**Identification and Damage Assessment of Architectural Heritage using Faster R-CNN with Inception ResNetV2 and DCN**

Peter Neil Malabay, Katrina Julie Ann Silao, Jay-R Zacarias

Computer Science Department, Technological Institute of the Philippines Quezon City

CS401: Thesis 2

Ms. Capule J. Thesis Adviser

May 2024

APPROVAL SHEET

This thesis entitled Identification and Damage Assessment of Architectural Heritage using Faster R-CNN with Inception ResNetV2 and DCN, prepared and submitted by Malabay, Peter Neil N., Silao, Katrina Julie Ann A., Zacarias, Jay-R V., in partial fulfillment of the requirements for the degree Bachelor of Science in Computer Science is hereby accepted.

| **Ms. Janice Capule** | |
| --- | --- |
| Research Paper Adviser | |
| **Ms. Hazel San L. Patilano** | |
| Chair, Defense Panel | |
| **Mr. Jess Garcia** | **Dr. Risty Acerado** |
| Panel Member | Panel Member |
|  |  |
|  |  |

Accepted and approved for the conferral of the degree **Bachelor of Science in Computer Science**

**Ms. Hazel San L. Patilano**

Dean, College of I.T. Education

Table of Contents

[APPROVAL SHEET ii](#_30j0zll)

[DEDICATION iii](#_19c6y18)

[ACKNOWLEDGMENTS iv](#_3tbugp1)

[TABLE OF CONTENTS](#_nmf14n)

[ABSTRACT x](#_37m2jsg)

[1 INTRODUCTION](#_1t3h5sf) 1

[LIST OF TABLES viii](#_1mrcu09)

[LIST OF FIGURES ix](#_46r0co2)

[1 INTRODUCTION](#_gjdgxs) 1

[1.1 Background of the Study](#_1fob9te) 2

[1.2 Objectives of the Study](#_3znysh7) 3

[1.3 Significance of the Study](#_2et92p0) 4

[1.4 Scope and Delimitations](#_tyjcwt) 5

[2 THEORETICAL FRAMEWORK](#_3dy6vkm) 8

[2.1](#_35nkun2) Review of Related Literature 8

[2.2](#_k5dbzw39r89a) Literature Map 18

[2.3 Conceptual Framework](#_3whwml4) 19

[2.4 Definition of Terms](#_2bn6wsx) 20

3 [OPERATIONAL FRAMEWORK](#_qsh70q) 24

[3.1 Project Design](#_3abgcdiocn1g) 24

[3.1.1 System Architecture](#_3as4poj) 24

[3.1.2 Algorithm Design](#_1pxezwc) 25

[3.2 Project Development](#_49x2ik5) 26

[3.2.1 Datasets and Sources](#_2p2csry) 26

[3.2.2 Data Processing Techniques](#_147n2zr) 26

[3.2.3 Algorithm](#_3o7alnk) 27

[3.2.4 Pseudocode](#_ihv636) 29

[3.2.5 Application Design](#_32hioqz) 30

[3.2.6](#_vx1227) Data 34

[3.2.7 User Interface Design](#_vx1227) 34

[3.2.8 Integration](#_3fwokq0) 35

[3.3 Testing and Operating Procedure](#_n1syvnbzdkqh) 36

[3.3.1 Hardware and Software](#_zaoku6rmi2k3) 36

[3.3.2 Comparative Analysis](#_zaoku6rmi2k3) 37

[3.3.3 Experimental Design 4](#_4f1mdlm)0

[3.3.4 Iterative Testing](#_xjzkmi9qnc9y) 41

[3.4 Project Evaluation](#_nm5h5c9a59w3) 42

[3.4.1 Evaluation Metrics](#_sfwrublpyexd) 42

[3.5 Work Plan](#_y1ovoyt794yj) 45

**Identification and Damage Assessment of Architectural Heritage using Faster R-CNN with Inception ResNetV2 and DCN**

## Introduction

Preserving the rich heritage of historical buildings in the Philippines faces multifaceted challenges, ranging from the aftermath of World War II to contemporary threats posed by urban expansion and environmental concerns (Cruz, 2019). In this complex landscape, the application of advanced technologies such as Deformable Convolutional Networks (DCNs) emerges as a promising solution. DCNs, a type of convolutional neural network, offer a flexible and adaptive approach to image analysis, making them particularly well-suited for the intricate details and irregular patterns often found in aged architectural structures.

