaluating deep learning models. There are various

types of loss functions but their usage and application differ depending on the type of task that is being performed by the deep learning model such as whether it is a regression task or a classification task (Terven et al., 2023). For classification tasks such as the research problem of this thesis which is a multi-class classification problem, the cross-entropy loss function is a major factor in determining how many training epochs is required to be used to train a model (Wu & Liu, 2009). Beneficial information such as early stopping can be determined when patterns of overfitting in the validation loss and training loss are detected (H. Li et al., 2024). This helps in determining the ideal number of epochs that a model has to be trained to reach a certain performance. More training epochs directly means more training time and computing resources spent and thus is an important metric to consider when evaluating deep learning models. On the other hand, training time can also be an important metric to consider when evaluating deep learning models. The constraint of having limited computational resources and tight work deadlines is a key factor for the optimization of deep learning models in terms of resource wastage (Zancato et al., 2020). The training time refers to the amount of time taken to build the classification model and depending on the method or architecture of the model, the length of the training time can be shorter or longer (Tetila et al., 2020). As such it is a key performance indicator of which model is better when considering the utilization of minimal hardware. Various studies on machine learning models have utilized training time as a valuable metric when evaluating their models (AlZoman & Alenazi, 2021; Loukili et al., 2022; Wang et al., 2018).

After reviewing and understanding more information about deep learning, object detection, and the important metrics to consider when evaluating deep learning models, it is also important to review other works that have been done regarding the use of deep learning models as tools in wildlife re-identification. There have been several works researched and

published on the use of these deep learning architectures as tools for wildlife re-identification in ecology and it is certainly interesting and important to review these implementation examples for useful insights and knowledge that may provide valuable guidance for this research work.

2.5 Deep Learning Tools in Wildlife Re-Identification

Currently researchers around the world are looking to develop machine learning models that can re-identify individual animals of a species. Re-identification of animals in ecological research, as discussed before in the previous sections, is a crucial process as it can provide information and understanding on the population, distribution, and migration of the individuals of an animal species. Without monitoring the population of the animal species being studied, it can be difficult for ecologists to implement interventive measures to manage the species for its benefit (Tuia et al., 2022).

Conventional methods such as the ones mentioned in the previous sections, requires considerable amount of human labour as field workers are still required capture the animal to attach, retrieve, and re-identify the tags that have been attached to the animals. Besides that, the presence of humans during these processes also introduces risks to the wild animals, their habitat, and to the field workers themselves (Tuia et al., 2022). Other options such as utilizing DNA profiling for re-identification are also largely unscalable, especially for animal populations that are widely distributed across a wide geographic area. Therefore, it is ideal that machine learning be used with another form of monitoring tool that is also widely used in ecology, that tool is a camera trap with its capabilities to capture photographic and videographic data. Machine learning models learn by taking in information in the form of input data and for this field of work, that can be performed by utilizing data that has

previously been obtained by animal monitoring work that is done using cameras such as camera traps.

Camera traps have been and still are one of the more popular methods of monitoring and examining the behaviours of animals. Camera traps have been used to monitor various types of wildlife such as bears (Proctor et al., 2022), possums (Moore et al., 2021), tigers (Mohd-Azlan & Sharma, 2003) as well as underwater fishes (Bilodeau et al., 2022). Camera traps that are motion triggered, equipped with sensors, solar rechargeable batteries, and panels, enable ecologists to film animals at remote sites without regular human visits that can disrupt the natural habitat (Kucera & Barrett, 2011). Now in the current era of internet-connected edge devices, it is possible as well to be able to stream live video and images directly from such cameras to provide real-time monitoring capabilities. Camera traps have also been noted as being more effective and cheaper than other forms of labour-intensive monitoring methods (Welbourne et al., 2015). As technological innovations improve and the cost of cameras and camera trapping equipment decreases, camera-based techniques have grown in popularity. Cameras such as camera traps are also able to capture picture and videographic data on individual animals, but also on large groups of animals that are within its camera detection range and capture frame (Caravaggi et al., 2017).



Figure 2.6: A pole-mounted trap camera positioned on a beach habitat (August 2022).

Trap cameras such as the one in Figure 2.6 has also been used to study *Chelonia mydas* sea turtles. Images of juvenile *Chelonia mydas* sea turtles have been collected and used for re-identification purposes by using the data that is captured off the trap cameras that are mounted on-site at beach habitats where the juvenile sea turtles can be found roaming on the sands (Hipiny et al., 2018). However, in that research, they utilized the Histogram of Oriented Gradients (HOG) technique for re-identification of the juvenile *Chelonia mydas* individuals instead of using AI-based deep learning models. HOG is a technique that uses the evaluation of feature descriptors in the form of image gradient orientation that is extracted from all localized grids of an image (Dalal & Triggs, 2005). Deep learning-based machine learning models on the other hand, utilizes AI architectures such as neural networks and attention mechanisms instead. Currently, the research is still ongoing regarding their possibility and capability to be used for the re-identification of animal individuals. However, with the data provided by cameras such as the camera traps used in monitoring and tracking of animals in ecology, datasets to train deep learning models is becoming increasingly available.

There are several examples of how AI models are currently being used for animal re-identification in ecology that can be looked into to understand how it is applied and utilized. Currently, extensive efforts are being made to create datasets of various animals into a database for animal re-identification and tracking (Kuncheva et al., 2022). However, as is normal with using trap cameras to capture video data, animals in such datasets can be incorrectly labelled especially when considering herd and communal type animals, where multiple individuals are situated very close together which makes annotating and re-identification difficult, cows and sheep herds for example. This is called label noise, but a carefully curated dataset that is done by a reliably skilled annotator can reduce the amount of incorrect labelling in such datasets thus ensuring that the dataset can be suitably used by researchers.

Once the dataset of a researched animal is collected, annotated, and labelled correctly; researchers can begin training their deep learning-based models on those datasets. CNN-based deep learning models have been used for individual re-identification of marine animals such as Koi fishes (Kuncheva et al., 2023). However, in that study, the authors also applied simple linear models on the same dataset and concluded that non-machine learning models can perform on the same level or even better than deep learning models. Besides that, land-based mammals such as giraffes have also been studied in the interest of conducting re-identification of individual giraffes by using CNN-based deep learning models (Miele et al., 2021). This is possible because, for the animal species of *Giraffa camelopardalis*, also known as giraffes, the coat pattern on their skin is uniquely different for each individual and is one of the stable visual features that can be used to re-identify giraffes in the wild or captivity (Lee et al., 2018). The authors of the paper first created a dataset of 5,019 images that were manually cropped out from 3,940 photographs captured

with their digital single-lens reflex (DSLR) cameras off wild giraffes. Afterward, a pre-trained CNN-based deep learning model was trained on that dataset and then tested with test data. They managed to achieve a 90% top 5 accuracy but remarked that to fully verify that the results were accurate, human verification was still required to filter out any false positives and true negatives in their research. Other than that, the re-identification of domesticated farm animals such as cows has also been attempted on multiple occasions with CNN deep learning-based models (Bello et al., 2020) and also with attention-based deep learning models (Chen et al., 2022). The former research utilized the body pattern of the individual cows as a stable feature while the latter research utilized the cows' facial pattern as a stable feature for re-identification. This showcases that various parts of the animal species can be used for re-identification purposes given that the feature is stable and is also unique enough for re-identification to be possible. Both the research papers mentioned have managed to achieve results of 90% and above consistently for re-identification of the cows. Considering this and the previous examples that have been reviewed in this section, they have demonstrated that re-identification of animal individuals are indeed possible by utilizing deep learning models. Next, discussions can be made on the use of deep learning tools for the re-identification of *Chelonia mydas* sea turtles to examine the possibility of using it with photo identification methods on the stable features that are presently available and uniquely individual for each turtle.

Table 2.1: Table of summary of deep learning tools in wildlife re-identification

Author	Year	Title	Summary
Kuncheva et al.	2023	An experiment on animal re-identification from video	The study focuses on experimenting with 25 different classifier models of various types which also includes one CNN Deep Learning model for the purpose of re-identifying Koi fishes. The authors noted that the simple linear models have much favourable results compared to the rest but remarked that the use of deep learning can be explored further.
Miele et al.	2020	Revisiting animal photo-identification using deep metric learning and network analysis	The study proposes using CNN deep learning models with object detection to perform re-identification of Giraffe individuals. The authors managed to reach good accuracies of up to 90% using the CNN model but mention that it may be possible to improve the results by photographing the giraffes in as many conditions as possible instead of their currently limited angles.
Bello et al.	2020	Image-based individual cow recognition using body patterns	The study proposes a system to re-identify individual cows based on their body patterns by using a deep learning model. Their methods managed to successfully achieve up to 89.95% accuracy on their testing data.

2.6 Existing Usage of Deep Learning Tools Used for *Chelonia Mydas* Re-Identification

Based on the previous section, it is established that re-identification of animals is possible with deep learning models; whether it be by using CNN models with or without attention, they both can be used to re-identify animal individuals. For *Chelonia mydas* sea turtles on the other hand, several works have been done to integrate deep learning models for the purpose of re-identifying the individual turtle based on its stable features. These stable features in the case of the *Chelonia mydas* species of sea turtles are unique to each individual and can be used to re-identify individual turtles. The stable features of the *Chelonia mydas*

sea turtle that have been used for this purpose are the facial pattern (scutes), and the carapace pattern of the individual turtle.



Figure 2.7: A Green Sea Turtle with distinct facial patterns (Witherington et al., 2006).

Several researchers have utilized the facial scutes of *Chelonia mydas* sea turtles as a photo re-identification criterion to determine the individual turtle. This type of identification is called photo identification and is meant to be incorporated as a minimally invasive method to re-identify the turtle individuals (Araujo et al., 2016). One of the reasons why facial scutes are used is because it has been identified that the facial scute pattern of *Chelonia mydas* turtles can remain stable and unchanged for up to 11 years but pigmentation color changes of the face of the individual turtle may shift over time (Carpentier et al., 2016). Re-identification can be done easily and manually by a human reviewer (Chew et al., 2015) or by utilizing pattern-matching software such as I³S Pattern (Araujo et al., 2016; Reisser et al., 2008), HotSpotter (Neves-Ferreira et al., 2023), and NaturePatternMatch (Long & Azmi,

2017). Re-identification of *Chelonia mydas* sea turtles by utilizing the facial scute pattern is demonstrated to be reliable and was regarded as a good alternative to tag-based identification. However, the software-based photo re-identification method listed above uses software algorithms such as scale-invariant feature transform (SIFT) and speeded-up robust features (SURF) instead of trained deep learning models. As such they can be used directly for image comparison in re-identification, unlike deep learning models which require a dataset to be created first to train the model before re-identification is possible.

Facial re-identification of *Chelonia mydas* individuals from images utilizing deep learning models is a new endeavour and studies on the feasibility and reliability of this method are limited. However, there has been recent attempts made by researchers to incorporate deep learning in the re-identification process of *Chelonia mydas* individuals by its facial scute patterns. Convolutional Neural Network-based models have been used such as AlexNet CNN models in conjunction with a multi-class support vector machine (SVM) and they managed to produce an overall classification accuracy of 62.9% (Yussof et al., 2020). It must be noted as well that in this research, the deep learning CNN model, AlexNet, was only used for feature extraction on the facial scute patterns of the *Chelonia mydas* turtles and classification was only performed using error correcting code approach via a multi-class SVM. However, its implementation still utilizes a deep learning model for feature extraction and showcases the utility of deep learning models in the re-identification process of *Chelonia mydas* individuals using facial scute patterns. Artificial Neural Networks (ANN) have also been implemented instead of Convolutional Neural Networks (CNN) for re-identification via facial scute patterns of *Chelonia mydas* individuals by comparing newly taken images of an individual and comparing it with existing images that are catalogued in a database (Carter et al., 2014).

However, after reviewing several works of literature on deep learning model application for re-identification of *Chelonia mydas* sea turtles, it has been found that the amount of research done utilizing the unique facial scute pattern of the turtle far outnumber the amount of research done utilizing the carapace pattern. There is a lacklustre amount of literature on the carapace pattern being used. Although the carapace pattern is indeed unique for each turtle and is a stable feature that can be used to re-identify *Chelonia mydas* individuals, especially for long-term scenarios (Tabuki et al., 2021).

Table 2.2: Table of summary on works utilizing images for turtle re-identification

Author	Year	Title	Summary
Araujo et al.	2016	Using minimally invasive techniques to determine green sea turtle <i>Chelonia mydas</i> life-history parameters	The study proposes the combine use of photo-identification and photogrammetry to determine life-history parameters of free-ranging <i>Chelonia mydas</i> sea turtles. In the study, the method of re-identification of an individual sea turtle is by using the facial features of the sea turtle, manual observation of the researchers, and the pattern matching software “I ³ S Pattern”. The study concludes that using photo-ID is a valid approach that can be cost-effective while also minimizing the risk of injury to the turtle.
Carpentier et al.	2016	Stability of facial scale patterns on green sea turtles <i>Chelonia mydas</i> over time: A validation for the use of photo-identification method	The study investigates the stability of facial scale patterns and shapes of foraging green sea turtles and the potential of being reliable for long-term identification. The study concludes that facial scale patterns can be used for the photo-ID approach for the re-identification of sea turtles.
Reisser et al.	2008	Photographic identification of sea turtles: Method description and validation, with an estimation of tag-loss	The research focuses on using mark-recapture technique and the facial profile photographs of individual turtles for identifying recaptured individual <i>Chelonia mydas</i> and <i>Eretmochelys imbricata</i> sea turtles. They were able to visually re-identify individual sea turtles

			based on the facial photographs of the turtles by manual inspection.
Neves-Ferreira et al.	2023	Photo-identification shows the spatio-temporal distribution of two sea turtle species in a Brazilian developmental foraging ground	The study uses photo-ID to investigate the spatio-temporal distribution of <i>Chelonia mydas</i> and <i>Eretmochelys imbricata</i> sea turtles. The pattern-recognition algorithm “Hotspotter” and photographs of the lateral head scale pattern of the turtles are used. The study concluded that it is possible to use photo-ID to detect resident individuals of sea turtles that is occupying a certain area.
Long and Azmi	2017	Using photographic identification to monitor sea turtle populations at Perhentian Islands Marine Park in Malaysia	The study proposes using photo-ID methods to identify sea turtle individuals and to determine their sex ratios, habitat use, and site fidelity. The researchers used a combination of manual visual matching and the software “NaturePatternMatch” to identify individuals. The facial feature of the turtle is used for re-identification while the carapace feature is used for sex identification. Of the total, only 4.3% of sea turtles cannot be identified. The researchers noted that the quality of photographs is critical and that training can help photographers take better photographic data.
Yussof et al.	2020	Photo identification of sea turtles using AlexNet and Multi-class SVM	The study proposes the use of photo-ID in identifying sea turtles using a modular approach that includes AlexNet and a Multi-class SVM. Photographs of the facial scute pattern of turtles were used and the research managed to reach an overall accuracy of 62.9% for re-identification of sea turtle individuals.

2.7 Summary

After reviewing the various literature in this section on the importance of animal monitoring, the current methods of animal tracking via physical tags, cameras such as camera traps, and the use of deep learning-based models for the re-identification of animals,

especially *Chelonia mydas* sea turtles; the conclusion is that there is indeed a clear benefit of applying deep learning models in re-identification of animals and *Chelonia mydas* sea turtles. It is reviewed in this chapter that there exist several downsides to the current method of re-identifying animals and sea turtles with physical tags such as GPS tags and Inconel tags; and that cameras such as camera traps can provide deep learning models with datasets that can be used to train those models. Therefore, it is acknowledged that there is insufficient study on the use of the *Chelonia mydas* carapace for deep learning-based re-identification and that there are already quite a few successful implementations of deep learning-based models used for re-identification of *Chelonia mydas* individuals albeit by utilizing the facial scute patterns instead. Due to this research gap, it is necessary to pursue the utilization of the *Chelonia mydas* carapace pattern as a feature for re-identification compared to the facial scute pattern as the carapace feature has also been examined to be suitable for photo identification to be possible in short and long-term scenarios.

CHAPTER 3

METHODOLOGY

3.1 Introduction

In this chapter, the research flowchart is explained, and the methods used to train, test, compare, and evaluate a deep learning model with attention mechanism and several CNN models for the re-identification of *Chelonia mydas* individuals via carapace pattern in a short-term direct re-identification scenario and a long-term scenario with simulated barnacle growth to indicate natural change in the carapace pattern of the sea turtles is discussed. The process used is detailed beginning from the start of the data collection and preparation phase, as well as the creation of the training dataset and its format used for training. The model that has an attention mechanism is compared with the traditional deep learning models based off the CNN architecture namely AlexNet, SqueezeNet and ResNet50 to evaluate the two different architectures in relation to several metrics such as accuracy, loss, and test sets re-identification results.

Discussions on the code functions, software, hardware, and ethics used in the methodologies that were used in this research are provided in this chapter. This is important because, compared to working with data that is readily available such a public online dataset; this research involves actual physical work with living free roaming wild animals that must be considered especially since the *Chelonia mydas* species of sea turtles are endangered. This chapter covers the above while the result of the methodologies used is covered in the next chapter.

3.2 Research Flowchart

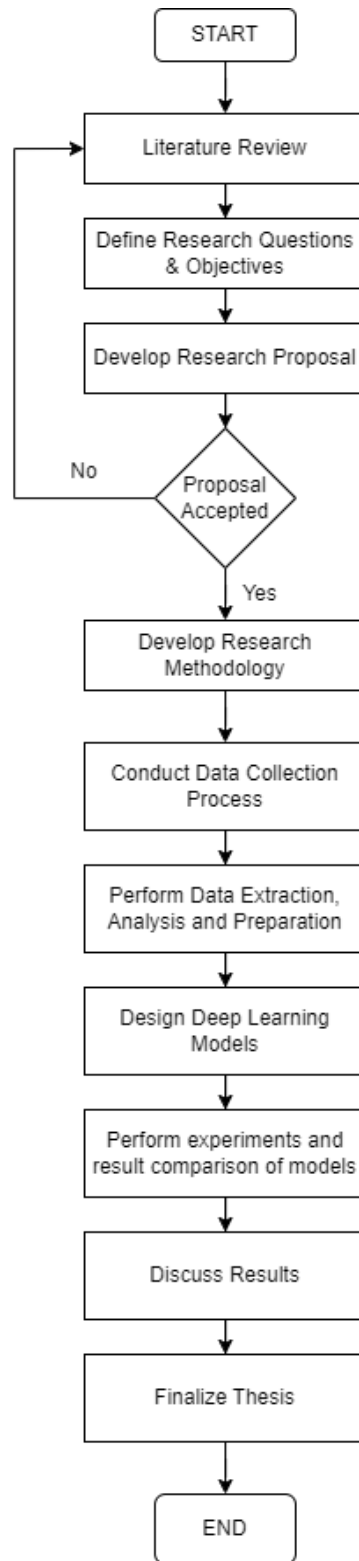


Figure 3.1: Research Flowchart of the study

The flowchart in Figure 3.1 is the research flowchart of the study and is used to guide the research work from start to end. It starts with conducting a literature review to understand and collect information regarding the works that has been done in the field of AI, deep learning, object detection, ecology, and the relation of all of them in the re-identification process of animals. Literature review on the subject animal of the study which is the *Chelonia mydas* green sea turtle is also done to understand the subject animal better. After the literature review process is done, the research objectives and questions can now be defined. The research proposal is then developed and depending on the results of a proposal defence evaluation, the research flow will either progress to the next step or to return to the literature review step for refinement and then re-evaluated in another proposal defence.

Once the research proposal passes the proposal defence, the research methodology is developed. The main focus of the study is regarding the use of a deep learning model with an attention mechanism to re-identify *Chelonia mydas* sea turtles' individuals in short- and long-term scenarios and as such several deep learning models without attention mechanisms and a model with attention mechanisms is to be developed and their performance are to be compared. The research methodologies developed include the methodology for collecting, automated extracting utilizing YOLO v7, and preparation of the *Chelonia mydas* data, as well as methodologies for developing the deep learning models and the method of comparing the performance of the models in short- and long-term scenarios.

After conducting the steps that have been planned out in the research methodology following the sequence as presented in the research flowchart Figure 3.1, the results of the experiments are recorded. The deep learning models are then compared in terms of several metrics such as accuracy and loss metrics during training, the efficiency of the model in terms of training time, and the accuracy in re-identification in short- and long-term scenario.

The results are discussed and summarized. The advantage or disadvantage of the deep learning model with attention mechanisms is discussed and the thesis is concluded. The more in-depth details of the research methodology can be found in the following subsections.

3.3 Sampling and Data Collection Techniques Innovated

Before beginning to train the deep learning models, it is first required that an image dataset of several different unique *Chelonia mydas* individuals is obtained. The data used in this research is entirely new and was acquired during several scheduled sampling sessions on Talang Besar Island off the coast of Sarawak. The research partner of this work, the Sarawak Forestry Corporation (SFC), a state statutory body governing the conservation efforts of the state's biodiversity and protected areas; has provided clearance and permits to access the protected areas where the *Chelonia mydas* sea turtles are known to come ashore for nesting. The areas permitted in the permit are Tanjung Datu, Satang Island, Talang Besar Island, and Talang Kecil Island. However, after visits to Tanjung Datu and discussions with the park wardens of Satang Island, Talang Besar Island, and Talang Kecil Island, it was concluded that Talang Besar Island is the most suitable island to conduct sampling activities. The reasons are due to several factors as listed below:

- Population of *Chelonia mydas* individuals: It was recorded and determined that the population density of *Chelonia mydas* individuals that land ashore during nesting season between July and September were the most populated at Talang Besar Island. Tanjung Datu, Satang Island, and Talang Kecil Island, only had at most several hundreds of turtles combined a season; Talang Besar Island by itself can reach up to a thousand or more depending on the weather conditions.