The impact of World War II left scars on the nation's historical landscape, destroying many significant buildings. Now, as the Philippines grapples with the complexities of modern times, including rapid urbanization and environmental factors, heritage structures face new risks (Cruz, 2019). Incorporating DCNs into preservation strategies provides a technologically advanced toolset for identifying and addressing structural vulnerabilities (Kee, et al. 2022). These adaptive neural networks excel in recognizing complex patterns, enabling more accurate detection of issues like cracks or erosion that may threaten the integrity of historical sites.

The ongoing challenge lies in balancing the imperative to safeguard cultural assets with the requirements for growth and development. The unique history and architectural treasures of the Philippines exemplify this delicate balance. By embracing technologies like DCNs, preservationists can bridge the gap between tradition and progress, ensuring that the essence of the nation's historical narratives, embedded in its buildings, perseveres amidst the challenges of the modern era.

## Background of the Study

The Philippines, with its rich historical and architectural heritage dating back to the Spanish colonial period (1565-1898), is home to numerous churches, forts, and colonial-era buildings (Cruz, 2019). These structures not only served as places of worship and governance but also stood as symbols of Spanish influence and the conversion to Christianity. However, this diverse architectural legacy faces significant challenges in the modern era, including urbanization, natural disasters, and limited resources. Preserving and conserving this historical and cultural heritage is crucial for economic development, cultural identity, and public safety. These structures contribute significantly to the economy by attracting tourists and providing livelihoods through various revenue sources, such as entrance fees, guided tours, accommodations, and local crafts. Despite their importance, heritage buildings are susceptible to damage from natural elements, environmental factors, and human activities, necessitating structural health monitoring (SHM) to ensure their longevity and safety (Mishra et. al, 2020).

Machine learning has emerged as a valuable tool for SHM in heritage buildings (Mishra et. al, 2020). It offers the potential to reduce uncertainties, enhance structural assessments, and support informed decision-making for interventions. This application of ML in heritage building assessment is particularly relevant in the context of the Philippines, where historical structures are at risk due to contemporary challenges. A study has been conducted and was focused on Crisologo Street's preservation and heritage assessment in Vigan, Ilocos Sur, employing a multifaceted approach that combines qualitative research methods. By utilizing policy analysis, interviews, site inspections, and the study of pertinent legislation, the researchers aim to gain a thorough understanding of the government's role in heritage preservation. Additionally, descriptive and historical research methods will be employed to assess the current conditions and historical influences on Crisologo Street (Estonanto et. al., 2019).

In the broader context of artificial intelligence (AI) and deep learning, the study recognizes the potential of these technologies in replicating human behavior and comprehending the visual world (McNeely-White, 2019). The exploration of Convolutional Neural Networks (CNNs), including architectures like Inception and ResNet, highlights their significance in image recognition challenges. Moreover, the study delves into the advancements in object detection within computer vision, driven by deep learning architectures such as YOLO, SSD, R-CNN, and Faster R-CNN (Sojasingaraya, 2022). The introduction of deformable convolutions as a technique in object detection further enhances the adaptability of CNNs (Mishra et. al., 2020).

As the Philippines grapples with the complexities of modernity and the challenges posed by urbanization, this study aims to contribute to the development of a system that can detect architectural heritage and assess current damage. By doing so, it seeks to address the gaps in current knowledge and provide valuable insights for researchers, practitioners, and policymakers involved in the preservation of the nation's historical and cultural treasures.

## Objectives of the Study

The main objective of this study is to design a dual processing model of Faster R-CNN with Inception ResNetV2 and DCN for damage assessment of architectural heritage based on captured images.