- Geographic suitability for sampling activities: Talang Besar Island has a large sandy beachhead that is concentrated outwards towards the sea. As such, there is more space to anticipate, observe, locate, and approach *Chelonia mydas* sea turtles that are digging in the sand to nest. The other beaches are either small or are elongated along the coastal shoreline meaning more traversal is needed to locate and find the sea turtles.
- Minimal interference/ No Tourist Presence: Talang Besar Island is a totally protected area that is closed off to the public. Access to the island is only permissible to the park rangers and wardens, and to researchers and scientists who are conducting ecological research on the biodiversity of the local fauna and flora of the area. This means Talang Besar Island is unlike the other locations that can be packed or congested with tourists. Talang Besar Island is free of outside human interference that might disrupt the sampling activities or that can disturb the natural environment of the *Chelonia mydas* sea turtles.



Figure 3.2: The red square showcases the restricted zone where visitors are not permitted to be within 2.8km of the national park without special arrangements with SFC (October 2023).

After an inquiry session with the park rangers and wardens, it is informed that the restricted zone helps keep visitors, and commercial or recreational boats away from the island as best as possible. There are several reasons for this but among them are that the boats can scare away the female *Chelonia mydas* individuals from coming ashore for nesting. Besides that, these boats can and may cause confusion and disorientation to young juvenile hatchlings. These *Chelonia mydas* juvenile hatchlings orient themselves to swim toward the sea by being attracted to the reflected moonlight over the ocean but the floodlights from fishing or recreational boats can confuse them (Truscott et al., 2017). Other than that, the low presence of human activity on Talang Besar Island ensures that the adult female *Chelonia mydas* turtles are not intimidated to land ashore for nesting. Further interviews with the park rangers of the island explained that the less visible the human activity done on the beaches, the more likely the turtles are to come ashore. The insight provided by them is that it is true that the female *Chelonia mydas* only come out for nesting when it is dark such as at night or in the early hours of morning, but the turtles are also already scouring the waters of the island during daytime to scout the beach conditions for predators. The less visible the human presence, the better the conditions to facilitate a larger chance for female *Chelonia mydas* sea turtles to come ashore for nesting, which is very beneficial for the research as more data can be collected during the sampling period done on the island.

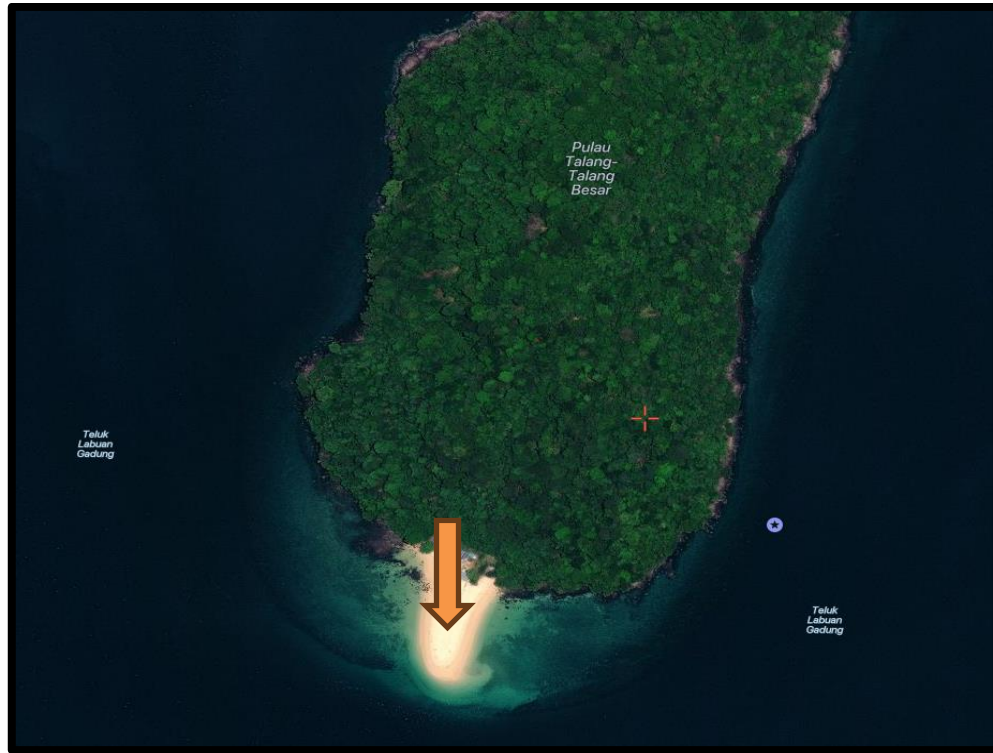


Figure 3.3: High-resolution satellite image of Talang Besar Island's beachfront (October 2023).

One of the plans for the data collection process was to set up a vantage point where most of the nesting beach areas are easy to observe and reach for sampling. Talang Besar Island is naturally perfect as the beachfront is elongated towards the sea and not along a coastline. As can be seen in Figure 3.2, the beach is narrow but lengthy and it is possible to stay on a single point to be able to see and observe a large swatch of the beachfront. The orange arrow in Figure 3.2 is where the monitoring station is set up and the equipment is prepared before the beginning of the sampling process for each female *Chelonia mydas* turtles that come ashore. The presence of the sea turtles is determined by observing for the spray of sand kicked up by the female individuals that are digging their nests on the beach.

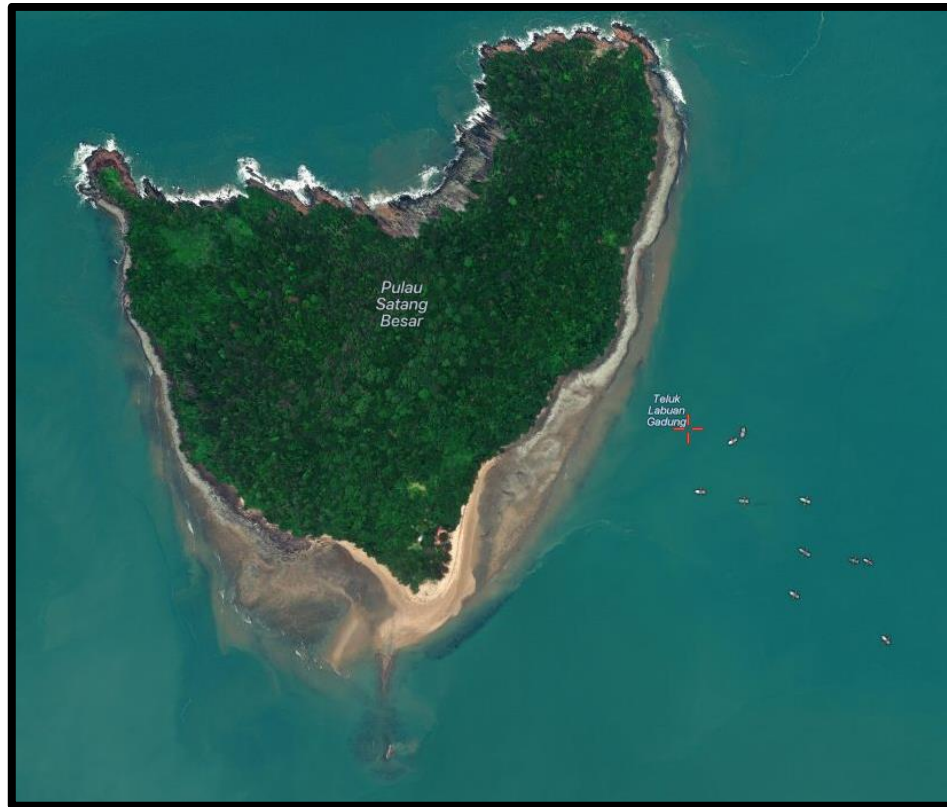


Figure 3.4: High-resolution satellite image of Satang Island’s beachfront (October 2023).

Based on Figure 3.3 of Satang Island, notice the stark difference in beach layout between the two. Also, note the significant presence of boats around Satang Island that are quite close to the beach area on the bottom right. These factors are what makes Talang Besar Island as the best place to conduct the research sampling activities instead of the other locations. The chance to obtain good amounts of data is of utmost importance to ensure that the deep learning models can learn well for re-identification purposes later.

Now that the sampling location has been defined, the hardware equipment used also needs to be determined as well. For this research, it was initially opted to use an off-the-shelf commercially available trap camera, the Bushnell NatureView HD trap camera, that can record videos or take pictures in low light environments either manually or automatically via

the use of motion sensors. However, after performing sampling activities on 15 different *Chelonia mydas* individuals, it was reviewed that the footages captured are grainy and prone to be not centred on the turtle's carapaces. The limitations of this method are readily apparent when reviewing the footage captured in detail.

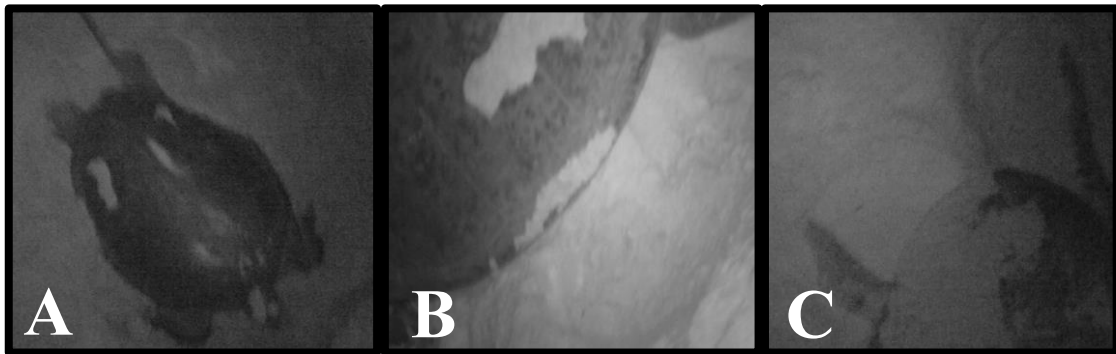


Figure 3.5: Three different sample images of a turtle's carapace taken from 3 different female *Chelonia mydas* individuals with Bushnell NatureView HD (August 2022).

As can be seen with the results of the initial sampling period, the images are considerably grainy and often miss capturing the whole carapace section of the individual turtles. The specific reason why the data quality captured with the Bushnell NatureView HD is dissatisfactory is because of the inherent hardware capabilities of the trap camera itself. Take for example section A in the figure above, the image appears grainy but the whole carapace of the turtle can still be identified. However, the detail of the turtle's carapace is difficult to ascertain as the image appears to be very dark. This is specifically due to the unremarkable infrared (IR) illumination capabilities on this trap camera for its complementary metal oxide semiconductor (CMOS) sensors to capture clear details of any object's surfaces. The Bushnell NatureView HD also suffers from a limited field of view (FOV) due to the nature of its camera lenses providing only 50° of FOV, thus requiring the trap camera to be positioned directly above the female *Chelonia mydas* individual to be able

to capture its carapace or risk capturing it partially such as example B and C in Figure 3.4 that is not usable as data because of its limited coverage of the desired region of interest. Besides that, there is no visual monitoring of what is being captured by the trap camera during sampling and recording because it uses the A/V output jack that requires a Radio Corporation America (RCA) input-supported secondary monitor which is not provided as is with the unit used. As such, the trap camera distance had to be estimated and adjusted variably by moving it up and down when performing sampling activities out at the beach to ensure a high chance that the camera would be able to capture the full carapace of the turtle. This is very problematic as in terms of ethics, this type of movement and camera handling disturbs the turtle and induces stress for the nesting *Chelonia mydas* female individuals being sampled. They would try to move away during the sampling as the camera has to be positioned very close to them. This is unsuitable as the work in this research must consider the appropriate and correct ethics when working with wild animals as informed by the park rangers and wardens. Due to these factors, innovation is required to design and implement a better camera setup to capture the data that is needed for the research. The current Bushnell NatureView HD is not suitable and because of that, a better setup that can alleviate the problems mentioned needs to be created.



Figure 3.6: The two different hardware and camera setups used in the sampling methods (August 2022).

The second hardware setup involves the use of much more conventional modern hardware such as a closed-circuit television (CCTV) camera, power banks, local area network (LAN) connectivity, and a computing platform, in this case, a laptop, for viewing live feed from the camera. Due to the limitations of the previous Bushnell NatureView HD trap cameras used, an overarching hanging pole with two legs to facilitate the limited optimal angle necessary to record the data is crafted. As shown in section A of Figure 3.5, the distance of the trap cameras from the ground level is very short at about 1.2 to 1.8 meters only thus making it too close in proximity to the turtle when recording as previously mentioned. The two-foot design of the first setup also limits the mobility of the setup to be moved around.

The new setup on the other hand, as shown in section B of Figure 3.5, is much more versatile as it can be more easily moved around the beach requiring a single person to carry the pole due to its single feet design, unlike the previous setup which requires 2 persons. This is important as it allows the setup to be able to be taller than the previous design making it ideal for sampling efforts later.

The camera used for setup in section B is a CCTV camera from HikVision, specifically the 5MP IR Fixed Network Bullet Camera, that comes equipped with 3D DNR noise reduction technology, a 2.8mm lens with 100° diagonal FOV that is more than twice the Bushnell NatureView HD's FOV, and various other built-in capabilities that is beneficial for the sampling activities on Talang Besar Island. Since this new camera used for this second setup uses a network-capable CCTV camera, it is possible to connect a laptop via LAN cable to the camera to fetch, stream, and view live video feed from the camera to a laptop. This was utilized during sampling to obtain even better data than the previous setup as it is now possible to view in real time what data is being recorded and saved into the laptop. This setup is also not limited to the small FOV limitation of the previous setup thus allowing the pole to be hoisted up higher and further away from the turtle during sampling. It is noticed that this causes less disturbance to the turtles as they do not seem to move as much as previously in the first sampling method.

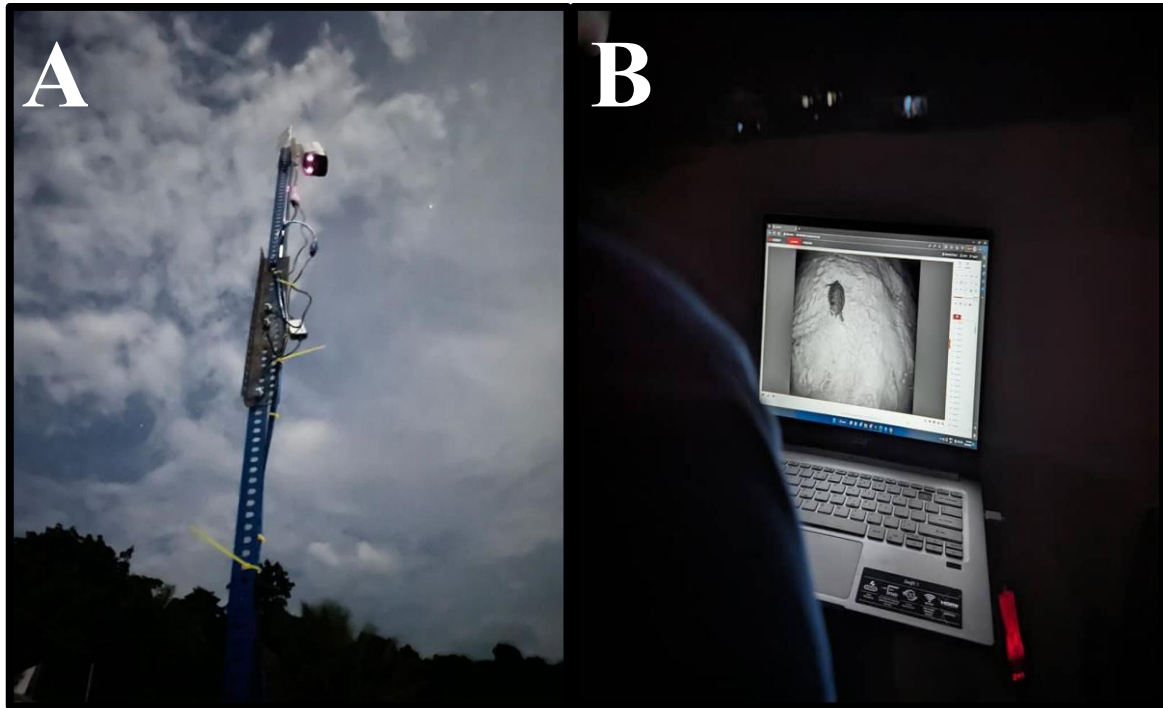


Figure 3.7: Image section A shows the IR illumination of the CCTV camera while image section B shows the live video feed of the CCTV camera of image section A (October 2022)

Sampling activities were done primarily at night during tagging operations with the accompaniment of the park rangers and wardens of Talang Besar Island. Due to ethical reasons as advised by the park rangers and warden, the sampling activities must adhere to the rules of not producing any bright lights at the beach during sampling. That means torches or flashlights should not be used to illuminate the nesting female *Chelonia mydas* individuals when performing the data collection and sampling process. That was one of the considerations as to how the CCTV capabilities were determined for the second setup. As shown in Figure 3.6 which showcases the IR illumination beam of the camera and its live feed in the dark during the sampling process, better data can be obtained as the IR illumination on the second setup is better than the first setup while the live feed ensures that

the camera settings can be adjusted accordingly to ensure the best footage possible is recorded during sampling.

49 different videos of various nesting *Chelonia mydas* individuals were able to be sampled and recorded during flipper ID tagging operations. This process took 4 weeks of sampling and data-gathering activities at Talang Besar Island using the second setup with the help of the park rangers and wardens. It is ensured that both the left and right Inconel flipper ID of each individual sea turtle that was sampled is recorded and kept. Several videos were recorded of individual turtles that were roaming about the beach area. However, since these turtles were not nesting and the rangers did not attach tags to them, this data is only kept as supplementary data. Ideally, for this research, it should be ensured that each turtle recorded is unique as the carapace recorded should correspond to each turtle based on its ID to avoid double sampling of the same individual. It was explained by the park rangers and warden that there can be multiple visits from the same turtle during a single nesting season and as such this ID tracking consideration was valid as to avoid recording the same individual twice by accident.

3.4 AI-Based Image Filtration, Analysis, and Extraction with YOLOv7

As mentioned in the previous section, there are a total of 49 videos recorded with the second setup. They are all IR-illuminated night mode videos of 36 different nesting female *Chelonia mydas* individuals that showcase multiple frames of the full carapace of the sea turtle. For this research, only a cropped-out segment of each frame where the carapace is shown with decent clarity is required. Other elements of the videos where the image frames show no turtle presence or are too blurry should be omitted. Therefore, it is important to perform analysis through all the image frames of the videos to crop out the necessary desired

carapace section of the turtle to create the dataset needed to perform training on the deep learning models later. But herein lies a particular problem, to perform manual cropping for all 49 videos is a very laborious process as the total duration length of all the 49 videos combined is about 1 hour and 38 minutes long. After examining the metadata of all the videos, it is calculated that approximately 120,000 image frames need to be individually analysed and hand-cropped manually whenever a good decently clear visual of the turtle carapace is present. Because of this, it is necessary to establish a method of automatic analysis, cropping, and extraction of turtle carapace images from the 49 videos.

For this method, You Only Look Once (YOLO) v7, a unified neural network model and architecture for object detection that can be used to draw bounding boxes of objects in a classification scenario (Redmon et al., 2016) is utilized. To perform automated cropping of the *Chelonia mydas* individuals that are present in the video frames, an already packaged program and code base by Rizwan Munawar under the GNU public license was used to perform cropping based on the level of confidence that can be manually set (Munawar, 2022). This provides options to tweak the custom model's classification capability to only crop out the object if it is certain that the object within its bounding box drawn is at or above the confidence threshold set. In theory, this would reduce blurry or vague image sections of the video where the turtle shape can be recognized but its carapace details are not present.

```

Start
Set globalcounter = 0
Repeat objectcropping
    if objectdetected
        Get objectcoordinate
        Set objectlocation = objectcoordinate
        Set savepathname = "crop" + (globalcounter + 1) + ".jpg"
        Write croppedobjectimage
    else
        Get nextframe
    end-if
end

```

Figure 3.8: Pseudocode of the cropping function in the YOLO v7 Object Cropping program

The figure above shows in pseudo-code the program that performs the automatic cropping function as it is used in this research as part of the methodology based on the current version of the YOLO v7 Object Cropping program when it is available as a distribution on GitHub. First, an object must be detected in the image frame for automated cropping to occur. If an object is detected, the position of the object in the form of the coordinates of the object within the context of the image frame is obtained. This coordinate is assigned as the location of the object and both the path and filename of the object to be cropped is then set. The cropped image is then written to the file and then the process continues to the next frame and repeats. The program will also accommodate multiple detections of the object in the same frame should it occur. But before automatic cropping can begin, the YOLO v7 pre-trained model that is provided by YOLO must be trained to be able to perform detection on *Chelonia mydas* individuals based on the dataset parameters. The dataset parameters, in this case, are IR-illuminated greyscale images of nesting *Chelonia mydas* individuals. As such, a total of 1,436 manually labelled and cropped images of *Chelonia mydas* individuals from the video

dataset that has been selected via human review and labour is used to create an initial dataset to train this YOLO v7 model into a custom model that can fit the use case. The image annotation SaaS (software as a service) Roboflow is used to create this dataset for training.