Specifically, the study aims to:

1. Design Faster R-CNN with Inception-ResNetV2 model to automatically identify architectural heritage and Deformable Convolutional Networks for damage assessment.
2. Implement the model on a working mobile application.
3. Evaluate the effectiveness of the model using metrics like Precision, Recall, Mean Average Precision, and F-1 Score.
4. Evaluate the system using ISO/IEC 25010 in terms of (1) Functionality, (2) Usability, and (3) Performance Efficiency.
5. **Significance of the Study**

The study on the identification and damage assessment of Architectural Heritage has significant benefits for the following:

***Government Agencies***

Those responsible for preserving culture often struggle with decision-making and urban planning due to incomplete or outdated information on heritage structures and their condition. The proposed system will provide government agencies with real-time information on architectural heritage and damage assessment. Enabling them to make well-informed decisions for enhanced cultural preservation.

***Preservation Organizations***

The organizations face challenges in prioritizing restoration efforts and allocating resources efficiently, often leading to delays in protecting cultural heritage. The system’s precise identification and damage assessment capabilities enable preservation organizations to prioritize restoration projects based on urgency. Ensuring efficient resource allocation, facilitating timely protection and conservation of cultural heritage.

***Tourism Industry***

This sector faces challenges due to varying assessments of architectural heritage sites, affecting the overall tourist and economic benefits derived from these locations. Accurate identification and assessment provided by the system enhance tourist experiences. This reliability contributes to increased satisfaction, and economic growth for regions hosting heritage sites within the industry.

***Local Communities***

Often lack of comprehensive knowledge about the historical and cultural significance of architectural landmarks within vicinity. The system educates and engages local communities by providing detailed information about nearby heritage structures. This fosters a sense of cultural awareness as individuals understand the historical value of these landmarks.

***Future Researchers***

This study will serve as a valuable research guide for future researchers, providing a foundational resource for their investigations. Researchers can utilize the insights gained from this study to enhance environmental aspects and contribute to sustainable practices in their respective area of focus.

## Scope and Delimitations

This study is focused on systems that use machine learning to classify and assess damage in architectural heritage structures. These sites, reflecting the history, can suffer damage from various factors. The system offers an automated, efficient, and precise solution for this assessment.

***Scope of the Study***

The study focuses on harnessing advanced machine learning techniques for the classification and assessment of damage at architectural heritage sites. The aim is to develop an automated, efficient, and precise solution capable of not only identifying heritage structures but also assessing the extent of their damage. To achieve this, the researchers will gather high-resolution images of architectural heritage structures. These images will undergo a pre-processing phase, which includes resizing and normalization. Resizing standardizes the images to consistent dimensions, essential for uniform processing by the machine learning model. Normalization follows, wherein the researchers would adjust the pixel values to a common scale, typically between 0 and 1. This step is crucial for reducing computational complexity and enhancing the model's training efficiency, ensuring that variations in lighting or camera specifications do not bias the analysis.

Moreover, for model development, the researchers plan to create a multi-stage machine learning model. This model may integrate technologies such as Faster R-CNN with Inception ResNetV2 and Deformable Convolutional Networks (DCN), focusing specifically on detailed damage assessment. The designed model will undergo training and validation using a comprehensive dataset, enabling it to acquire the necessary knowledge for accurate damage assessment.

For evaluating the performance of the system, the researchers will utilize metrics such as Precision, Recall, Mean Average Precision, and F-1 Score. These metrics have been selected to ensure the high accuracy and reliability of the system. Furthermore, the researcher will engage in iterative testing procedures to continually refine the model. This method aims to improve both the accuracy and efficiency of the system, ensuring its effectiveness and reliability in assessing damage at architectural heritage sites.

***Delimitation of the Study***

The system faces certain limitations in its ability to detect structures such as modern buildings, houses, and dams. Additionally, environmental factors present during the image capture process, such as poor lighting conditions or adverse weather, can significantly compromise the accuracy of identification. Furthermore, users operating on limited hardware resources may encounter challenges. Specifically, those with less than 4GB of RAM or using an 8MP camera might experience slower processing times, alongside a noticeable reduction in detection accuracy. These constraints highlight the need for optimal conditions and adequate hardware capabilities to ensure the system functions effectively.