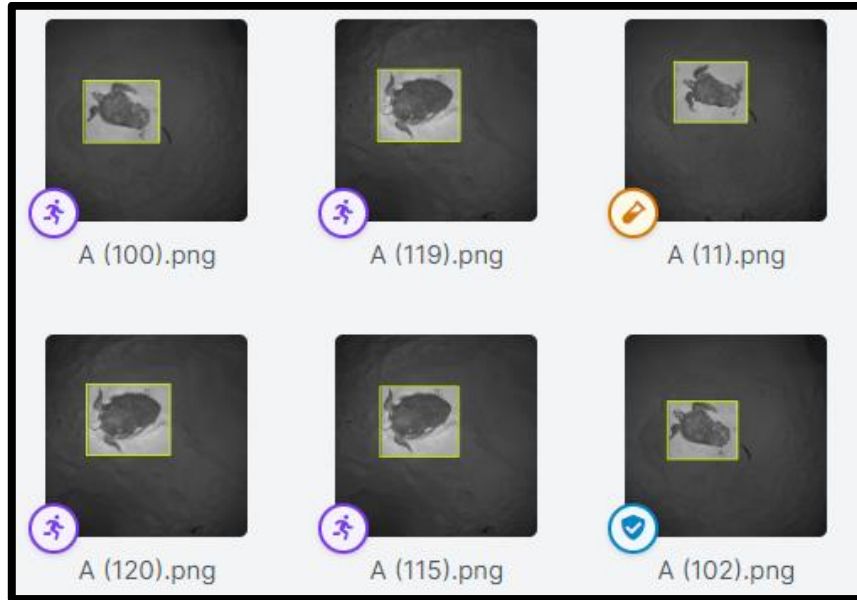


Figure 3.9: Several example images of the annotated dataset by using Roboflow

Once the images were annotated as shown in Figure 3.8, the YOLO v7 pre-trained model was then trained with the dataset at several different epochs. Once the suitable epoch with the optimal accuracy and precision is found (30 epoch), automated cropping is then performed on the remaining videos of the data that was collected from sampling efforts at Talang Besar Island at different confidence threshold configurations and it was found that the confidence threshold of 80% to 90 % works best for processing the videos as it outputted correctly the cropped segment of the image frame where the carapace is identifiable with minimal blurriness. At the end of the automated cropping process, a total of 31,563 cropped-out images of *Chelonia mydas* turtles from 35 different individuals are left. This amount is about 26% of the total image frame initially in all the videos and has been reduced to a manageable amount for a human reviewer to perform final pruning. Since cropping has

already been performed, no cropping is necessary anymore for the human reviewer to perform; only final human verification on the quality of the cropped image is required on the 31,563 cropped images.

After final human verification is performed, only 4,236 images from 12 different individuals are selected to be in the final dataset to be used for the next methods on utilizing deep learning models with and without attention mechanism tests. The remaining images from the leftover 23 individuals are kept for possible future use should the images from the 12 individuals selected for further testing be insufficient. The reason these 12 individuals and their total number of images being 4,236 is selected was to ensure that the hardware utilized in the following methods would be able to accommodate the training requirements for the different deep learning architectures later.

3.5 Creating Datasets with Python SplitFolders

The remaining images are then to be segmented into a proper dataset before any training of the deep learning models can begin. The traditional way of doing this is by performing a three-part split on the data; by categorizing each into the training set, the validation set, and the test set which is considered a good approach that can be used to build good predictive models (Muraina, 2022). These parts are also known individually as a sub-dataset, and each of the three plays a specific role in the design aspect of the deep learning models. This method is known as the holdout validation approach, whereby the available data is partitioned into the three subsets for training, for performing evaluation and for testing. They are determined at the start and do not change; which is vastly different than another also rather popularly used method known as the cross-validation approach where the subsets are instead crossed over one another in the successive rounds of model training and

validation (Awwalu, 2019). To further clarify the use of the sets in the holdout method, the training set is used to train the model's parameters, the validation set is used to define the hyper-parameters that control how the model is being trained, and lastly, the test set is used to estimate the skill level of the final tuned model (Kahlout & Ekler, 2021). The Python package, SplitFolders version 0.5.1, is utilized to perform randomized splitting of the image data into the 3 sub-datasets mentioned. For this research, it was decided to use only the data from 12 different *Chelonia mydas* sea turtles; this thus ensured that each sea turtle individual is a labelled class of its own and is also a class folder that must be present in the 3 different sub-folders of the main dataset folder.

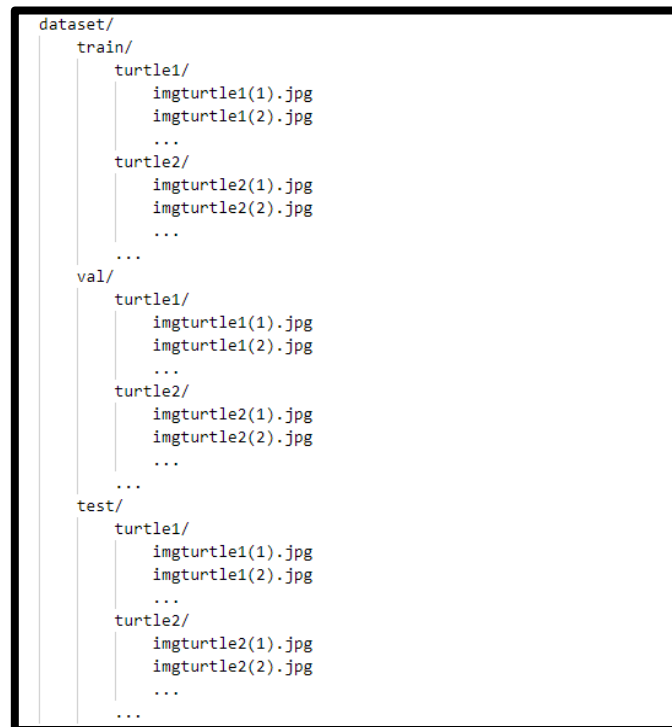


Figure 3.10: Example folder structure of the dataset

As shown in Figure 3.9, the main dataset would have 3 sub-folders and they are the training set (train) folder, the validation set (val) folder, and the test set (test) folder. Each of these 3 sub-folders has 12 inner sub-folders that are labelled correspondingly to the 12

Chelonia mydas sea turtle individuals that have been chosen. Inside these 12 inner folders is where the cropped-out data images are stored.

```
pip install split-folders
import splitfolders
splitfolders.ratio('Dataset',output="12turtlesplit",seed=69,ratio=(.7,0.2,0.1))
```

Figure 3.11: Code snippet of SplitFolders v0.5.1

As can be seen from the figure above, the distribution of the dataset for the 3 subsets are divided in the ratio of 7:2:1 which allocates the data to 70% for training set, 20% for validation set and a remaining 10% for the test set. The randomization seed is set to 69 and this value is recorded so that the randomization of the data split can be replicated by re-using the same seed value should it be required. As mentioned previously, this method uses the hold-out validation approach and as such are not re-substituting or re-using the data from the training set for validation and testing to avoid overfitting which can occur due to re-using the same data which is known also as “re-substitution validation” (Yadav & Shukla, 2016). At the end of this process, the final compiled and organized dataset is obtained and can be used for the next method which is training the deep learning models.

3.6 CNN and Attention-Based Deep Learning Models for Re-Identification

For this research, a deep learning model with an attention mechanism that can perform re-identification on the *Chelonia mydas* sea turtle individuals must be designed. As a metric for comparing its performance, several CNN based deep learning models are also trained. An AlexNet, a SqueezeNet, and a ResNet50, deep learning models are trained to

compare with the attention-based deep learning model to see if the attention-based model has any notable advantages or benefits compared to the traditional CNN-based approach.

To begin, an attention-based deep learning model is developed. For this project, it was decided to utilise the use of Transformers, which is a type of model architecture that relies on attention mechanisms to establish global dependencies between its inputs and outputs instead of utilizing convolutions like in CNN models or sequence-to-sequence alignments like in RNN (Recurrent Neural Network) models (Vaswani et al., 2023). Initially developed by Vaswani et. al. in their breakthrough paper “Attention Is All You Need” for NLP (natural language processing) tasks, the attention-based deep learning model in this research is developed based on the follow-up work on Transformers done by Dosovitskiy et. al. that has managed to adapt the Language-Transformers that is now a standard for natural language processing to be workable with vision-related tasks. This new self-attention-based architecture of transformers for computer vision is known as vision transformers and can utilise the benefits of self-attention like how the original transformer architecture design is also capable of (Dosovitskiy et al., 2021).

To build the vision transformer model with self-attention capabilities, PyTorch, the open-source machine learning development framework built on the Python programming language which also incorporates the open-source machine learning library of tools and packages from Torch, was used. Other necessary tools and software packages were also downloaded to ensure that PyTorch would be able to run on the hardware. The two important hardware components that are especially crucial for this phase is the CPU and GPU. The CPU and GPU used for the research is an AMD (Advanced Micro Devices) 5600X 6-core desktop processor CPU and an Nvidia RTX 3060TI GPU with only 8GB of DDR5 memory.

It is important to be noted on the physical limits of the hardware as the design of the deep learning models especially for vision transformer models are highly dependent on the capabilities of the hardware. As it is known, the more complex the model is, there is an exponential increase in the time taken for model training and a more demanding need for better hardware (Tufail et al., 2023). As such the model parameters for the vision transformer model with self-attention mechanisms must be able to conform with the capabilities of the CPU and GPU while making sure that the parameters used are also replicable with the design of the AlexNet, SqueezeNet, and ResNet 50 CNN-based models.

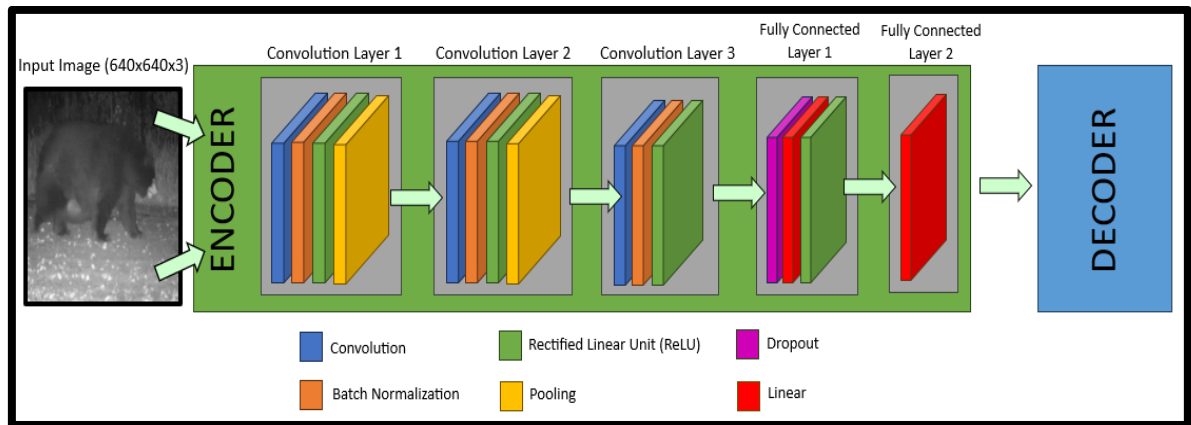


Figure 3.12: Example of an AlexNet CNN encoder

In vision transformer architectures, images are not loaded directly as inputs and undergo convolution for feature map extraction like traditional CNN models do, instead, images in vision transformer architectures are loaded as a split of several fixed-sized patches that are then projected linearly with their positions embedded to be fed to a transformer encoder. Understanding this is important to understanding how attention is achieved in vision transformers. Based on Figure 3.11 which shows a standard implementation of a CNN model architecture for vision-related tasks, it is seen that the input image is loaded at the start usually as a 2-dimensional array with a set dimension (in this case it is 640x640) and type

(x3 which means RGB) with (or without as in Figure 3.11) some pre-processing done before being passed on to the encoder to perform convolution at the successive convolutional layers within the encoder. This is different than how vision transformers' encoders handle their input.

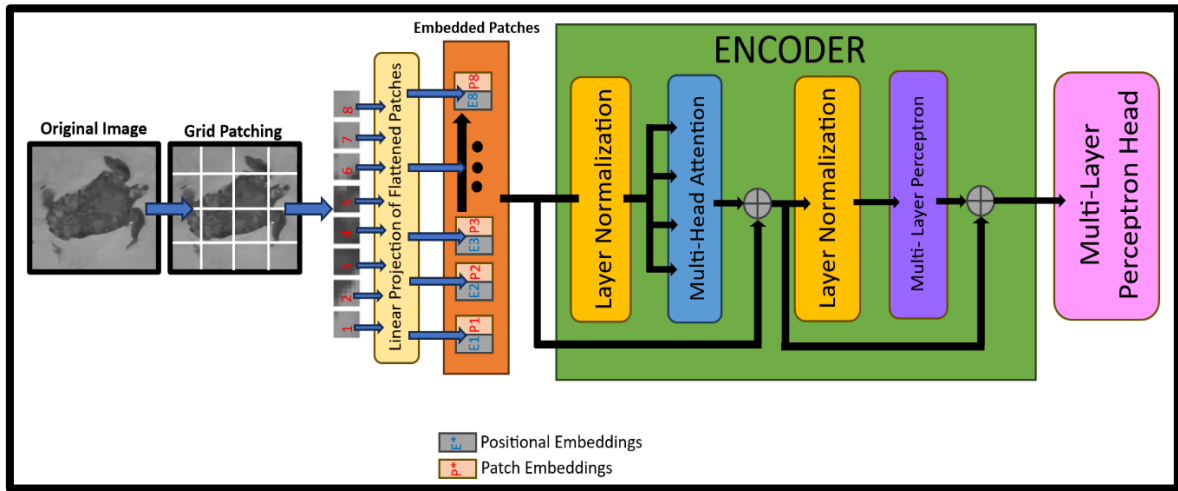


Figure 3.13: The implementation of the vision transformer architecture in this research

The figure above shows the general overview of how the vision transformer is designed in this research. The original image is first loaded with a fixed dimension like how it is in CNN architectures. For this research, the dimension values of **128x128** for the image was used. The difference is instead of loading the image as a 2D array and passing it to the encoders directly, the image is instead separated into a preset grid of dimensions. This preset grid of dimensions can be set into several different grid size values such as 16, 32, or 64. For this research, a grid size of **16** patches per image is used. A single grid square of the image is now known as a patch which is then projected as a linear projection of patches along with its position in the global context of the image embedded with it before being passed to the encoder. This ensures the positional information of each patch is kept and allows the model to identify the position of the respective patches. This positional embedding (PE) allows the

self-attention mechanisms of vision transformers to perform permutations on the PE to identify the tokens (which are the embedded patches) from their different positions in the global context of the full image (Yu et al., 2022). The attention mechanism, specifically self-attention, is achieved at the encoder level where the multi-head attention layer calculates the pair-wise attention between the tokens and the spatial information of those tokens to assign a value of how important it is for that token to pay attention to itself and the other tokens (Y. Li et al., 2023). Self-attention can be achieved with just a single attention head in the encoder, but adding more attention heads can help provide better training stability since single-head attention would require more layers to attend to the positions (Liu et al., 2021).

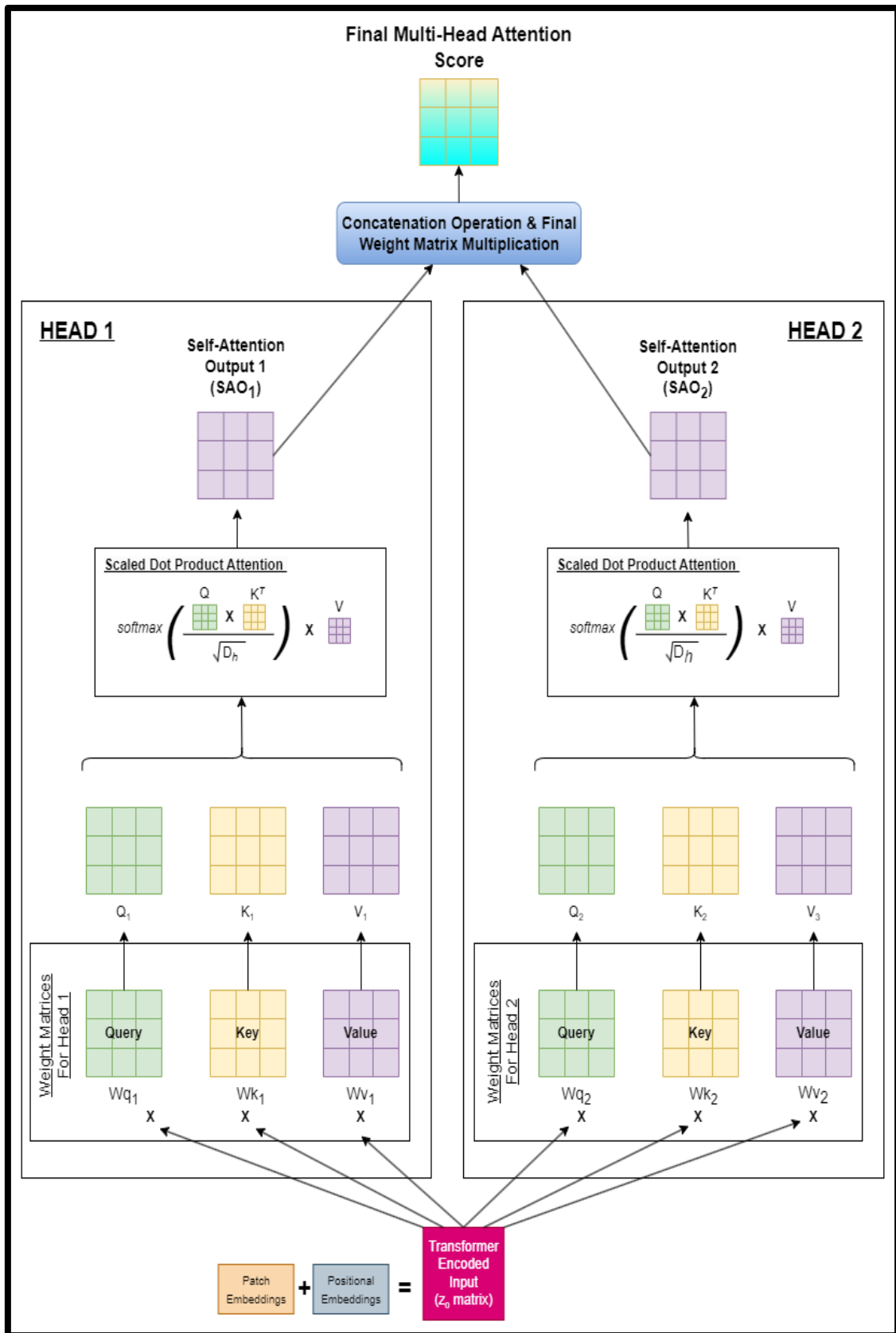


Figure 3.14: Example on the Design of the Multi Head Attention

For this research, the amount of multi-head attention is set to only be **8** heads. Figure 3.13 shows an example of how the heads in the vision transformer model create an attention mechanism in the model architecture. The output of each patch that undergoes the embedding process creates a transformer-encoded input of each patch to be fed to the vision transformer encoder. The encoder then creates many duplicates of this transformer-encoded input and passes 3 to every attention head in the multi-head attention layer. Each of the transformer encoder input matrix, once they are in an attention head, is multiplied with the 3 weight matrices W_q , W_k , and W_v . The 3 represents the query(q), key(k), and value (v). These 3 matrixes are randomly initialized and are thus different for every attention head. This is one of the key parts of what makes multi-head attention establish attention. Due to the random initialized nature of the weights in each attention head, it allows the model to treat and put attention values to each patch differently than one another. Multi-head attention enables the model to be able to simultaneously attend to information from the different representation subspaces at different positions (Vaswani et al., 2023). For an analogy, if attention head 1 treats transformer encoded input 1 as important because it is in the middle of every different image, attention head 2 might instead treat transformer encoded input 1 as important because it is always between transformer encoded input 2 and 3 in different-different images. The weights are only randomly set during initialization, after that they are regularly updated into a learned matrix via backpropagation during training. At the end of each attention head is the self-attention output matrix. This matrix is concatenated with all the other self-attention output matrix of all the other remaining attention heads and multiplied on another set of learned weight matrix to generate the final multi-head attention score of that patch. The parameter of the vision transformer model design is as in the following attached figure.

```

# Training device:
device = torch.device("cuda")

# Hyperparameters:
batch_size = 16
epochs = 25 #25 for training session 1 and later another 25 for training session 2
lr = 3e-5
gamma = 0.7
seed = 69
IMG_SIZE = 128
patch_size = 16
num_classes = 12

# Linear Transformer:
efficient_transformer = Linformer(dim=128, seq_len=64+1, depth=12, heads=8, k=64)

# Vision Transformer Model:
model = ViT(dim=128, image_size=128, patch_size=patch_size, num_classes=num_classes, transformer=efficient_transformer, channels=3).to(device)

# loss function
criterion = nn.CrossEntropyLoss()

# Optimizer
optimizer = optim.Adam(model.parameters(), lr=lr)

# Learning Rate Scheduler for Optimizer:
scheduler = StepLR(optimizer, step_size=1, gamma=gamma)

```

Figure 3.15: The hyperparameters used in the vision transformer model

Key things to re-iterate here as was previously mentioned are the image size dimensions of **128** which are fixed for every image, the patch size of **16** meaning each image is patched into 16 pieces to be passed to the encoder, the number of attention heads is set to **8**, and lastly two of the more important things that was not mentioned in the previous section is the fixed epoch of **50** and batch size of **16** for training the vision transformer model. In the code in Figure 3.14, it is 25 but that was because it was split into two separate sessions of training as training 50 epochs in one go was not feasible due to the amount of strain and training time it taxed on the training hardware. The batch size is set to 16 as more batch sizes for training the vision transformer was not viable for the training hardware as the next considerable jump for batch sizes such as 32 and 64 was untrainable due to insufficient GPU memory available on the hardware. A batch size of 8 was a possibility to be considered but

it was decided that a batch size of 16 can be used as a form of controlled parameter that can be compared with the CNN models that are to be developed later. It was known that in CNN models especially too small batch sizes can cause the network to bounce forward and backward without attaining any good performance (Kandel & Castelli, 2020). As such since it was planned to not use a batch size that is too low such as 8 for the CNN model training parameters and added with the fact that the vision transformer can only be capped at 16 batch size; it was decided that the batch size for all the models is to be set at 16. The training epoch of 50 is also a controlled parameter for the method and as such all the CNN models are also capped at 50 epochs for each model. The architecture for the CNN models (the AlexNet, SqueezeNet, and ResNet50 models) are standard and the parameters for each CNN deep learning model used are as in the figures below.

```
# Hyperparameters:
batch_size = 16
epochs = 50
lr = 3e-5
gamma = 0.7
seed = 69
IMG_SIZE = 128
num_classes = 12

# loss function
criterion = nn.CrossEntropyLoss()

# Optimizer
optimizer = optim.Adam(model.parameters(), lr=lr)

# Learning Rate Scheduler for Optimizer:
scheduler = StepLR(optimizer, step_size=1, gamma=gamma)
```

Figure 3.16: The model parameters that are shared for all 3 CNN models (AlexNet, SqueezeNet, and ResNet50)

At the end of this section, 4 different deep-learning models were trained and developed.