# 2 THEORETICAL FRAMEWORK

This chapter will discuss existing research and scholarly works relevant to this study. The purpose is to provide more information and a comprehensive overview of the current state of knowledge in this field. This chapter will also cover the conceptual framework where the researchers used the IPO framework to understand the flow of information within the system. Lastly, the definition of terms is established in this chapter to clarify the use of key terms within the context of the study, creating a common ground for the specialized language can bridge the gap between the researcher and the audience.

## Review of Related Literature

### *CNN-Powered Monument Detection Webapp for Preserving India’s Cultural Heritage*

In this paper, the researcher developed a web application that identifies Indian Monuments. The ***dataset*** used is gathered from online resources using an extension for Chrome called ‘Download All Images’. When using this extension, the researchers used specific keywords related to the monument to get accurate results. Filtering the downloaded data is necessary to remove insignificant and low-quality images while keeping high-quality images that clearly show the specific monument. The collected images are split into training and testing folders, the desired ratio for the split is 80:20, for training and testing. Annotating the gathered data with labels indicating the corresponding monument category is important in identifying the monuments.

This study by Anshul et. al (2023) also included the data pre-processing steps done for the dataset, it is done using a Keras library called Data Image Generator. The pre-processing steps included rescaling, shear transformation, zoom transformation, and horizontal flip. Then for the directories of the data, it used the data directory and class mode to specify the directories for the training and testing data. The parameter for the desired image input dimension is defined as well as the number of images that will be used for training and evaluation. Lastly, the class mode is defined as categorical to specify the organization of the images into multiple classes.

### *CNN-based Extraction of Filling Station Fuel Price Information from Captured Images*

The thesis of Fonseca et al. (2022) requires gathering and annotation of ***custom datasets*** for fuel price boards. The researchers had a target of 200 images per fuel brand for training and validation, having 600 images in total. To ensure the balance of the dataset, the collected images vary in distance, angles, and time of day. Then, the images are processed using the “Roboflow” website for annotation and data augmentation. The 600 images are split into 80:20 for training and validation, and the data augmentation applied to the raw dataset includes flip, crop, rotation, etc. resulting in a total of 1440 training images and 120 validation images.

To expand the diversity of the dataset, the researchers applied data augmentationtechniques using the “Roboflow” website. The said website has multiple augmentation techniques available that can be applied to improve the dataset, the researchers used flip, crop, rotation, shear, grayscale, brightness, and more data augmentation techniques to develop a more diverse dataset the “outputs per training example:3” which means that the website will generate three (3) versions of the sample image with varying data augmentations.

### *Automatic Detection of Welding Defects Using Faster R-CNN*

Image data preprocessing was used in research by Oh et al. (2020), where the authors discussed the need to segment high-quality radiographic testing images of welds to reduce their size, leading to improvements in learning rate and performance in a machine learning task. This segmentation process involved removing noise and focusing on the relevant defect data, resulting in an increased number of data points for training and validation.

According to Oh et al. (2020), studies have shown that Inception-ResNet has a high recognition and learning rate, making it an excellent feature extractor for algorithms aimed at detecting welding defects. The Inception-ResNet model learns faster than either Inception or ResNet when used independently. There are two versions of the model: Inception-ResNet V1 and V2. V1 combines the efficient Inception V3 model, which includes advanced features like factorized convolutions, with ResNet's deep network training capabilities and residual connections that prevent gradient issues. V2, on the other hand, merges the more advanced Inception V4, which offers more refined and uniform modules for improved feature extraction, with ResNet's ability to train deep networks, aiming for enhanced accuracy and the ability to train deeper networks without a loss in performance.

### *A comprehensive survey of the R-CNN family for object detection*

The R-CNN family of algorithms for object detection was discussed by Hmidani & Alaoui (2022) which evaluated the different R-CNN algorithms using mean average precision (mAP) and detection speed. The techniques used in the paper include region proposal, feature extraction, and classification using SVM and linear regression models. The experimental results show that the ***Faster R-CNN*** algorithm achieves improved accuracy and detection speed compared to R-CNN and Fast R-CNN with the use of Region Proposal Network (RPN).