- i. The vision transformer model, which is an attention-based deep learning model.
- ii. An AlexNet model, which is based on the architecture of a deep convolutional neural network was first published in the paper “ImageNet Classification with Deep Convolutional Neural Networks (Krizhevsky et al., 2017).
- iii. A SqueezeNet model, based on a deep convolutional neural network architecture that has far fewer architecture parameters than AlexNet but can maintain an accuracy level that is comparable to AlexNet (Iandola et al., 2016).
- iv. A ResNet50 model, equipped with a deep residual learning framework that can perform residual mapping by adding identity mapping layers that help reduce degradation especially when the networks start to get deeper (He et al., 2016).

The re-identification capabilities of all the models when applied to the unaltered test set to simulate direct re-identification are tested as well as with the long-term scenario test set that is to be described in the next section. The training and testing results are presented and discussed in the next chapter of this research thesis.

3.7 Testing Developed CNN Models and Attention-Based Model with Artificially Aged *Chelonia Mydas* Test Images for Long-Term Scenario Simulation

Now that all the models are ready, a new set of test data can be created that simulates a long-term scenario whereby the sea turtle carapace exhibits some details of natural change that can be considered normal for a *Chelonia mydas* sea turtle to experience over time. It is understood during the literature review that the carapace pattern feature of a *Chelonia mydas*

individual is itself stable even after more than 10 years as has been stated previously by Tabuki et al. But what is also known based on the experience of working and interviews with the park rangers and wardens, is that the barnacle growth that can be found dotted around the *Chelonia mydas* individual's carapace is a form of random change that can occur over time to the carapace. The park wardens and rangers have exclaimed that some barnacles once removed from an individual can regrow during the next nesting season in different quantities and sizes. This presents a good opportunity to utilise barnacle growth to simulate a long-term scenario. Different sizes and numbers of barnacles can be applied on the carapace of the test data as a simulation for long-term scenario re-identification tests.

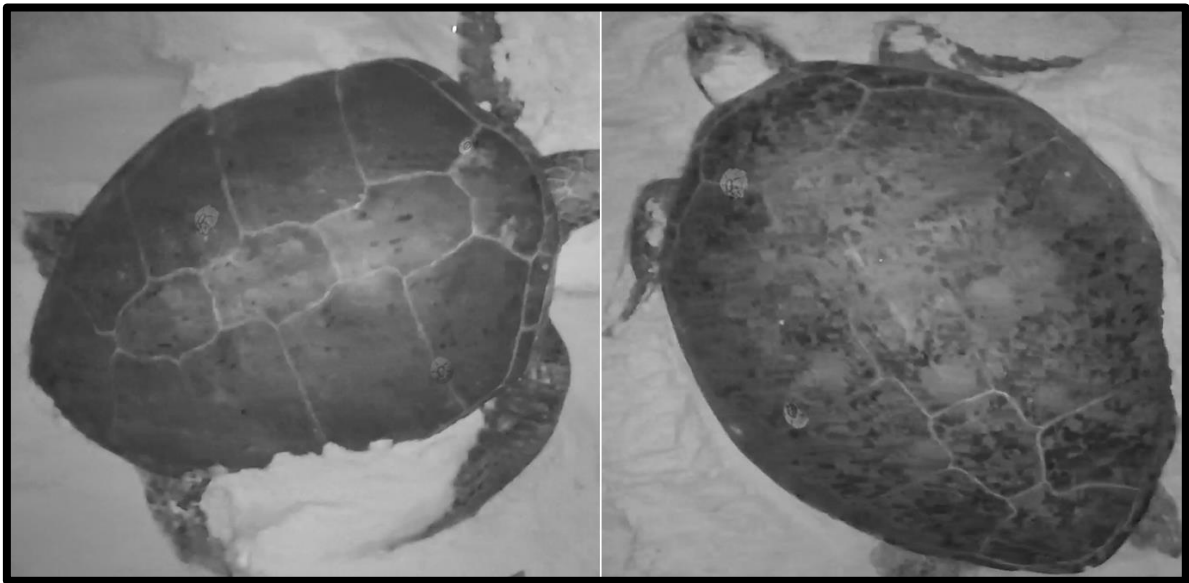


Figure 3.17: Shows two images of simulated barnacle growth on 2 individual carapaces (September 2022).

As such a test set is prepared that attempts to simulate a long-term scenario for re-identification of individual *Chelonia mydas* females that has barnacle growth simulated on its carapace. The dataset consists of 6 “barnacled carapace” images per turtle for each of the 12 *Chelonia mydas* individuals making a total of 72 images of “barnacled carapaces” to be

re-identified belonging to which turtle. The image manipulation software GIMP (GNU Image Manipulation Software) is used to attach the barnacles. When attaching the barnacles to the carapaces of each turtle, it was ensured that each turtle has 2 simulated carapaces with 1 barnacle, 2 simulated carapaces with 2 barnacles, and 2 simulated carapaces with 3 barnacles; to be fair that none of the carapaces would have significantly more simulated carapace change than the other that can impact re-identification results. This creates a form of severity level based on how much barnacles are on the carapace of the turtle individual with 1 barnacle indicating light changes, 2 barnacles indicating moderate changes, and 3 barnacles indicating large changes to the carapace. Also, one thing to note is that the barnacle size must be realistic. The mean average size for barnacles on *Chelonia mydas* green sea turtle individuals can be around 0.5cm to 4cm at the largest extent, with the majority being around 0.8cm to 1cm for adult females (Lim et al., 2020). Larger sizes than this are not simulated. The results of the experiments for the long-term barnacle growth scenario are presented and discussed in the following chapter along with the results of the other methods from the previous section of this chapter.

3.8 Summary

In summary, the research methodology has been defined and several research works that has been discussed in the research flow has been done in this Chapter. The deep learning models have been designed and their parameters have been defined. The deep learning models consists of 3 CNN models without attention mechanisms which are the AlexNet, ResNet and SqueezeNet model. The deep learning model with attention mechanisms is designed based on the vision transformer architecture. The planning and data collection required for training the models has been detailed in this chapter which involves sampling

activities on Talang Besar Island of Sarawak, Malaysia. The data gathered is then analysed and extracted using an automated cropping method using a YOLOv7 model that was trained with a selected number of images from our data collected. The data collected is used to prepare a dataset used to train the models while another set will be used to test the re-identification accuracy of the models in short- and long-term scenarios. The long-term test will utilize a set of images from the dataset that has been altered by attaching barnacles on the images with several levels of severity. This attempts to simulate a long-term scenario where the carapace is susceptible to barnacle growth over time. The next chapter will discuss on the results of the experiments and the comparison of the several metrics between the 4 deep learning models.

CHAPTER 4

FINDINGS AND DISCUSSIONS

4.1 Setup of Experiments

For the experiments in this chapter, training and re-identification testing was performed on the models as described in Chapter 3: Methodology by using the data that was also discussed in detail in that chapter. The graphs of the training results for the models and the classification results for direct re-identification and re-identification in long-term barnacle growth scenario were included in the form of the confusion matrices. All the models were trained based on the same dataset and the test dataset used for re-identification testing was consistent across all tests. The parameters for the models were kept as similar as possible between the models as described in the methods which are re-iterated here:

- 50 Epochs for each model; 16 Batch Size; 128 Dimensions for Images

All the models were trained on the same machine with the same training hardware as discussed in the methods which was an AMD 5600X CPU and an Nvidia RTX3060TI GPU. The training machine operates using the Windows 10 operating system from Microsoft and the programming and training for the models was done with the PyTorch machine learning framework running on the Jupyter Notebook development platform. The model is trained in sequence one after the other on the training machine starting with the attention-based model which was the vision transformer model first; then followed by the CNN models starting with the AlexNet model, the SqueezeNet model and the ResNet50 model being the last. After each model is trained and tested, the training machine is turned off to allow it to rest for up to 30 minutes before restarting and resuming training and testing on the next model.

4.2 Comparison of Training Loss Metric Among the Models

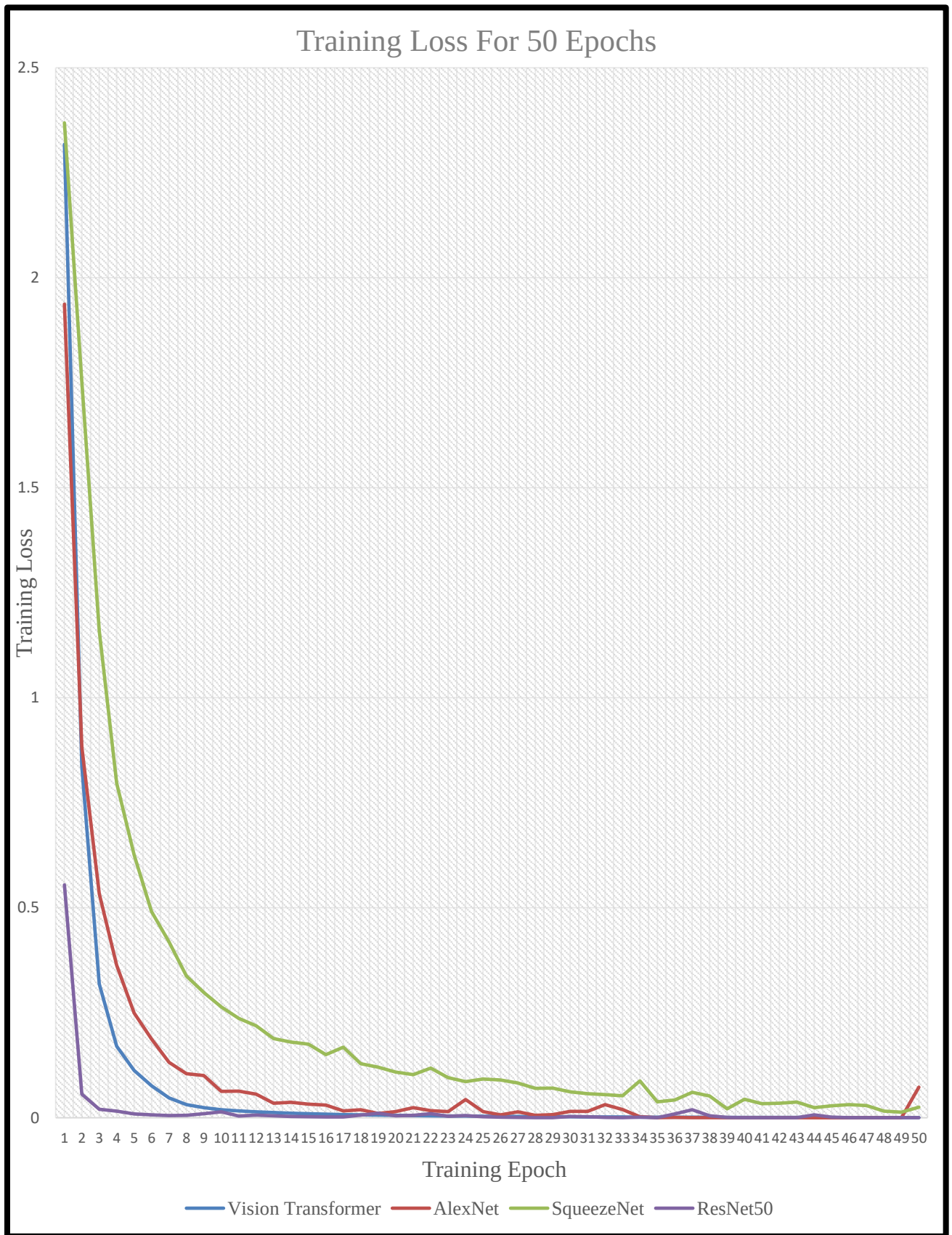


Figure 4.1: Graph of Training Loss for All Models At All Epochs During Training

One of the first important metrics to be investigated and discussed is the training loss of the deep learning models. The training loss for every single epoch from 1 to 50 was recorded and a line graph on the loss value of each epoch during training was made. The graph can be seen in Figure 4.1 while the exact data can be found in a tabulated form attached in the appendix section of this thesis.

The training loss is a loss function used in machine learning during training as a method to gauge the measure of error for the model with regards to the training data. The ideal scenario for this value is that it gets minimized further and further nearing zero as the training epoch gets longer. This scenario indicates that the model is improving its capability to make better predictions on the training set while making fewer erroneous predictions. There are various types of loss functions used in machine learning but since the method in this thesis is performing a form of image classification task, the loss function used should be the Cross Entropy Loss function.

Based on the results obtained for this metric, it was observed and accurately determined that the vision transformer model with its attention mechanisms can perform better than the AlexNet and SqueezeNet CNN models for this metric but below the performance of the ResNet 50 CNN model for the first 15 epochs. The ResNet50 model was able to minimize its training loss much quicker than the vision transformer attention-based model. The AlexNet model comes third while the SqueezeNet model performed the worst in terms of training loss as it seems that the SqueezeNet model has difficulties in making accurate predictions even way after many epochs has passed. It is also evident that not one of the models seems to be underfitting as none of the models seem to maintain a fairly high training loss throughout the 50 epochs with minimal random training loss spikes visible,

especially for the vision transformer model as it appears to have a quite stable gradual descent for its training loss indicating that it is gradually learning quite well.

4.3 Comparison of Validation Loss Metric Among the Models

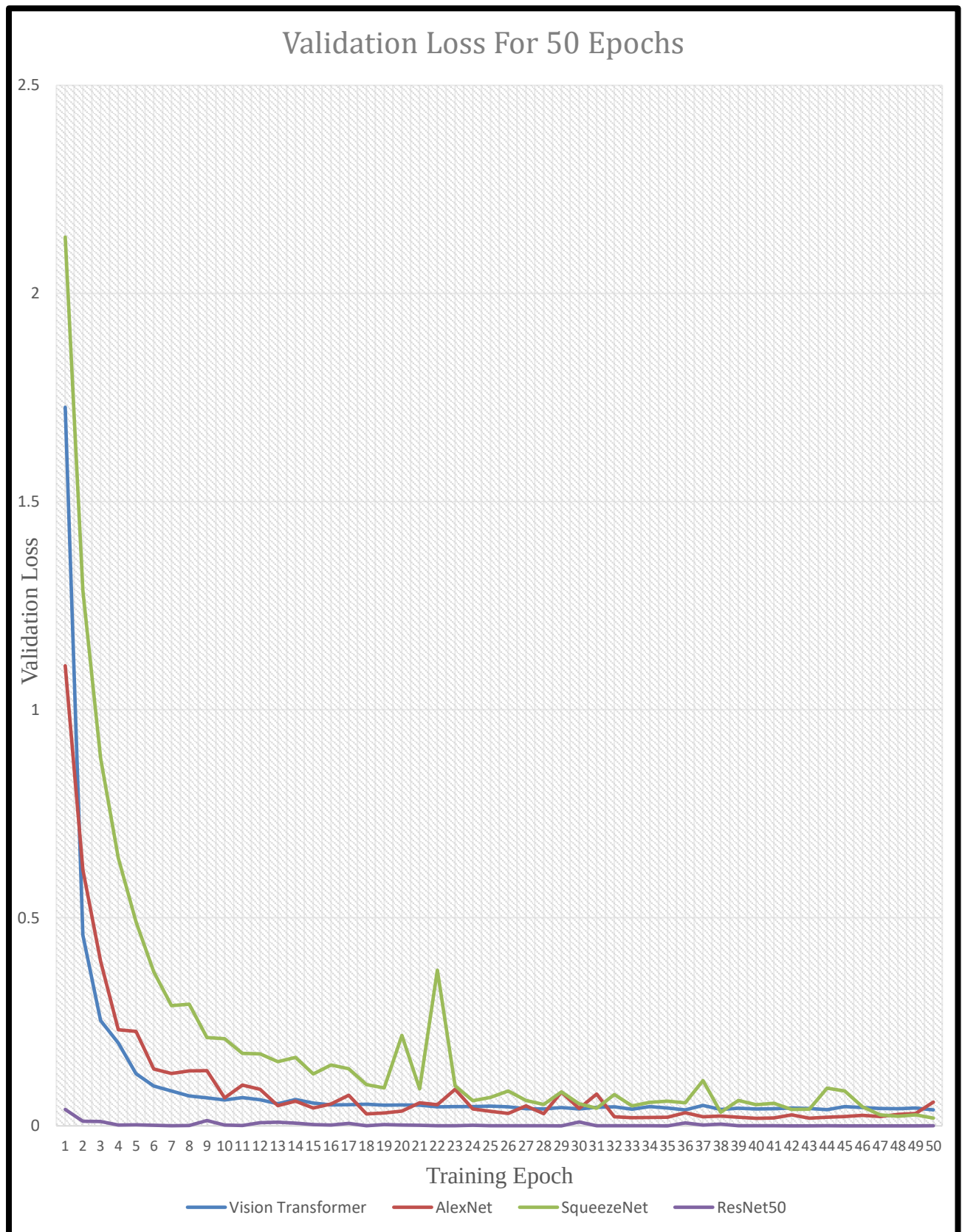


Figure 4.2: Graph of Validation Loss for All Models At All Epochs During Training

The next metric that should be discussed is the validation loss metric. Like training loss, validation loss is also an attempt to measure the degree of error for the model but for this instance, the measurement is made with regards to the validation data instead of training data. For this metric, the ResNet50 model outperforms every other model in this research by a large margin as it managed to minimize its validation loss quickly within the first 10 epochs to be nearing zero while the other models appear to struggle. What can be observed here is that the ResNet50 model is optimally fitted to be able to correctly make its prediction with the validation set with only a few epochs of training needed. The other models on the other hand are struggling a bit. The SqueezeNet model is the worst performer of the 4 models as it struggles to stabilize its validation loss even after 20 epochs of training. The random spikes can be attributed to having an unlucky batch of images being loaded for validation during those epochs from 20 until 23 but the other 3 models seemed to just be performing much better throughout the 50 epochs for this metric than the SqueezeNet model.

One impressive thing to note here is how stable the gradual descent of the vision transformer model's validation loss is compared to the others. Although it seems to reach its optimal stable point around epoch 15 and refuses to decrease much further after epoch 36, this stability tells us that the vision transformer can achieve a consistent performance of classification on the validation set for every epoch at and after its stable point. That is great as it means that the vision transformer model is well fit. It is not too overfitted to cause too many errors for the validation data; whilst being reliable enough to be used for classification at lower epochs compared to the SqueezeNet or AlexNet models that only seems to be reaching stability after epoch 38 and epoch 32 respectively.

4.4 Comparison of Training Accuracy Metric Among the Models

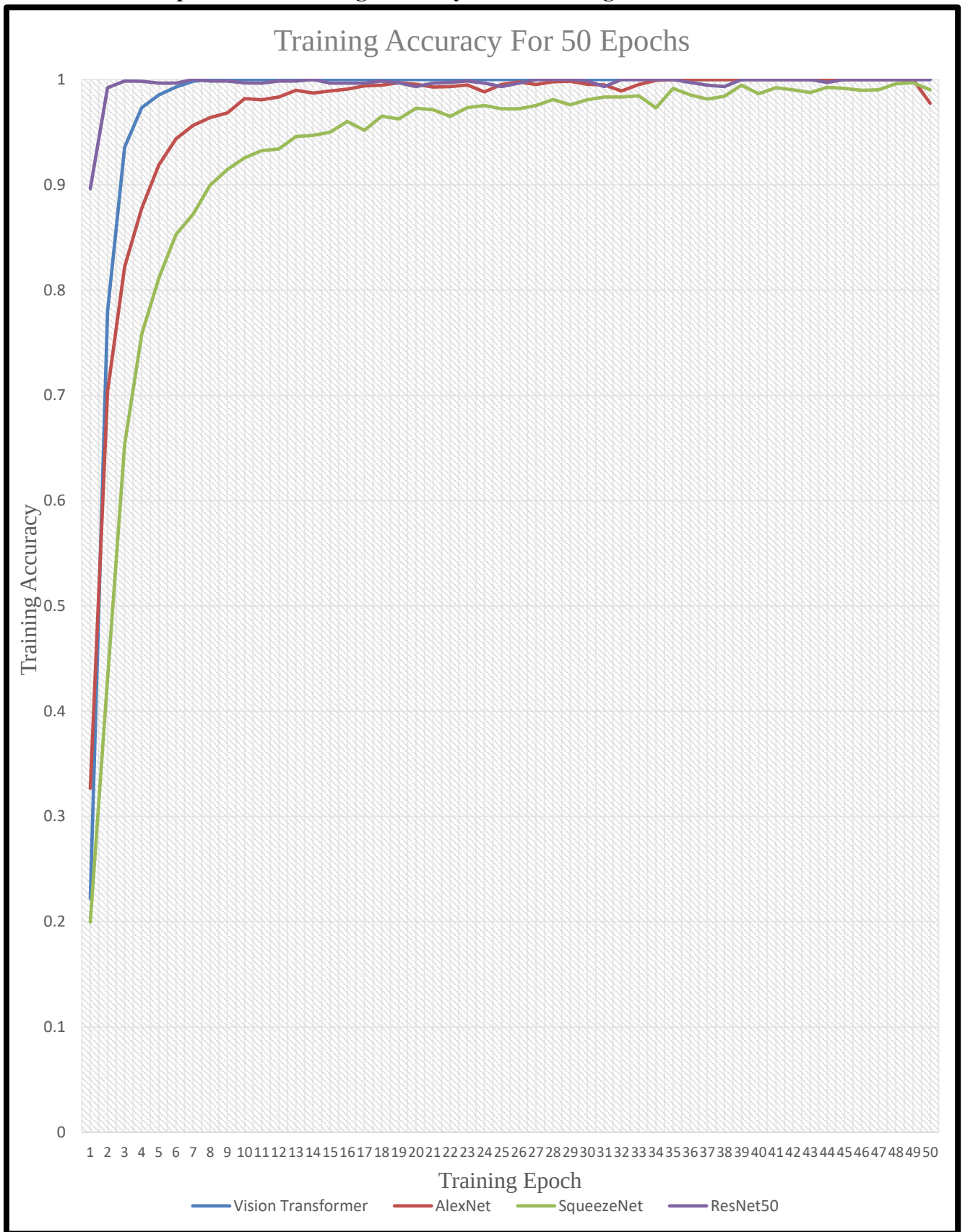


Figure 4.3: Graph of Training Accuracy for All Models At All Epochs During Training

The next metric to discuss is the training accuracy of the 4 models. Training accuracy is one of the metrics that enable us to understand how well the model can make correct predictions on the labelled data when presented with the data from the training set. It refers to how well the model can accurately predict the correct label for the training set data that it is shown. Lower values indicate that the model is unable to perform well, while higher values indicate a fitted model. Exceptionally high values however may indicate overfitting of a model. But for the case of this research, the high training accuracy is better as it conforms to the notion that the model can be very specific in assigning the correct labels to what it believes in when making a prediction. Specificity is more important than generalisation when it comes to re-identification as it is preferable to avoid an underfitted model that would just assume a *Chelonia mydas* is which individual based on general characteristics of the turtle.

For this metric, again it is observed that the ResNet50 CNN model and the vision transformer model with attention mechanisms outperform the AlexNet and SqueezeNet CNN models. The training accuracy metric for ResNet50 is the best but the vision transformer model is very effective performance-wise by being not too far behind to reach higher values of training accuracy in the training cycle. By epoch 10, both have very good training accuracy values, unlike the AlexNet and SqueezeNet CNN models which seem to be underperforming even after 20 epochs with the SqueezeNet CNN model being the worst offender of the 4 models. It appears that the vision transformer is consistently performing better than some CNN architectures so far and by doing so at lower training epochs.

4.5 Comparison of Validation Accuracy Metric Among the Models

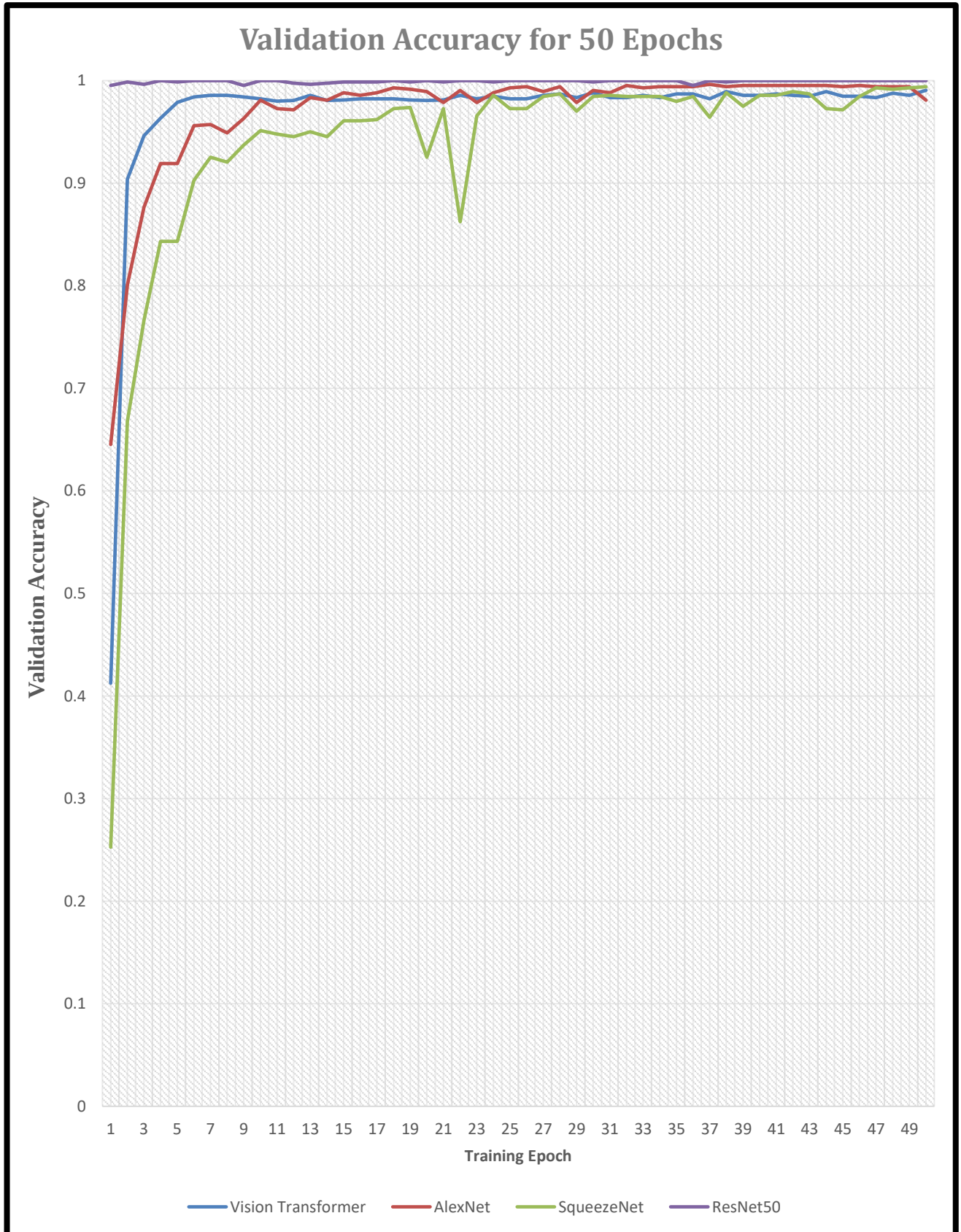


Figure 4.4: Graph of Training Accuracy for All Models At All Epochs During Training

Besides that, the validation accuracy metric of the models can also be discussed. Validation accuracy is like training accuracy but with regards to the model's predictive accuracy on the validation set instead of the training set. Again, like what has been seen earlier, the ResNet50 CNN model and the vision transformer model with attention mechanisms seem to be the leading two for this metric by reaching quite a reasonably high level of accuracy as soon as 10 epochs. The SqueezeNet model is again consistently being the worst performer of the 4 models. For this metric, high validation accuracy is ideal as it provides us with an indication that the model would be able to make accurate predictions on new unseen data (validation set). So far, the validation accuracy for all models seems to be quite well as a high validation accuracy value indicates that overfitting is not a concern for any of the models. The ResNet 50 model is outstanding for this metric as it doesn't seem to need many training epochs to be able to perform correct predictions on the validation set. The other 2 two CNN models seem to require much more training epochs before they can achieve a good validation accuracy and stability making early stopping an unideal scenario should it be attempted.

The vision transformer model seems to take lesser training epochs than the two to reach a stable point before being consistently able to achieve a constant validation accuracy value. However, this may cause concern as it means that after a certain point, in this case, it is epoch 10, the vision transformer doesn't seem to be able to improve its validation accuracy after any more successive training and is being outpaced by the AlexNet model after epoch 14 and then by SqueezeNet model after epoch 45. This tells us that after a certain point, there isn't much more reason to keep training the vision transformer model. Combined with the results of the vision transformer model's training accuracy that also achieved stabilization at earlier training epochs, it can be said for certain that training the vision transformer after

epoch 10-15 does not really improve the model much in any way or capacity. Early stopping would have saved us precious training time as optimal performance has been achieved earlier anyways.

4.6 Comparison of Training Time After Every Epoch Metric Among the Models

Table 4.1: Table of Total Training Time Taken to Reach Every Epoch

Epoch	Total Training Time Taken to Reach Every Epoch (Hours: Minutes: Seconds)			
	Vision Transformer	AlexNet	SqueezeNet	ResNet50
1	0:22:50	0:00:36	0:00:34	0:00:37
5	1:51:19	0:02:37	0:02:40	0:03:00
10	3:41:56	0:05:09	0:05:20	0:05:54
15	5:32:37	0:07:42	0:07:59	0:08:54
20	7:23:46	0:10:16	0:10:40	0:11:51
25	9:15:17	0:12:49	0:13:25	0:14:54
30	11:04:04	0:15:20	0:16:04	0:17:52
35	12:52:10	0:17:54	0:18:40	0:20:59
40	14:39:26	0:20:28	0:21:14	0:24:00
45	16:26:08	0:23:00	0:23:50	0:26:59
50	18:14:19	0:25:30	0:26:29	0:30:02

One of the most overlooked metrics when it comes to training deep learning models is the time taken for the models to train. Loss and accuracy values are important metrics to gauge a model's performance but knowing whether the effort taken to achieve the set performance is reasonable or not is as valuable as those other metrics as well. If efficient models take too much time to train, and the option of using a slightly less performant but faster-to-deploy model is available; it would be a justifiable reason to select one model over the other. Based on the results of the running total duration of the 4 models after training each epoch, there is a dramatic difference that can be seen for this metric between CNN models and the vision transformer model with attention mechanisms. In the case of this research, the vision transformer model with its attention mechanism takes longer durations to train than any of the CNN models. To reach 50 epochs, the 3 CNN models would need in

average only around 27 minutes to finish training. However, for the vision transformer model, it would need a staggering 18 hours 14 minutes and 19 seconds to reach 50 training epochs. But one thing that was discovered in the results based on comparing the other metrics such as the training loss, validation loss, training accuracy, and validation accuracy is that technically there wasn't much reason to train the vision transformer model after around epoch 10 to 15 since the metrics did not justify that more training is necessary. Still, there is a caveat since even if training was stopped at epoch 10, it would still take 3 hours to train. This is still considerably longer than what the other models took to reach the same epoch or even 50 epochs on that matter.

4.7 Re-Identification Capabilities of All Models with Direct Re-Identification on Test Set

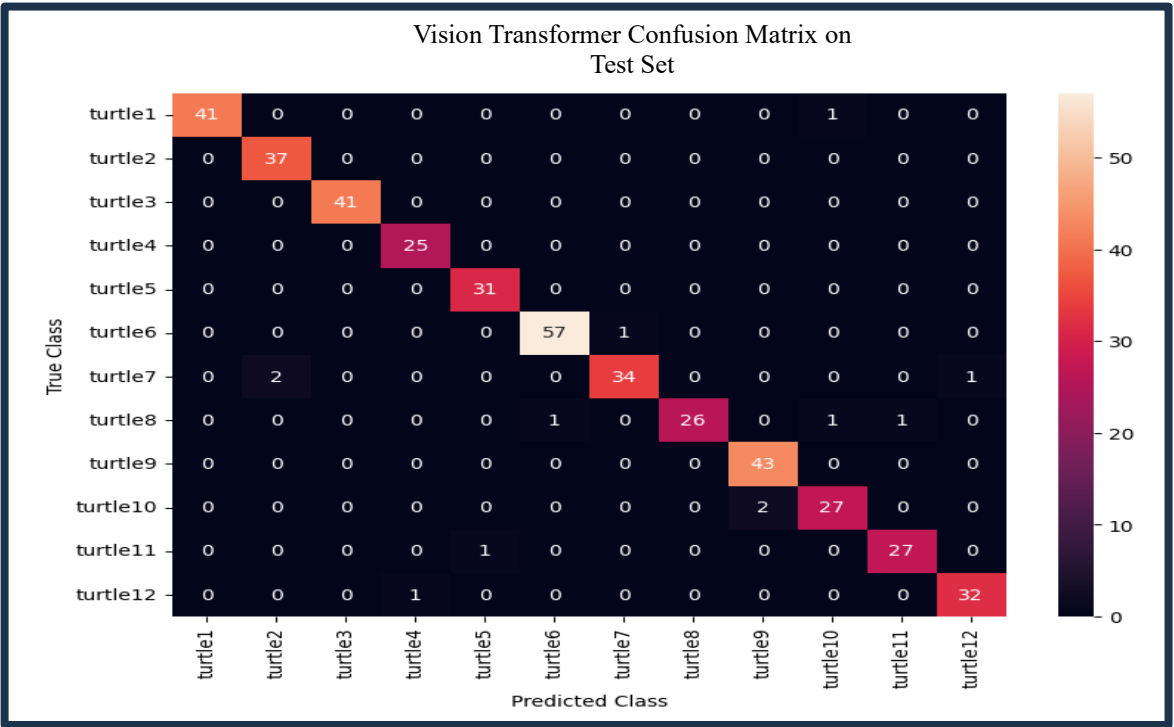


Figure 4.5: Confusion Matrix for Vision Transformer Model on Direct Re-Identification

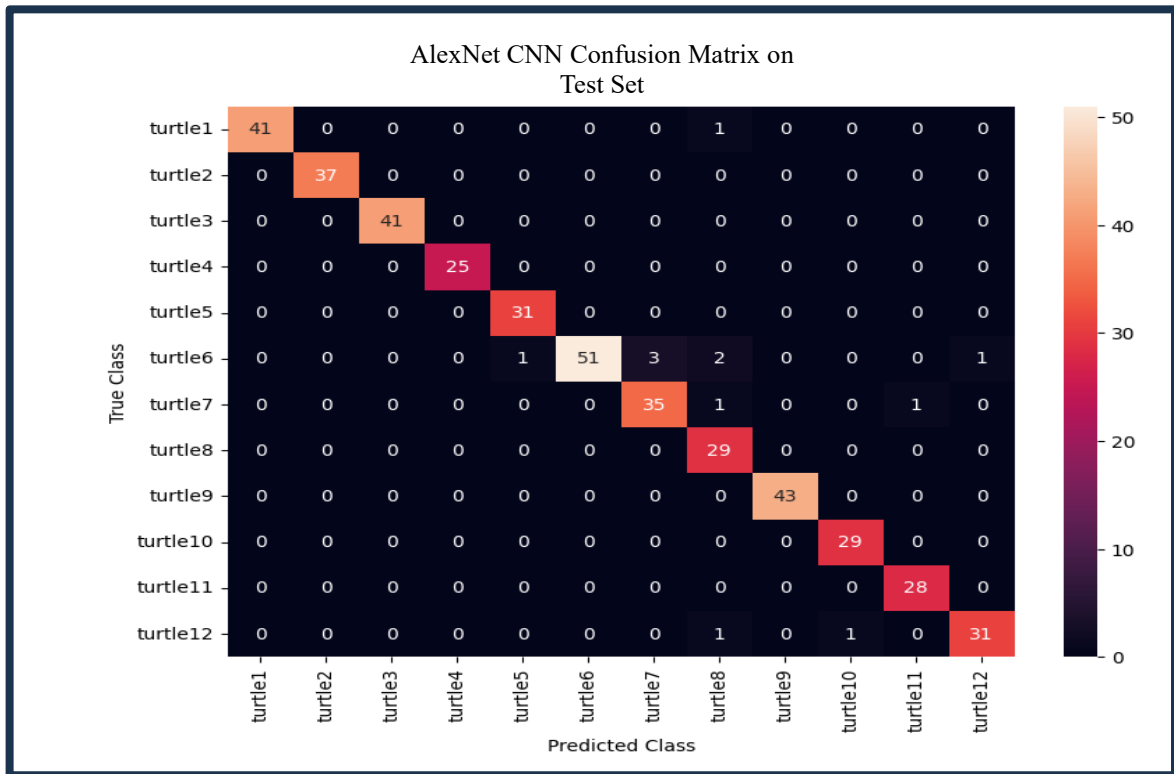


Figure 4.6: Confusion Matrix for AlexNet CNN Model on Direct Re-Identification

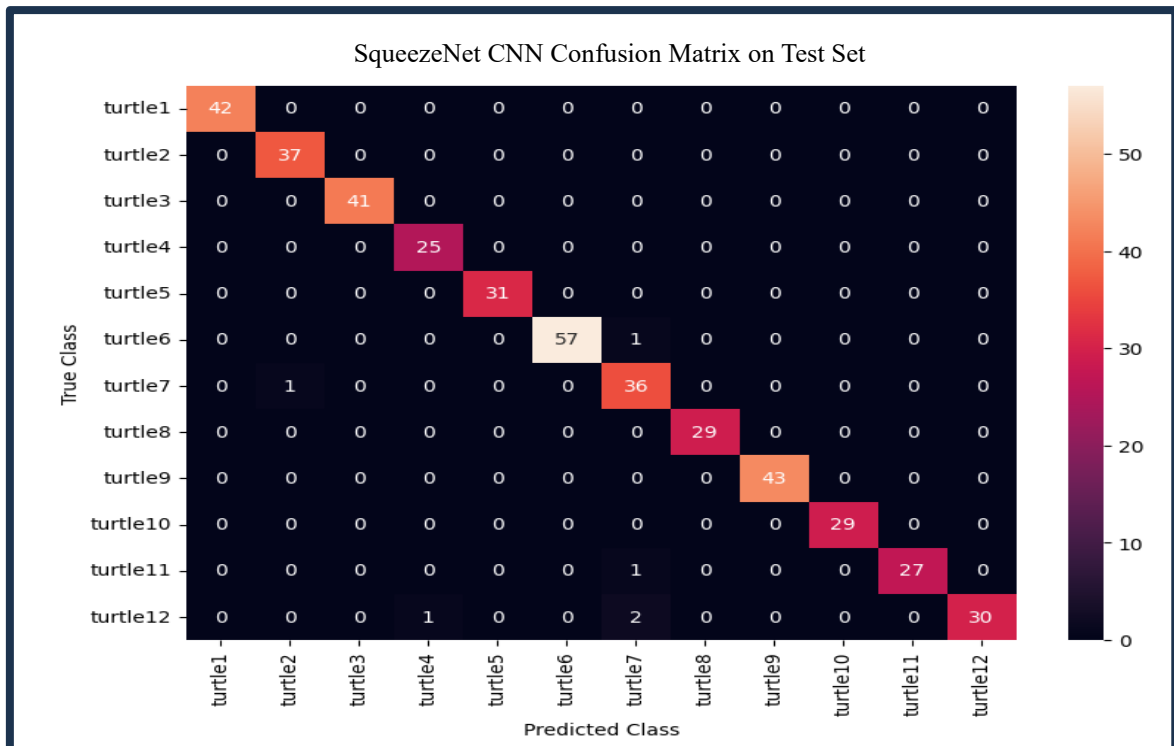


Figure 4.7: Confusion Matrix for SqueezeNet CNN Model on Direct Re-Identification