### *Automatic Gun Detection From Images Using Faster R-CNN*

Alaqil et al. (2020) Faster R-CNN architectures for detecting guns in images, highlighting the Inception-ResNetV2 variant which achieved the highest mean Average Precision (mAp) at 82%, marking it as the most accurate but also the slowest, a potential limitation for real-time applications. The feature extractors used in this study were pre-trained on ImageNet for classification tasks. The last layers related to the classification task were removed to adapt them for Faster R-CNN. In contrast, the YOLOv2 model was noted for its speed, outperforming the Faster R-CNN with VGG16, which was the quickest among the R-CNN variants yet still behind YOLOv2. MobileNetV2 was specifically highlighted for mobile applications due to its smaller architecture, making it well-suited for mobile and embedded systems. The study underscored the trade-offs between precision and speed among the models, with some offering high precision at the expense of speed.

### *An Optimized Faster R-CNN Method Based on DRNet and RoI Align for Building Detection in Remote Sensing Images*

Bai et al. (2020) study explores the *Faster R-CNN* algorithm's utility in large-scale target detection. This deep learning method is celebrated for its high accuracy and efficiency when dealing with substantial target areas. It consists of two core components: the Fast R-CNN detection module and the Region Proposal Network extraction module. In this process, the RPN generates quality regions from feature maps, while the Fast R-CNN performs target detection. The workflow involves image input, feature map generation, RoI Pooling for feature extraction, and classification/regression, concluding with Non-Maximum Suppression (NMS) for result refinement. While Faster R-CNN has excelled in target detection, Bai acknowledges certain limitations. These include the gradual loss of edge texture information and inaccuracies in feature map extraction. To enhance performance, Bai suggests adopting deeper networks like ResNet and leveraging DenseNet to preserve texture details, aiming to improve building detection accuracy.

### *COVID-19 Detection in Chest X-Rays Using Inception Resnet-V2*

### In a study conducted by Badrahadipura et al. (2021) Utilizing the Inception-ResNet-v2 model, originally trained on the ImageNet database with over a million images across a hundred classes, the study replaces the traditional convolutional neural network approach. The adaptation and fine-tuning of the *Inception-ResNet-v2* model in this context demonstrates a notable accuracy, emphasizing the potential of AI in medical image diagnostics.

### *A Two-Stage Industrial Defect Detection Framework Based on Improved-YOLOv5 and Optimized Inception-ResNetV2 Models*

### Li et al. (2022) study focuses on detecting surface defects in steel, addressing the challenges posed by their complex shapes, small sizes, and similar characteristics. To achieve this, they propose a two-stage defect detection framework. Initially, they employed the Improved-YOLOv5 model to identify suspected defect areas on the steel surface. Subsequently, they use the *Optimized-Inception-ResnetV2* model for second-stage recognition within these areas, completing the surface inspection. The study emphasizes enhancements made to both models, resulting in significant performance improvements. In evaluations using the Enriched-NEU-DET dataset, the two-stage framework outperformed single-stage models, achieving a higher recognition accuracy. Real-world testing with robots and industrial cameras further validated the framework's superiority in defect detection. Future research will explore data augmentation and other methods to enhance recognition accuracy, particularly for defects with subtle differences.

### *A Real-Time Bridge Crack Detection Method Based on an Improved Inception-Resnet-v2 Structure*

### The research article by Wang, et al. (2022) introduces a real-time crack detection method based on an improved *Inception-Resnet-v2* structure. The proposed convolutional neural network model comprises four key modules: a feature extraction backbone network utilizing Inception-Resnet-v2, a multi-scale context information fusion module, a GKA clustering algorithm module, and a Dropout module. The Inception-Resnet-v2 backbone extracts crack features, incorporating the improved Inception structure and residual networks to prevent gradient divergence. The multi-scale context information fusion module enhances the detection of small cracks. Additionally, the GKA module accurately identifies target areas, reducing computational complexity, while the Dropout module combats overfitting. The paper explores how these components work together to improve bridge crack detection.

***Implementation of Faster R-CNN Inception ResNet V2 Algorithm for Human Body Pieces Detection***

A research article presented by Nabilah, et al. (2023) in implementing the ***Faster R-CNN*** Inception-ResNet V2 Algorithm for human body piece detection. The model is separated into several components, with the Inception-ResNet V2 renowned for its deep and efficient architecture that serves as the feature extractor and the Region Proposal Network (RPN) of Faster R-CNN generating rectangular region proposals from the feature map. The synergy of these two advanced technologies in the model offers a robust and precise solution, as evidenced by the impressive accuracy rate achieved in their experiments.