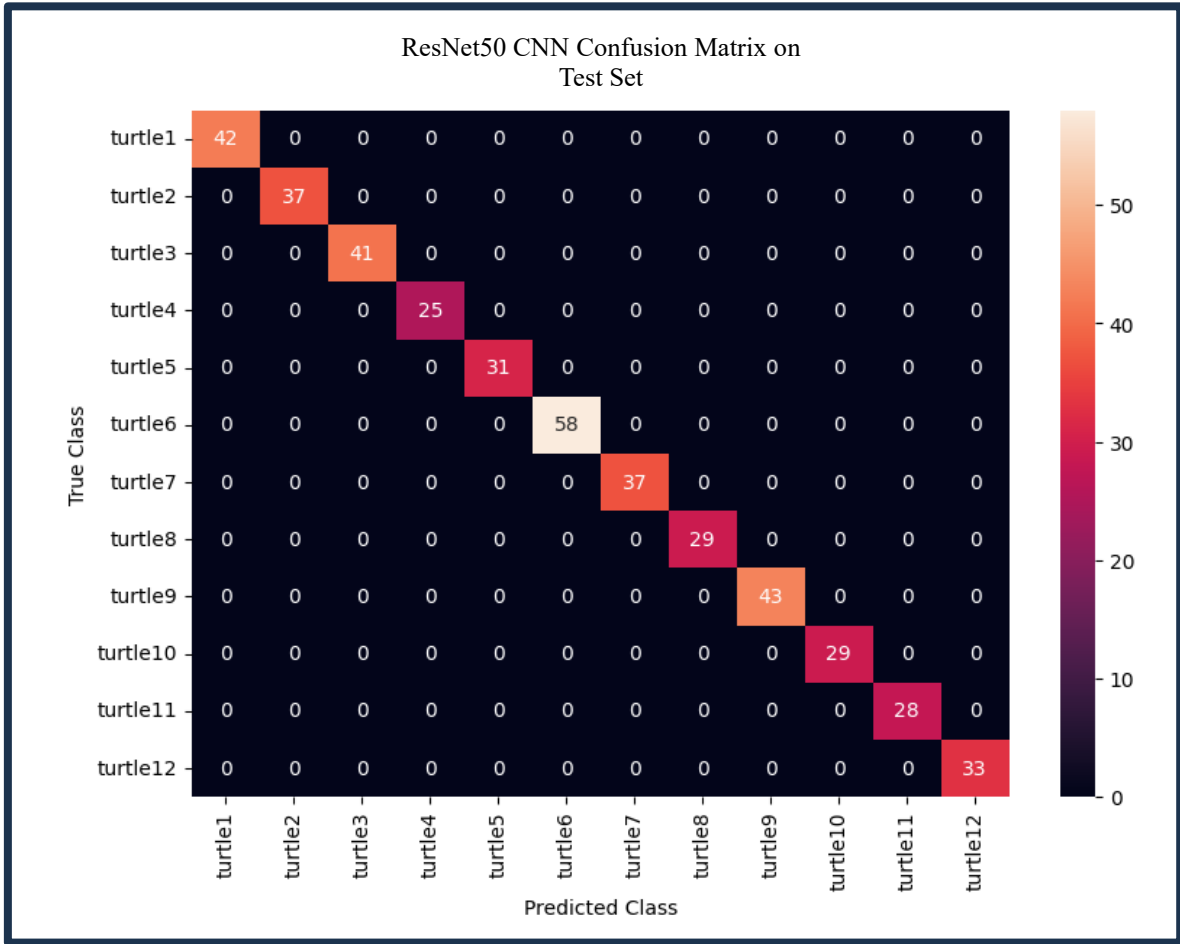


Figure 4.8: Confusion Matrix for ResNet50 CNN Model on Direct Re-Identification

Based on the results of the confusion matrix on classification testing with the test set data, it is observed that all the 4 models are able to perform a high degree of correct re-identification on each of the carapace images in the test set. In total, there are 433 carapace images of *Chelonia mydas* from the 12 individuals that is not included in the training set and validation set. The classification accuracy of each model based on the confusion matrix results is calculated using the following formula and the result is included in the table below:

$$Accuracy = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

Table 4.2: Accuracy results of confusion matrixes on the test set

	Vision Transformer	AlexNet	SqueezeNet	ResNet50
Accuracy (Decimal)	0.9723	0.9723	0.9861	1.0
Accuracy (%)	97.23%	97.23%	98.61%	100%

It appears that when performing re-identification on unseen images of carapace data in the test set, the SqueezeNet CNN model appears to be able to re-identify which carapace in the test belongs to which individual turtle better than the AlexNet and the vision transformer models. ResNet50 performs spectacularly with correct re-identification of whom the carapaces in the test data belong to while the vision transformer and the AlexNet model are not as perfect as that, but they still provide a decent result that is certainly applicable to be used for direct re-identification scenario.

One thing to note and understand is that the confusion matrix also tells us that for the vision transformer, AlexNet and SqueezeNet models; each of them are unique in that they have some re-identification errors on different individuals instead of the same one. The vision transformer has mispredictions on turtles 2 and 9, AlexNet has some issues with turtles 7 and 8 while SqueezeNet has issues with turtle 7 but no issues with turtle 8.

4.8 Re-Identification Capabilities of All Models with Simulated Long-Term Scenario of Barnacle Growth

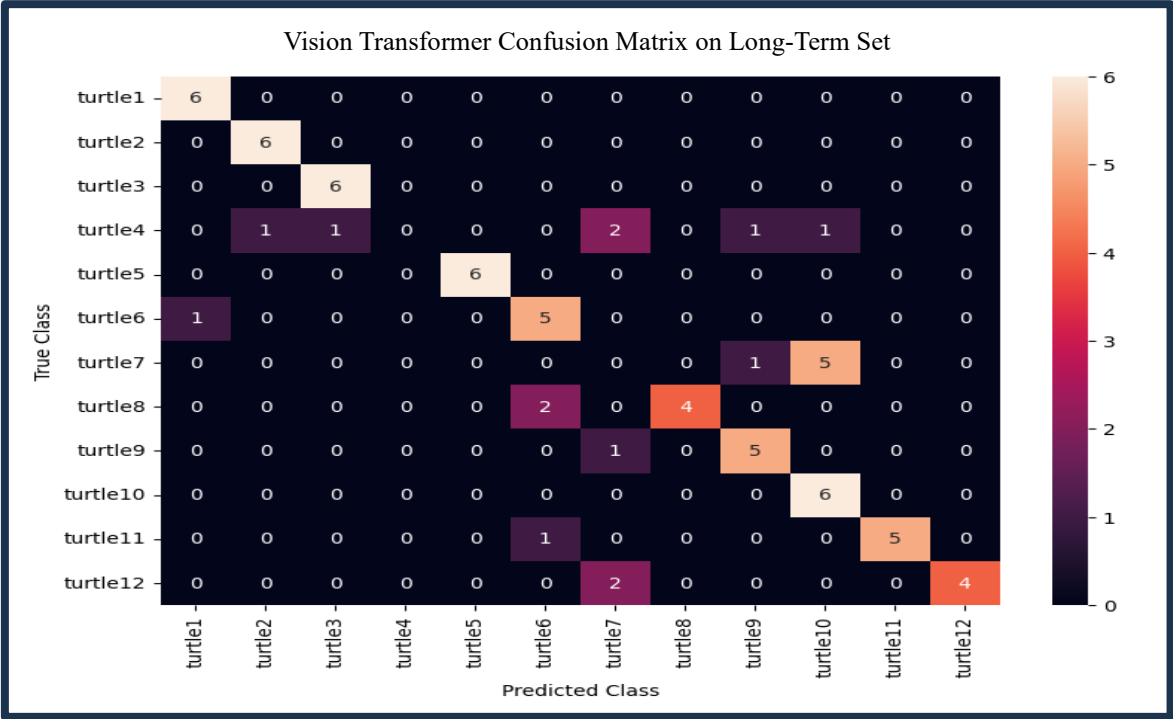


Figure 4.9: Confusion Matrix for Vision Transformer Model on Long-Term Set

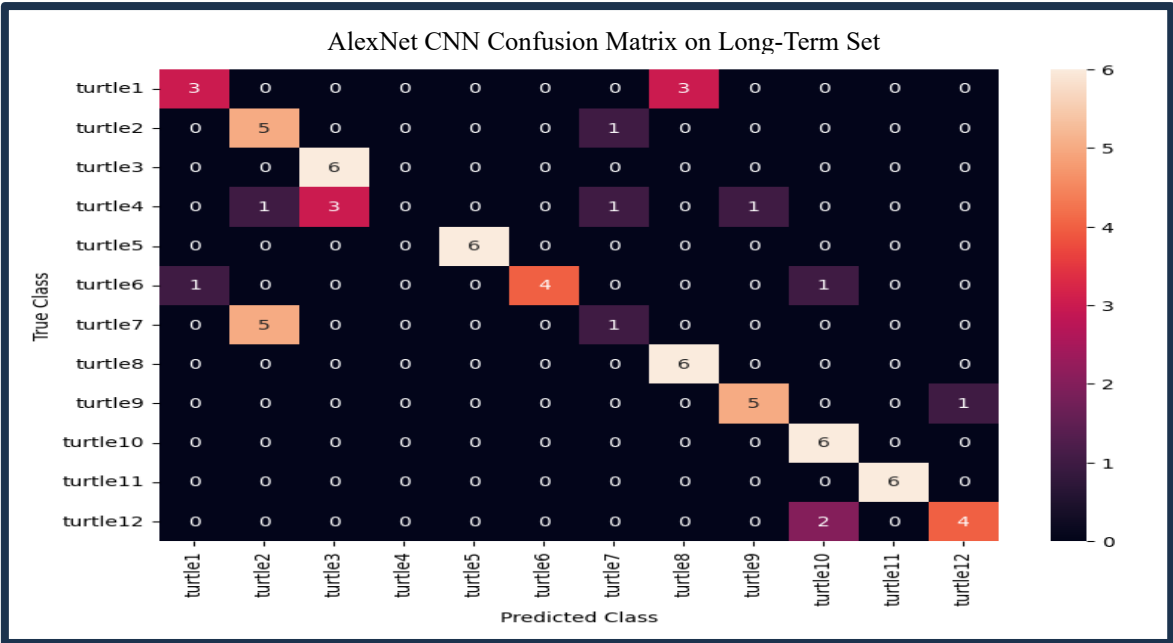


Figure 4.10: Confusion Matrix for AlexNet CNN Model on Long-Term Set

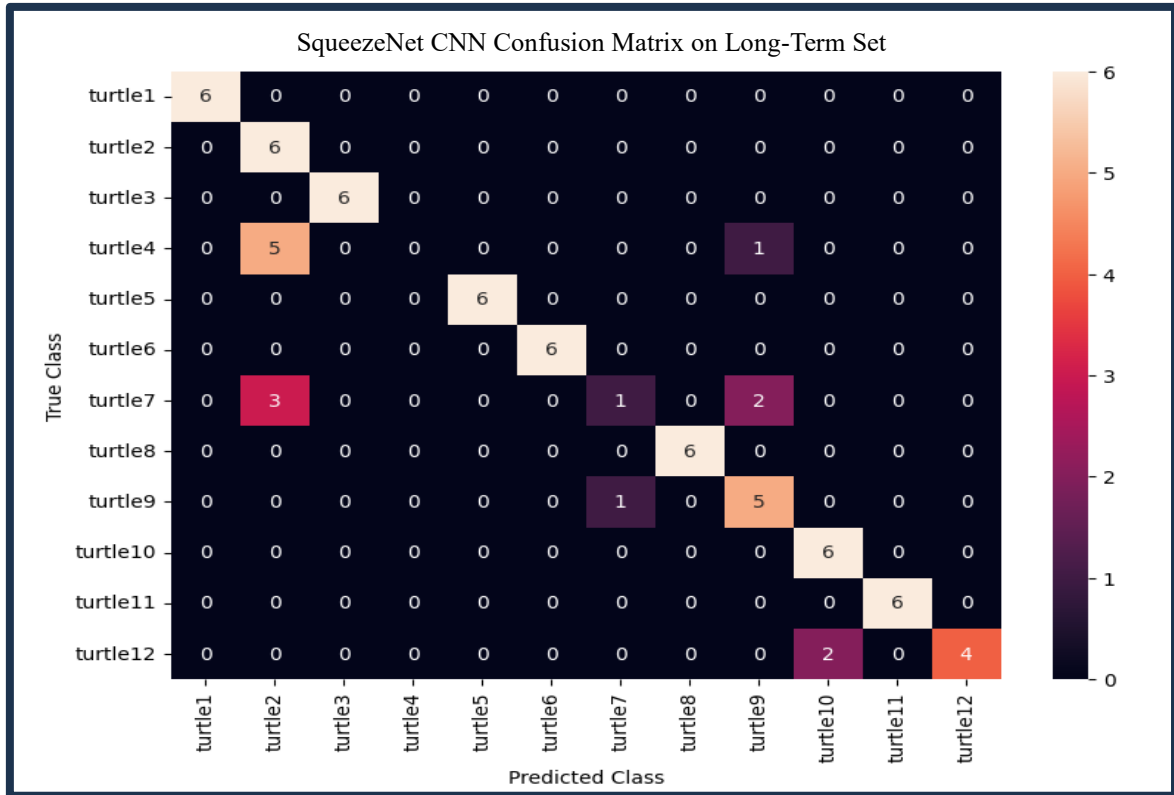


Figure 4.11: Confusion Matrix for SqueezeNet CNN Model on Long-Term Set

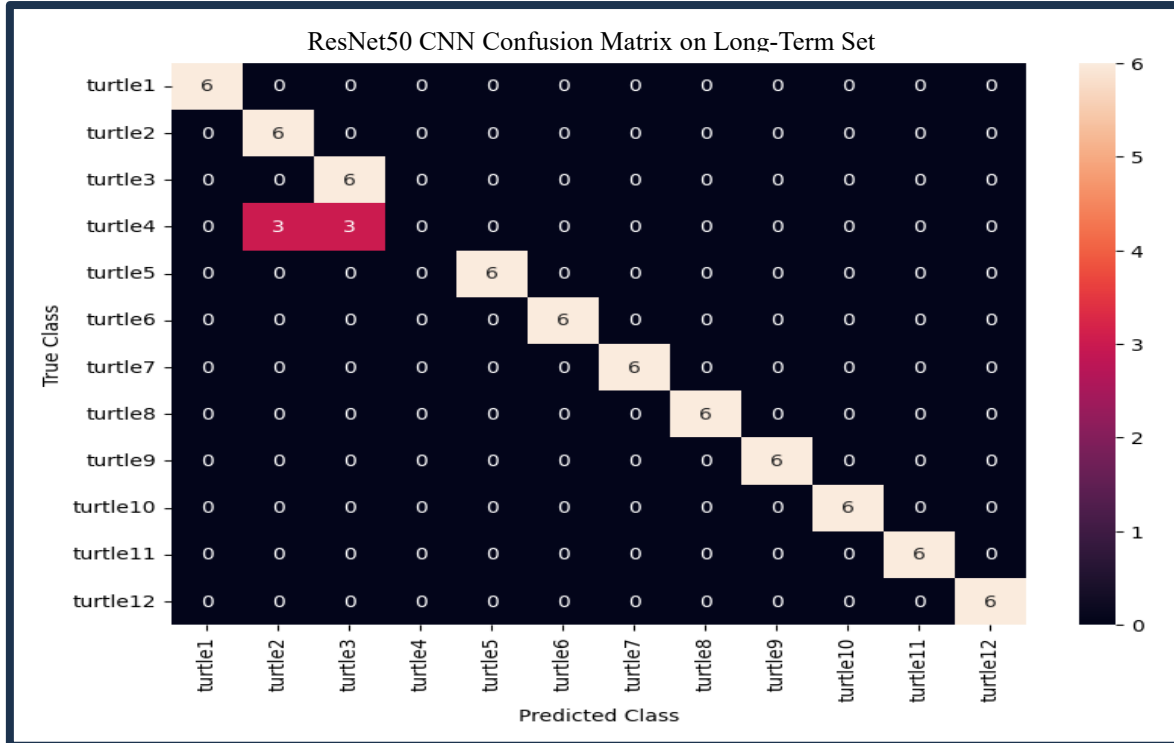


Figure 4.12: Confusion Matrix for ResNet50 CNN Model on Long-Term Set

Table 4.3: Accuracy results of confusion matrixes on the test set

	Vision Transformer	AlexNet	SqueezeNet	ResNet50
Accuracy (Decimal)	0.7361	0.7222	0.8056	0.9167
Accuracy (%)	73.61%	72.22%	80.56%	91.67%

Based on the accuracy calculations from the confusion matrixes in this long-term scenario test where barnacle is simulated on the carapace of the turtles; it was concluded that the actual best model for re-identification was the ResNet 50 CNN model. The SqueezeNet model was 2nd, while the vision transformer model with its attention mechanisms was 3rd followed by the AlexNet CNN model last.

It is possible to infer some useful information from this. Firstly, the SqueezeNet model is perhaps more underfitted than the vision transformer model thus it was able to handle more generalization of the features on what to expect on the *Chelonia mydas* carapaces in the long-term set. The vision transformer model and AlexNet model, on the other hand, performed well during training and validation but when tested on the long-term scenario of barnacle growth, suffered from significant difficulty in re-identifying which carapace belongs to whom when barnacles were added. Another thing to note is that it appears none of the models were able to predict accurately in the long-term scenario simulation for turtle 4 in the long-term set even though they managed to perform well on direct re-identification with un-simulated carapaces on turtle 4 test set.

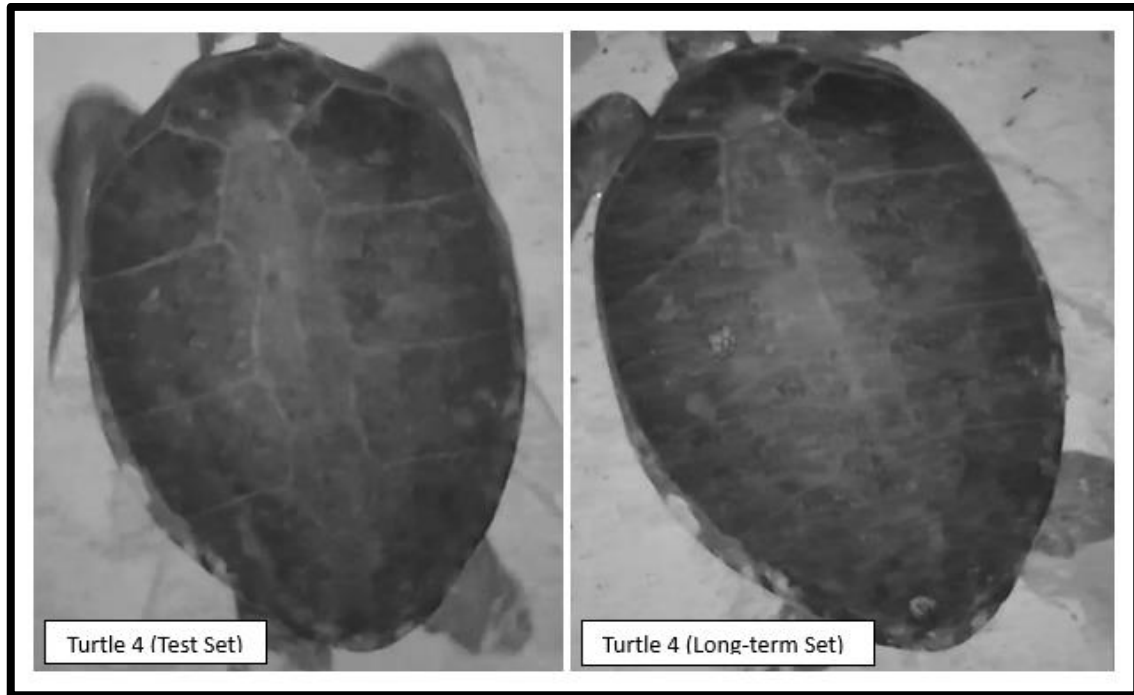


Figure 4.13: These are 2 images that were used to test the re-identification capabilities of the models for direct and long-term scenario (September 2022).

For a human, it is possible to identify that these two carapaces belong to the same turtle even though there are two barnacles present in the long-term set. One of the key indicators that a human reviewer can use to identify how these two carapaces belong to the same individual is by the distinct markings and patterns on the top section and middle left section of the carapace. There are clear colouration and pattern markings that makes this carapace identifiable by a human. However, the vision transformer model with its attention mechanisms and the CNN models were not able to. This can be attributed to all models being overly fitted on turtle 4 that it was unable to accept even the change of adding two barnacles as acceptable to be labelled as turtle 4. This does not however occur for all the turtles. This issue is specific to turtle 4 as none of the deep learning models were able to identify any of the simulated long-term test carapaces to turtle 4.

Another key takeaway that can be inferred from the confusion matrix but not from the accuracy data that was calculated is that the vision transformer model, when it can re-identify a turtle that has been simulated to have barnacle growth, can do so consistently. Notice how for every turtle that the vision transformer model was able to re-identify in the long-term set, the model can identify each turtle in 4 or more simulated images of barnacle growth. Unlike the AlexNet or SqueezeNet models, the vision transformer model confidently re-identifies the same individual, regardless of the severity of the barnacle changes on the carapace. The vision transformer model does show promise in this regard when compared to the AlexNet and SqueezeNet models.

However, in the end, ResNet50 is the best performer of all the models in re-identification capability on the long-term set. It was only not able to re-identify turtle 4 while performing well on everything else.

4.9 Summary

In summary, the vision transformer model performs well generally with respectable accuracy and is comparable with some CNN models on several metrics as was discussed in this chapter. But based on the results, it was determined that some of the CNN models can indeed outperform a vision transformer model that has attention mechanisms for re-identification of *Chelonia mydas* individuals in direct re-identification or in a long-term simulated scenario such as barnacle growth on the carapace of the turtle. Before ending this chapter, it must be mentioned that to compare the vision transformer model to the CNN-based models, 3 different CNN models (AlexNet, SqueezeNet, ResNet50) were used which are unique to one another out of a myriad of other CNN architecture choices such as DenseNet, EfficientNet, ShuffleNet and many others. The vision transformer model might

perform better (or worse) in comparison should these other CNN architectures were chosen instead of the 3 that was used in this research. Due to difficulties in attempting training and testing on many more CNN-based models, we can say that should there be a limitation where training and testing a bunch of CNN models to find the best one be a troublesome or time-consuming procedure, then vision transformer models can be a safe alternative to use for re-identification purposes, whether it be on *Chelonia mydas* sea turtles or for other use cases.

CHAPTER 5

CONCLUSION

5.1 Summary

5.1.1 Data collection of *Chelonia Mydas* sea turtles and Automated Pipeline for Cropping ROIs from Field Video Recordings

This work utilized a machine learning model to extract data from field video recordings that was captured during sampling work that was done at Talang Besar Island. The model has helped to automate the process of creating the dataset that was used for this research. The YOLO v7 model trained using just 1,436 manually annotated carapace images was able to perform automated cropping of the carapace regions of interests in our field video recordings. It was able to prune through 49 different videos and crop out carapace images of 35 different *Chelonia mydas* individuals. As such the research has demonstrated that it is possible to utilise deep learning models such as the YOLOv7 model to help in filtering and extracting cropped images from animal field recordings. The automated cropping process that was done by the YOLO v7 model has made our work in creating the dataset for designing and making our deep learning models much easier and shows potential for such methods to be used for similar future works. There are also 2 papers published based on the work that was done from this part of the research in scientific journals focused on Artificial Intelligence. The list of publications is as below:

- The paper titled “*Chelonia mydas* detection and image extraction from field recordings” published in Volume 13, No 2 of the IAES International Journal of

Artificial Intelligence (IJ-AI) that is published by the author of this thesis as a first author.

- The paper titled “An Instance Segmentation Method for Nesting Green Sea Turtle’s Carapace using Mask R-CNN” published in Volume 16, Issue 1 of the International Journal of Computing and Digital Systems that is published by the author of this thesis as a second author.

5.1.2 Measuring the Efficiency and Effectiveness of Standard CNN-based Models Versus Our Proposed Method

In this thesis, several deep learning models were designed such as the vision transformer model with attention mechanisms, the AlexNet, SqueezeNet, and the ResNet 50 model. The efficiency and effectiveness of these models are then measured and compared based on several metrics such as training and validation loss, training and validation accuracy, and training time. Re-identification tests on a test set for direct re-identification scenario and for long-term scenario of simulated barnacle growth was also performed. The results of this thesis of which have been compiled and discussed has shown that the vision transformer model can indeed perform as well or better than some standard CNN models such as the AlexNet, and SqueezeNet model but not as well as some models like the ResNet 50 model. Lastly, it can be mentioned that the advantage of the deep learning vision transformer model with attention mechanisms compared to the CNN models is that it doesn’t require as many training epochs as some of the CNN models to perform better on some training and validation metrics. Besides, it can be implied that even though the vision transformer model isn’t the best; it is still a decent performer in the re-identification tests. Compared to some CNN models that did not perform as well as the vision transformer model

in the training and validation metrics as well as in the classification tests, we can proudly say that if we were to choose only one model to develop, it would be the vision transformer model. We perhaps got lucky that one of the models we selected was ResNet50 and it turned out to be a great CNN architecture for this research task, but it could have been a different CNN architecture selected such as an underperforming DenseNet and perhaps our vision transformer model would have been the overall best beating all the CNN based models. However, it was not so. Vision transformer models with their attention mechanisms have shown promise here, but perhaps more future work needs to be done to make it the best deep learning architecture to use for re-identification purposes.

5.2 Conclusion

Animal re-identification for ecological purposes is certainly a field where new technologies such as deep learning models can be incorporated to improve the field. The literature review done in this research has shown that there is indeed a practical beneficial advantage to incorporate such new technologies to replace traditional methods such as physical tags and tracking collars. In the case of *Chelonia mydas* sea turtles, deep learning models can indeed be explored and researched for potential use in individual re-identification to replace the use of GPS tags or Inconel flipper tags that possess certain disadvantageous characteristics.

5.3 Future Work

There is still more to be done with vision transformers models and CNN models for re-identification of *Chelonia mydas* individuals. In the future, perhaps we can propose for a more robust set of hardware equipment to create a better training machine to train our

models. We do understand that AI hardware is undoubtedly very expensive at current times with AI oriented GPUs easily costing tens and thousands of ringgits. But should the opportunity arise in the future, it will be interesting to see what we can do with more wiggle room provided by the hardware on the number of parameters we can change in our design of the vision transformer model. Perhaps increasing the number of patches to larger numbers such as 32 or 64 patches can provide more interesting results. Maybe perhaps increasing the image dimensions to a larger quality in the future can help in providing more context for the vision transformer to learn. The possibilities are numerous and future work can indeed provide us with better answers to these questions.

Lastly, it would have been great to compare the vision transformer model with attention mechanisms to even more CNN model architectures, not just AlexNet, SqueezeNet, and ResNet50. There are a wide variety of CNN architectures that we can test out and compare, and it would be very interesting to try and examine their capabilities to re-identify *Chelonia mydas* sea turtles as well. As time passes, new model architectures are made, and more innovations appear in the field of AI; it certainly will be interesting to see how well vision transformer models fare against the future AI models to come. For now, we are proud of our work and believe that the future is indeed bright for AI and for the *Chelonia mydas* sea turtles at Talang-Satang as well.

REFERENCES

- AlZoman, R. M., & Alenazi, M. J. F. (2021). A Comparative Study of Traffic Classification Techniques for Smart City Networks. *Sensors*, 21(14), 4677. <https://doi.org/10.3390/s21144677>
- Amjoud, A. B., & Amrouch, M. (2023). Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review. *IEEE Access*, 11, 35479–35516. <https://doi.org/10.1109/ACCESS.2023.3266093>
- Araujo, G., Montgomery, J., Pahang, K., Labaja, J., Murray, R., & Ponzo, A. (2016). Using minimally invasive techniques to determine green sea turtle *Chelonia mydas* life-history parameters. *Journal of Experimental Marine Biology and Ecology*, 483, 25–30. <https://doi.org/10.1016/j.jembe.2016.06.004>
- Awwalu, J. (2019). On Holdout and Cross Validation. *International Journal of Trend in Research and Development*, 6(2), 235–239. https://www.researchgate.net/publication/332350661_On_Holdout_and_Cross_Validation_A_Comparison_between_Neural_Network_and_Support_Vector_Machine
- Azmi, N., Kamarudin, L. M., Zakaria, A., Al Kanagaraj, E., Ahmad, R. B., Yusoff, M. H., Abdullah, E. F. H. S., & Jamal, A. A. (2019, July). *Radio Frequency Identification (RFID) Range Test for Animal Activity Monitoring*. <https://ieeexplore.ieee.org/document/8940097/>
- Balazs, G. H. (1999). Factors to Consider in the Tagging of Sea Turtles. *Research and Management Techniques for the Conservation of Sea Turtles*, 4, 1–10. http://www.seaturtle.org/library/BalazsGH_1999_InResearchandManagementTechniquesfort_p1-10.pdf

- Bello, R.-W., Zawawi, A., Sufril, A., A., D., & Noah, F. (2020). Image-based Individual Cow Recognition using Body Patterns. *International Journal of Advanced Computer Science and Applications*, 11(3). <https://doi.org/10.14569/IJACSA.2020.0110311>
- Bilodeau, S. M., Schwartz, A. W. H., Xu, B., Paúl Pauca, V., & Silman, M. R. (2022). A low-cost, long-term underwater camera trap network coupled with deep residual learning image analysis. *PLOS ONE*, 17(2), e0263377. <https://doi.org/10.1371/journal.pone.0263377>
- Brothers, J. R., & Lohmann, K. J. (2018). Evidence that Magnetic Navigation and Geomagnetic Imprinting Shape Spatial Genetic Variation in Sea Turtles. *Current Biology*, 28(8), 1325-1329.e2. <https://doi.org/10.1016/j.cub.2018.03.022>
- Caravaggi, A., Banks, P. B., Burton, A. C., Finlay, C. M. V., Haswell, P. M., Hayward, M. W., Rowcliffe, M. J., & Wood, M. D. (2017). A review of camera trapping for conservation behaviour research. *Remote Sensing in Ecology and Conservation*, 3(3), 109–122. <https://doi.org/10.1002/rse2.48>
- Carpentier, A. S., Jean, C., Barret, M., Chassagneux, A., & Ciccione, S. (2016). Stability of facial scale patterns on green sea turtles *Chelonia mydas* over time: A validation for the use of a photo-identification method. *Journal of Experimental Marine Biology and Ecology*, 476, 15–21. <https://doi.org/10.1016/j.jembe.2015.12.003>
- Carter, S. J. B., Bell, I. P., Miller, J. J., & Gash, P. P. (2014). Automated marine turtle photograph identification using artificial neural networks, with application to green turtles. *Journal of Experimental Marine Biology and Ecology*, 452, 105–110. <https://doi.org/10.1016/j.jembe.2013.12.010>

- Chahal, A., & Gulia, P. (2019). Machine Learning and Deep Learning. *International Journal of Innovative Technology and Exploring Engineering*, 8(12), 4910–4914. <https://doi.org/10.35940/ijitee.L3550.1081219>
- Chen, R., & Mao, H. (2022). The Impact of Autopilot on Tesla. *BCP Business & Management*, 31, 89–95. <https://doi.org/10.54691/bcpbm.v31i.2540>
- Chen, X., Yang, T., Mai, K., Liu, C., Xiong, J., Kuang, Y., & Gao, Y. (2022). Holstein Cattle Face Re-Identification Unifying Global and Part Feature Deep Network with Attention Mechanism. *Animals*, 12(8), 1047. <https://doi.org/10.3390/ani12081047>
- Chew, V. Y.-C., Liew, H.-C., & Joseph, J. (2015). Photographic Identification of Green Turtles (*Chelonia mydas*) at Redang Island, Malaysia. *Marine Turtle Newsletter*.
- Costa-Pereira, R., Moll, R. J., Jesmer, B. R., & Jetz, W. (2022). Animal tracking moves community ecology: Opportunities and challenges. *Journal of Animal Ecology*, 91(7), 1334–1344. <https://doi.org/10.1111/1365-2656.13698>
- Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1, 886–893. <https://doi.org/10.1109/CVPR.2005.177>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale* (arXiv:2010.11929). arXiv. <http://arxiv.org/abs/2010.11929>
- Dupont, B. (2010). *Green Turtle (Chelonia mydas)* [Photograph]. [https://commons.wikimedia.org/wiki/File:Green_Turtle_\(Chelonia_mydas\)_\(6083206500\).jpg](https://commons.wikimedia.org/wiki/File:Green_Turtle_(Chelonia_mydas)_(6083206500).jpg)

- Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for Multi-Class Classification: An Overview* (arXiv:2008.05756). arXiv. <http://arxiv.org/abs/2008.05756>
- Hassan, R., & Yahya, N. K. (2022). Green Sea Turtle (*Chelonia Mydas*): A Historical Review With Relevance to Population Size in Sarawak. *International Journal of Biology and Biomedical Engineering*, 16, 221–232. <https://doi.org/10.46300/91011.2022.16.28>
- Hays, G. C., & Hawkes, L. A. (2018). Satellite Tracking Sea Turtles: Opportunities and Challenges to Address Key Questions. *Frontiers in Marine Science*, 5, 432. <https://doi.org/10.3389/fmars.2018.00432>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Hindarto, D. (2023). Enhancing Road Safety with Convolutional Neural Network Traffic Sign Classification. *Sinkron*, 8(4), 2810–2818. <https://doi.org/10.33395/sinkron.v8i4.13124>
- Hipiny, I., Ujir, H., Mujahid, A., & Yahya, N. K. (2018). Towards Automated Biometric Identification of Sea Turtles (*Chelonia mydas*). *Journal of ICT Research and Applications*, 12(3), 256. <https://doi.org/10.5614/itbj.ict.res.appl.2018.12.3.4>
- Hj Wan Yussof, W. N. J., Shahrudin, N., Hitam, M. S., Awalludin, E. A., Rusli, M. U., & Hoh, D. Z. (2020). Photo Identification of Sea Turtles Using AlexNet and Multi-Class SVM. In H. Fujita, A. Selamat, & S. Omatu (Eds.), *Frontiers in Artificial Intelligence and Applications*. IOS Press. <https://doi.org/10.3233/FAIA200549>

- Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). *SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size* (arXiv:1602.07360). arXiv. <http://arxiv.org/abs/1602.07360>
- IUCN. (2004). *Chelonia mydas*: Seminoff, J.A. (Southwest Fisheries Science Center, U.S.): *The IUCN Red List of Threatened Species 2004: e.T4615A11037468* [Dataset]. <https://doi.org/10.2305/IUCN.UK.2004.RLTS.T4615A11037468.en>
- Kahlout, K. M., & Ekler, P. (2021). Algorithmic Splitting: A Method for Dataset Preparation. *IEEE Access*, 9, 125229–125237. <https://doi.org/10.1109/ACCESS.2021.3110745>
- Kandel, I., & Castelli, M. (2020). The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. *ICT Express*, 6(4), 312–315. <https://doi.org/10.1016/j.icte.2020.04.010>
- Kayid, A., Yasmeen Khaled, & Elmahdy, M. (2018). *Performance of CPUs/GPUs for Deep Learning workloads*. <https://doi.org/10.13140/RG.2.2.22603.54563>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>
- Kucera, T. E., & Barrett, R. H. (2011). A History of Camera Trapping. In A. F. O’Connell, J. D. Nichols, & K. U. Karanth (Eds.), *Camera Traps in Animal Ecology* (pp. 9–26). Springer Japan. https://doi.org/10.1007/978-4-431-99495-4_2
- Kuncheva, L. I., Garrido-Labrador, J. L., Ramos-Pérez, I., Hennessey, S. L., & Rodríguez, J. J. (2023). An experiment on animal re-identification from video. *Ecological Informatics*, 74, 101994. <https://doi.org/10.1016/j.ecoinf.2023.101994>

- Kuncheva, L. I., Williams, F., Hennessey, S. L., & Rodríguez, J. J. (2022). A Benchmark Database for Animal Re-Identification and Tracking. *2022 IEEE 5th International Conference on Image Processing Applications and Systems (IPAS)*, 1–6. <https://doi.org/10.1109/IPAS55744.2022.10052988>
- Lamont, M. M., Fujisaki, I., Stephens, B. S., & Hackett, C. (2015). Home range and habitat use of juvenile green turtles (*Chelonia mydas*) in the northern Gulf of Mexico. *Animal Biotelemetry*, 3(1), 53. <https://doi.org/10.1186/s40317-015-0089-9>
- Lee, D. E., Cavener, D. R., & Bond, M. L. (2018). Seeing spots: Quantifying mother-offspring similarity and assessing fitness consequences of coat pattern traits in a wild population of giraffes (*Giraffa camelopardalis*). *PeerJ*, 6, e5690. <https://doi.org/10.7717/peerj.5690>
- Li, H., Rajbahadur, G. K., Lin, D., Bezemer, C.-P., Ming, Z., & Jiang. (2024). Keeping Deep Learning Models in Check: A History-Based Approach to Mitigate Overfitting. *IEEE Access*, 12, 70676–70689. <https://doi.org/10.1109/ACCESS.2024.3402543>
- Li, Y., Wang, J., Dai, X., Wang, L., Yeh, C.-C. M., Zheng, Y., Zhang, W., & Ma, K.-L. (2023). *How Does Attention Work in Vision Transformers? A Visual Analytics Attempt* (arXiv:2303.13731). arXiv. <http://arxiv.org/abs/2303.13731>
- Lim, K. K., Syed Hussein, M. A., & Palaniappan, P. (2020). Abundance, placement and sexual identity of the epizoic barnacle *Chelonibia testudinaria* relative to the size and species of host turtles in Mabul Island, Malaysia. *Journal of the Marine Biological Association of the United Kingdom*, 100(8), 1299–1309. <https://doi.org/10.1017/S0025315420001198>
- Liu, L., Liu, J., & Han, J. (2021). *Multi-head or Single-head? An Empirical Comparison for Transformer Training* (arXiv:2106.09650). arXiv. <http://arxiv.org/abs/2106.09650>

- Long, S. L., & Azmi, N. A. (2017, August). *Using photographic identification to monitor sea turtle populations at Perhentian Islands Marine Park in Malaysia*. *Herpetological Conservation and Biology*.
https://www.researchgate.net/publication/319956126_Using_photographic_identification_to_monitor_sea_turtle_populations_at_Perhentian_Islands_Marine_Park_in_Malaysia
- Loukili, S. E., Ezzati, A., Alla, S. B., & Zraibi, B. (2022). Reducing the training time of deep learning models using synchronous SGD and large batch size. *2022 International Conference on Intelligent Systems and Computer Vision (ISCV)*, 1–3.
<https://doi.org/10.1109/ISCV54655.2022.9806117>
- Matapurkar, P. (2023). *Animal Classification and Recognition Using Deep Learning*. 12(04).
- Mayne, B., Tucker, A. D., Berry, O., & Jarman, S. (2020). Lifespan estimation in marine turtles using genomic promoter CpG density. *PLOS ONE*, 15(7), e0236888.
<https://doi.org/10.1371/journal.pone.0236888>
- Mazlan, A. G., Zaidi, C. C., Wan-Lotfi, W. M., & Othman, B. H. R. (2005). On the current status of coastal marine biodiversity in Malaysia. *Indian Journal of Marine Sciences*, 34(1), 76–87.
- Miele, V., Dussert, G., Spataro, B., Chamaillé-Jammes, S., Allainé, D., & Bonenfant, C. (2021). Revisiting animal photo-identification using deep metric learning and network analysis. *Methods in Ecology and Evolution*, 12(5), 863–873.
<https://doi.org/10.1111/2041-210X.13577>

- Misischia, C. V., Poecze, F., & Strauss, C. (2022). Chatbots in customer service: Their relevance and impact on service quality. *Procedia Computer Science*, 201, 421–428. <https://doi.org/10.1016/j.procs.2022.03.055>
- Mohd-Azlan, J., & Sharma, D. S. K. (2003). Camera Trapping The Indochinese Tiger, *Panthera Tigris Corbetti*, In A Secondary Forest In Peninsular Malaysia. *The Raffles Bulletin of Zoology*, 51(2), 421–427. <https://lkcnhm.nus.edu.sg/wp-content/uploads/sites/10/app/uploads/2017/04/51rbz421-427.pdf>
- Moore, J. F., Soanes, K., Balbuena, D., Beirne, C., Bowler, M., Carrasco-Rueda, F., Cheyne, S. M., Coutant, O., Forget, P., Haysom, J. K., Houlihan, P. R., Olson, E. R., Lindshield, S., Martin, J., Tobler, M., Whitworth, A., & Gregory, T. (2021). The potential and practice of arboreal camera trapping. *Methods in Ecology and Evolution*, 12(10), 1768–1779. <https://doi.org/10.1111/2041-210X.13666>
- Munawar, R., 2022. *YOLO v7 Object Cropping*. GitHub. Available at: <https://github.com/RizwanMunawar/yolov7-object-cropping>
- Muraina, I. O. (2022). Ideal Dataset Splitting Ratios In Machine Learning Algorithms: General Concerns For Data Scientists And Data Analysts. *7th International Mardin Artuklu Scientific Researches Conference*, 496–504. https://www.researchgate.net/publication/358284895_IDEAL_DATASET_SPLITTING_RATIOS_IN_MACHINE_LEARNING_ALGORITHMS_GENERAL_CONCERNS_FOR_DATA_SCIENTISTS_AND_DATA_ANALYSTS
- National Band & Tag Company. (2018). *Sea Turtle Flipper Tags* [Photograph]. <https://www.nationalband.com/sea-turtle-flipper-tags/>
- National Oceanic and Atmospheric Administration. (2023, December 13). *Tagging Efforts for Sea Turtle Research | NOAA Fisheries* (West Coast, National, International).

- NOAA Fisheries. <https://www.fisheries.noaa.gov/west-coast/science-data/tagging-efforts-sea-turtle-research>
- National Wildlife Federation. (2024). *Green Sea Turtle*. National Wildlife Federation. <https://www.nwf.org/Home/Educational-Resources/Wildlife-Guide/Reptiles/Sea-Turtles/Green-Sea-Turtle>
- Neves-Ferreira, I., Mello-Fonseca, J., & Ferreira, C. E. L. (2023). Photo-identification shows the spatio-temporal distribution of two sea turtle species in a Brazilian developmental foraging ground. *Marine Biology*, 170(7), 83. <https://doi.org/10.1007/s00227-023-04226-z>
- Pant, Y., Shah, G., Ojha, R., Thapa, R., & Bhatta, B. (2023). Comparison of CNN Architecture of Image Classification Using CIFAR10 Datasets. *International Journal on Engineering Technology*, 1(1), 37–52. <https://doi.org/10.3126/injet.v1i1.60898>
- Patel, S., & Patel, A. (2021). Object Detection with Convolutional Neural Networks. In A. Joshi, M. Khosravy, & N. Gupta (Eds.), *Machine Learning for Predictive Analysis* (Vol. 141, pp. 529–539). Springer Singapore. https://doi.org/10.1007/978-981-15-7106-0_52
- Pfaller, J. B., Williams, K. L., Frick, M. G., Shamblin, B. M., Nairn, C. J., & Girondot, M. (2019). Genetic determination of tag loss dynamics in nesting loggerhead turtles: A new chapter in “the tag loss problem.” *Marine Biology*, 166(7), 97. <https://doi.org/10.1007/s00227-019-3545-x>
- Piacenza, J., Piacenza, S., Mayoral, S., Kenney, A., & Shields, N. (2018). Design Opportunities for Sea Turtle Satellite Tracking Devices. *Volume 4: 23rd Design for Manufacturing and the Life Cycle Conference; 12th International Conference on*

Micro- and Nanosystems, V004T05A024. <https://doi.org/10.1115/DETC2018-85583>

- Proctor, M. F., Garshelis, D. L., Thatte, P., Steinmetz, R., Crudge, B., McLellan, B. N., McShea, W. J., Ngoprasert, D., Nawaz, M. A., Te Wong, S., Sharma, S., Fuller, A. K., Dharaiya, N., Pigeon, K. E., Fredriksson, G., Wang, D., Li, S., & Hwang, M. (2022). Review of field methods for monitoring Asian bears. *Global Ecology and Conservation*, 35, e02080. <https://doi.org/10.1016/j.gecco.2022.e02080>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You Only Look Once: Unified, Real-Time Object Detection* (arXiv:1506.02640). arXiv. <http://arxiv.org/abs/1506.02640>
- Reisser, J., Proietti, M., Kinas, P., & Sazima, I. (2008). Photographic identification of sea turtles: Method description and validation, with an estimation of tag loss. *Endangered Species Research*, 5, 73–82. <https://doi.org/10.3354/esr00113>
- Rivalan, P., Godfrey, M. H., Prévot-Julliard, A.-C., & Girondot, M. (2005). Maximum Likelihood Estimates Of Tag Loss In Leatherback Sea Turtles. *Journal of Wildlife Management*, 69(2), 540–548. [https://doi.org/10.2193/0022-541X\(2005\)069\[0540:MLEOTL\]2.0.CO;2](https://doi.org/10.2193/0022-541X(2005)069[0540:MLEOTL]2.0.CO;2)
- Santafe, G., Inza, I., & Lozano, J. A. (2015). Dealing with the evaluation of supervised classification algorithms. *Artificial Intelligence Review*, 44(4), 467–508. <https://doi.org/10.1007/s10462-015-9433-y>
- Sarawak Forestry Corporation. (2023). *Location of Pulau Satang and Pulau Talang in Sarawak* [Drawn Figure]. https://www.sarawakforestry.com/seaturtle/Turtle_Handbook.pdf

- Sarker, I. H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, 2(6), 420. <https://doi.org/10.1007/s42979-021-00815-1>
- Schlägel, U. E., Grimm, V., Blaum, N., Colangeli, P., Dammhahn, M., Eccard, J. A., Hausmann, S. L., Herde, A., Hofer, H., Joshi, J., Kramer-Schadt, S., Litwin, M., Lozada-Gobilard, S. D., Müller, M. E. H., Müller, T., Nathan, R., Petermann, J. S., Pirhofer-Walzl, K., Radchuk, V., Rillig, M. C., Roeleke, M., Schäfer, M., Scherer, C., Schiro, G., Scholz, C., Teckentrup, L., Tiedemann, R., Ullmann, W., Voigt, C. C., Weithoff, G., & Jeltsch, F. (2020). Movement-mediated community assembly and coexistence. *Biological Reviews*, 95(4), 1073–1096. <https://doi.org/10.1111/brv.12600>
- Sea Turtle Conservancy. (2023). *Green sea turtle being released with a satellite transmitter* [Photograph]. <https://stcturtle.org/sea-turtle-tracking-works/>
- Shetty, D., C.A, H., Varma, M. J., Navi, S., & Ahmed, M. R. (2020). Diving Deep into Deep Learning: History, Evolution, Types and Applications. *International Journal of Innovative Technology and Exploring Engineering*, 9(3), 2835–2846. <https://doi.org/10.35940/ijitee.A4865.019320>
- Song, H., Sun, D., Chun, S., Jampani, V., Han, D., Heo, B., Kim, W., & Yang, M.-H. (2021). *ViDT: An Efficient and Effective Fully Transformer-based Object Detector* (arXiv:2110.03921). arXiv. <http://arxiv.org/abs/2110.03921>
- South African Tourism. (2009). *Cheetah With Tracking Collar* [Graphic]. https://commons.wikimedia.org/wiki/File:Cheetah_with_tracking_collar,_Kgalagadi_Transfrontier_Park,_Northern_Cape_%286253206536%29.jpg

- Tabuki, K., Nishizawa, H., Abe, O., Okuyama, J., & Tanizaki, S. (2021). Utility of carapace images for long-term photographic identification of nesting green turtles. *Journal of Experimental Marine Biology and Ecology*, 545, 151632. <https://doi.org/10.1016/j.jembe.2021.151632>
- Terven, J., Cordova-Esparza, D. M., Ramirez-Pedraza, A., & Chavez-Urbiola, E. A. (2023). *Loss Functions and Metrics in Deep Learning* (arXiv:2307.02694). arXiv. <http://arxiv.org/abs/2307.02694>
- Tetila, E., Bressen, K., Astolfi, G., Sant'Ana, D. A., Pache, M. C., & Pistori, H. (2020). *System for Quantitative Diagnosis of COVID-19-Associated Pneumonia Based on Superpixels With Deep Learning and Chest CT*. <https://doi.org/10.21203/rs.3.rs-123158/v1>
- Trappes, R. (2023). How tracking technology is transforming animal ecology: Epistemic values, interdisciplinarity, and technology-driven scientific change. *Synthese*, 201(4), 128. <https://doi.org/10.1007/s11229-023-04122-5>
- Truscott, Z., Booth, D. T., & Limpus, C. J. (2017). The effect of on-shore light pollution on sea-turtle hatchlings commencing their off-shore swim. *Wildlife Research*, 44(2), 127. <https://doi.org/10.1071/WR16143>
- Tufail, S., Riggs, H., Tariq, M., & Sarwat, A. I. (2023). Advancements and Challenges in Machine Learning: A Comprehensive Review of Models, Libraries, Applications, and Algorithms. *Electronics*, 12(8), 1789. <https://doi.org/10.3390/electronics12081789>
- Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., Van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I. D., Van Horn, G., Crofoot, M. C., Stewart, C. V., & Berger-Wolf, T.

- (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications*, 13(1), 792. <https://doi.org/10.1038/s41467-022-27980-y>
- Uno, K.-A. (2020). Wildlife Animal Tracking System using GPS and GSM. *International Journal of Engineering Research*, 8(17).
- Urban, P., Sabo, P., & Plesník, J. (2021). How to define ecology on the basis of its current understanding? *Folia Oecologica*, 48(1), 1–8. <https://doi.org/10.2478/foecol-2021-0001>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need* (arXiv:1706.03762). arXiv. <http://arxiv.org/abs/1706.03762>
- Wang, W.-C., Chen, L.-B., & Chang, W.-J. (2018). Development and Experimental Evaluation of Machine-Learning Techniques for an Intelligent Hairy Scalp Detection System. *Applied Sciences*, 8(6), 853. <https://doi.org/10.3390/app8060853>
- Welbourne, D. J., MacGregor, C., Paull, D., & Lindenmayer, D. B. (2015). The effectiveness and cost of camera traps for surveying small reptiles and critical weight range mammals: A comparison with labour-intensive complementary methods. *Wildlife Research*, 42(5), 414. <https://doi.org/10.1071/WR15054>
- Witherington, B., Bresette, M., & Herren, R. (2006). *Chelonia mydas* – Green Turtle. *Chelonian Research Monographs*, 3, 90–104. https://www.researchgate.net/publication/351945057_Chelonia_mydas_-_Green_Turtle
- Wu, X., & Liu, J. (2009). A New Early Stopping Algorithm for Improving Neural Network Generalization. *2009 Second International Conference on Intelligent Computation Technology and Automation*, 15–18. <https://doi.org/10.1109/ICICTA.2009.11>

- Wyneken, J., Epperly, S. P., & Higgins, B. (2010). *PIT Tag Migration in Seaturtle Flippers*.
- Yadav, S., & Shukla, S. (2016). Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, 78–83.
<https://doi.org/10.1109/IACC.2016.25>
- Yu, R., Wang, Z., Wang, Y., Li, K., Zhao, Y., Zhang, J., Song, G., & Chen, J. (2022). *Position Embedding Needs an Independent Layer Normalization* (arXiv:2212.05262). arXiv. <http://arxiv.org/abs/2212.05262>
- Zancato, L., Achille, A., Ravichandran, A., Bhotika, R., & Soatto, S. (2020). Predicting Training Time Without Training. *Neural Information Processing Systems*.
<https://doi.org/abs/2008.12478>.

APPENDICES

Appendix A: Tables of Results

Table 1: Training Results of Vision Transformer Model

Epoch	Vision Transformer Results			
	Train Loss	Train Acc	Validation Loss	Validation Acc
1	2.3177	0.222	1.7266	0.4125
2	0.8367	0.7797	0.4583	0.9039
3	0.3187	0.9358	0.2529	0.9464
4	0.17	0.9733	0.1991	0.9634
5	0.1124	0.9855	0.1252	0.9788
6	0.0761	0.9929	0.0957	0.9841
7	0.0472	0.9983	0.0841	0.9858
8	0.0313	1	0.0717	0.9858
9	0.0238	1	0.0673	0.9841
10	0.0193	1	0.0622	0.9823
11	0.0162	1	0.0681	0.9801
12	0.0141	1	0.0631	0.9806
13	0.0122	1	0.053	0.9858
14	0.0108	1	0.0635	0.9806
15	0.0096	1	0.0549	0.9811
16	0.0085	1	0.0503	0.9823
17	0.0077	1	0.0511	0.9823
18	0.0069	1	0.052	0.9823
19	0.0063	1	0.0495	0.9811
20	0.0057	1	0.0502	0.9806
21	0.0052	1	0.05	0.9811
22	0.0047	1	0.0457	0.9858
23	0.0043	1	0.0463	0.9823
24	0.0039	1	0.0464	0.9847
25	0.0036	1	0.0475	0.9823
26	0.0033	1	0.0459	0.9823
27	0.003	1	0.0415	0.9858
28	0.0027	1	0.0407	0.987
29	0.0025	1	0.0441	0.9835
30	0.0023	1	0.0405	0.9882
31	0.0021	1	0.0448	0.9835
32	0.0019	1	0.0457	0.9835
33	0.0017	1	0.0401	0.9858
34	0.0016	1	0.0466	0.9835
35	0.0015	1	0.0431	0.987
36	0.0013	1	0.0386	0.987

37	0.0012	1	0.0491	0.9823
38	0.0011	1	0.0396	0.9894
39	0.001	1	0.0422	0.9858
40	0.0009	1	0.0407	0.9858
41	0.0009	1	0.0414	0.987
42	0.0008	1	0.0431	0.9858
43	0.0007	1	0.0416	0.9847
44	0.0007	1	0.039	0.9894
45	0.0006	1	0.0466	0.9847
46	0.0005	1	0.044	0.9847
47	0.0005	1	0.042	0.9835
48	0.0005	1	0.041	0.9877
49	0.0004	1	0.0427	0.9858
50	0.0004	1	0.0384	0.9906

Table 2: Training Results of AlexNet Model

Epoch	AlexNet Model Results			
	Train Loss	Train Acc	Validation Loss	Validation Acc
1	1.9371	0.3267	1.1055	0.6453
2	0.8826	0.7034	0.6156	0.8007
3	0.5333	0.8223	0.3958	0.8766
4	0.3629	0.8777	0.2308	0.9193
5	0.2495	0.9189	0.2268	0.9193
6	0.1876	0.9436	0.1365	0.9561
7	0.1318	0.9564	0.1258	0.9573
8	0.105	0.9639	0.1324	0.949
9	0.1004	0.9682	0.1328	0.9632
10	0.0626	0.9818	0.0675	0.981
11	0.0635	0.9807	0.0981	0.9727
12	0.0559	0.9834	0.0877	0.9715
13	0.0347	0.9899	0.0489	0.9834
14	0.037	0.9872	0.0596	0.981
15	0.0323	0.9892	0.0429	0.9881
16	0.0304	0.9909	0.0528	0.9858
17	0.0164	0.9939	0.0737	0.9881
18	0.0191	0.9946	0.0285	0.9929
19	0.01	0.997	0.0313	0.9917
20	0.0147	0.9956	0.0354	0.9893
21	0.024	0.9929	0.0557	0.9786
22	0.0166	0.9932	0.0507	0.9905
23	0.0146	0.9949	0.088	0.9786
24	0.0433	0.9882	0.0406	0.9881
25	0.0147	0.9953	0.0352	0.9929
26	0.0067	0.9976	0.0298	0.9941

27	0.0141	0.9953	0.048	0.9893
28	0.0055	0.998	0.0296	0.9941
29	0.0073	0.9983	0.0809	0.9786
30	0.0154	0.9956	0.042	0.9905
31	0.0154	0.9949	0.0764	0.9881
32	0.0311	0.9892	0.0219	0.9953
33	0.0198	0.9953	0.0196	0.9929
34	0.0023	0.9993	0.0202	0.9941
35	0.0004	1	0.0206	0.9941
36	0.0007	1	0.0316	0.9941
37	0.0003	1	0.022	0.9964
38	0.0004	1	0.0238	0.9941
39	0.0005	1	0.0206	0.9953
40	0.0002	1	0.0182	0.9953
41	0.0004	1	0.019	0.9953
42	0.0002	1	0.0263	0.9953
43	0.0001	1	0.0188	0.9953
44	0	1	0.0207	0.9953
45	0	1	0.0224	0.9941
46	0	1	0.0255	0.9953
47	0.0001	1	0.0226	0.9941
48	0	1	0.0275	0.9941
49	0	1	0.0299	0.9941
50	0.0725	0.9774	0.0574	0.981

Table 3: Training Results of SqueezeNet Model

	SqueezeNet Results			
Epoch	Train Loss	Train Acc	Validation Loss	Validation Acc
1	2.3687	0.1997	2.1354	0.2527
2	1.7529	0.4304	1.2864	0.6679
3	1.159	0.6534	0.8838	0.7663
4	0.7965	0.7581	0.6435	0.8434
5	0.6261	0.8115	0.4899	0.8434
6	0.491	0.8527	0.3699	0.9027
7	0.419	0.872	0.2891	0.9253
8	0.3376	0.8997	0.2922	0.9205
9	0.2983	0.9145	0.2125	0.9371
10	0.2636	0.9257	0.2095	0.9514
11	0.2366	0.9324	0.1743	0.9478
12	0.219	0.9341	0.1732	0.9454
13	0.1887	0.9459	0.1543	0.9502
14	0.1802	0.947	0.1643	0.9454
15	0.175	0.95	0.1245	0.9609
16	0.1501	0.9601	0.1464	0.9609

17	0.1679	0.9517	0.1372	0.962
18	0.1286	0.9652	0.0991	0.9727
19	0.1205	0.9625	0.0911	0.9739
20	0.1089	0.9726	0.2174	0.9253
21	0.1026	0.9713	0.0888	0.9727
22	0.1179	0.9649	0.3741	0.8624
23	0.0955	0.9733	0.0965	0.9656
24	0.0858	0.9753	0.0607	0.9858
25	0.0921	0.9723	0.0684	0.9727
26	0.0897	0.9723	0.0838	0.9727
27	0.0829	0.9753	0.0609	0.9846
28	0.07	0.9811	0.0514	0.987
29	0.0705	0.976	0.0813	0.9703
30	0.0616	0.9807	0.0529	0.9846
31	0.0574	0.9834	0.0426	0.9858
32	0.0553	0.9834	0.0752	0.9846
33	0.0521	0.9845	0.0479	0.9846
34	0.0879	0.973	0.0568	0.9846
35	0.0381	0.9916	0.0594	0.9798
36	0.0424	0.9855	0.0557	0.9846
37	0.0608	0.9814	0.1088	0.9644
38	0.052	0.9841	0.033	0.9881
39	0.0211	0.9946	0.061	0.9751
40	0.0437	0.9865	0.0511	0.9858
41	0.0333	0.9922	0.0543	0.9858
42	0.0346	0.9902	0.0394	0.9893
43	0.0376	0.9875	0.0403	0.987
44	0.0242	0.9926	0.0906	0.9727
45	0.0282	0.9916	0.0839	0.9715
46	0.0312	0.9899	0.0456	0.9846
47	0.0292	0.9902	0.0266	0.9929
48	0.0155	0.9959	0.0231	0.9917
49	0.0133	0.997	0.0263	0.9929
50	0.0249	0.9902	0.0187	0.9941

Table 4: Training Results of ResNet50 Model

	ResNet50 Results			
Epoch	Train Loss	Train Acc	Validation Loss	Validation Acc
1	0.5544	0.8963	0.0395	0.9953
2	0.0563	0.9922	0.0112	0.9988
3	0.02	0.9986	0.0104	0.9964
4	0.016	0.9983	0.0022	1
5	0.0091	0.9966	0.0028	0.9988
6	0.0068	0.9966	0.0014	1

7	0.0052	1	0.0006	1
8	0.0058	0.9986	0.001	1
9	0.0096	0.9986	0.0128	0.9953
10	0.0138	0.9969	0.0021	1
11	0.004	0.9966	0.0008	1
12	0.0063	0.9986	0.0076	0.9976
13	0.0046	0.9986	0.009	0.9964
14	0.003	1	0.0066	0.9976
15	0.0026	0.9966	0.0033	0.9988
16	0.0021	0.9966	0.0021	0.9988
17	0.0018	0.9966	0.0058	0.9988
18	0.0064	0.9986	0.0003	1
19	0.0109	0.9969	0.003	0.9988
20	0.0045	0.9932	0.0019	1
21	0.0059	0.9966	0.0013	0.9988
22	0.0096	0.9973	0.0005	1
23	0.0038	0.9986	0.0003	1
24	0.0051	0.9966	0.0017	0.9988
25	0.0023	0.9932	0.0001	1
26	0.0011	0.9966	0.0001	1
27	0.0008	1	0.0001	1
28	0.0005	1	0.0001	1
29	0.0003	1	0	1
30	0.0032	0.9986	0.0096	0.9988
31	0.0023	0.9932	0.0001	1
32	0.0007	1	0.0005	1
33	0.0005	1	0.0001	1
34	0.0016	1	0.0001	1
35	0.0004	1	0	1
36	0.009	0.9973	0.0072	0.9953
37	0.0189	0.9946	0.0023	1
38	0.0045	0.9932	0.0042	0.9988
39	0.0008	1	0.0001	1
40	0.0005	1	0.0001	1
41	0.0004	1	0.0001	1
42	0.0003	1	0	1
43	0.0004	1	0	1
44	0.0067	0.9973	0.0001	1
45	0.0014	1	0	1
46	0.0003	1	0	1
47	0.0003	1	0	1
48	0.0001	1	0	1
49	0.0003	1	0	1
50	0.0002	1	0.0002	1

Table 5: Time Taken to Complete Each Epoch in Seconds for All Models

Epoch	Time Taken Per Epoch (seconds)			
	Vision Transformer	AlexNet	SqueezeNet	ResNet50
1	1370	36	34	37
2	1330	29	30	35
3	1325	31	32	39
4	1327	28	32	37
5	1327	33	32	32
6	1325	30	32	34
7	1331	31	32	32
8	1325	31	32	38
9	1325	30	32	35
10	1331	30	32	35
11	1328	31	33	35
12	1326	31	30	36
13	1328	31	32	37
14	1332	30	32	36
15	1327	30	32	36
16	1334	31	32	36
17	1328	31	33	36
18	1333	31	32	35
19	1336	31	32	36
20	1338	30	32	34
21	1336	31	33	36
22	1342	31	33	36
23	1339	30	33	36
24	1340	30	33	37
25	1334	31	33	38
26	1337	30	33	37
27	1294	31	32	35
28	1292	30	32	35
29	1306	30	30	35
30	1298	30	32	36
31	1301	30	31	38
32	1302	30	31	38
33	1289	31	32	38
34	1291	30	31	37
35	1303	33	31	36
36	1296	31	31	37
37	1285	31	32	36
38	1281	31	31	36
39	1286	31	31	36
40	1288	30	29	36
41	1293	31	32	36
42	1216	30	30	36
43	1300	29	31	35

44	1294	31	32	36
45	1299	31	31	36
46	1300	30	31	35
47	1299	30	32	36
48	1305	30	32	36
49	1270	30	32	37
50	1317	30	32	39

Table 6: Total Training Time Taken to Reach Every Epoch in Seconds for All Models

	From Start to The End of that Epoch (s)			
Epoch	Vision Transformer	AlexNet	SqueezeNet	ResNet50
1	1370	36	34	37
2	2700	65	64	72
3	4025	96	96	111
4	5352	124	128	148
5	6679	157	160	180
6	8004	187	192	214
7	9335	218	224	246
8	10660	249	256	284
9	11985	279	288	319
10	13316	309	320	354
11	14644	340	353	389
12	15970	371	383	425
13	17298	402	415	462
14	18630	432	447	498
15	19957	462	479	534
16	21291	493	511	570
17	22619	524	544	606
18	23952	555	576	641
19	25288	586	608	677
20	26626	616	640	711
21	27962	647	673	747
22	29304	678	706	783
23	30643	708	739	819
24	31983	738	772	856
25	33317	769	805	894
26	34654	799	838	931
27	35948	830	870	966
28	37240	860	902	1001
29	38546	890	932	1036
30	39844	920	964	1072
31	41145	950	995	1110
32	42447	980	1026	1148
33	43736	1011	1058	1186

34	45027	1041	1089	1223
35	46330	1074	1120	1259
36	47626	1105	1151	1296
37	48911	1136	1183	1332
38	50192	1167	1214	1368
39	51478	1198	1245	1404
40	52766	1228	1274	1440
41	54059	1259	1306	1476
42	55275	1289	1336	1512
43	56575	1318	1367	1547
44	57869	1349	1399	1583
45	59168	1380	1430	1619
46	60468	1410	1461	1654
47	61767	1440	1493	1690
48	63072	1470	1525	1726
49	64342	1500	1557	1763
50	65659	1530	1589	1802