The Regional Proposal Network within Faster R-CNN utilizes anchor boxes of varying sizes and aspect ratios to scan the feature map, which is derived from the Inception-ResNet v2, for potential objects. Each position on the feature map can generate multiple region proposals through these anchor boxes. RPN has identified possible regions, Faster R-CNN takes over classifying the objects within these regions and refines the bounding box coordinates, ensuring a precise fit around each detected object. This step is for accurate object detection and involves both classification of the object type and regression to determine the exact bounding box.

For training this model, a pre-trained Inception-ResNet v2 network is utilized to initialize the convolutional layers of the RPN, leveraging the powerful feature extraction capabilities of the network. This pre-training helps in transferring knowledge from a large dataset to the specific task of human detection, providing a solid foundation for the model to learn from.

The entire modeling process is facilitated by Google Collaboratory, a cloud-based platform that offers free access to powerful computational resources, including GPU. This environment is beneficial for training deep learning models like Faster R-CNN, which require significant computational power, especially when working with large datasets and complex network architectures like Inception-ResNet v2.

### *Automatic Detection of Welding Defects Using Faster R-CNN*

In a study conducted by Oh, et al. (2020) The authors utilized ***Faster R-CNN*** for the automatic detection of welding defects in radiographic images. The author compared two high-performing architectures from ImageNet, ResNet, and Inception-ResNet V2 as backbone networks for Faster R-CNN used for feature extraction in the object detection process. The authors trained both ResNet and Inception-ResNet V2 using their training data. To improve the accuracy of their proposed algorithm, they applied data augmentation to each algorithm and evaluated their performance. They analyzed the convergence of the loss values for the training data and the mean Average Precision (mAP) of the algorithm for the evaluation data.

In the evaluation of the Faster R-CNN for welding defect detection conducted by Oh, et al. (2020), data augmentation was found to enhance detection accuracy significantly, with an improvement of 0.074. A key factor in the detection process was the adjustment of anchor box sizes to accommodate the small size of the defects. Among the models tested, ResNet with data augmentation emerged as the most effective, achieving a mean Average Precision (mAP) of 0.532. It was noted that the algorithm was more identifying consistency compared to the lack of fusion defects. There are plans to refine the accuracy of the algorithm by increasing the diversity of the image data and the scope of the algorithm's application beyond the shipbuilding industry where it was initially tested.

***An improved Faster R-CNN for defect recognition of key components of transmission line***

In a research article presented by Ni et al. (2021), Faster R-CNN, a renowned object detection framework, is discussed. This framework inputs an image into a convolutional neural network. After processing through shared convolution layers, the feature map bifurcates: one path leads to the Region Proposal Network (RPN) to generate candidate regions, while the other produces a higher-dimensional feature map. The RPN pinpoints potential object regions, which are subsequently directed to the ROI (Region of Interest) pooling layer. The study indicates a shift from the conventional feature extraction backbone of Faster R-CNN to Inception-ResNet-V2, a deep convolutional network known for its prowess in image classification, blending the Inception architecture with residual connections. This integration aims to extract intricate features from images, proving particularly effective for identifying transmission line defects. The researchers also searched into the optimization of the Inception-ResNet-V2 network to cater to their specific task, aiming to boost network efficiency and pinpoint accuracy for smaller targets. The top of these methodologies resulted in a recognition accuracy of 98.65%, underscoring the substantial advantages of incorporating Inception-ResNet-V2.

### *Cracks identification using mask region-based denoised deformable convolutional network*

A specific variation of Deformable Convolutional Network architecture is presented by Kee, et al. (2022) for crack identification. It combined instance segmentation and image classification and resulted in a Mask Region-Based Denoised Deformable Convolutional Network. The modified variation of DCN incorporates denoised deformable convolution, which improves the modeling capability of the convolutional layer. The proposed Mask R-DDCN achieved a lower validation loss and improved mean average precision (mAP) and crack classification accuracy. The study suggested that a more powerful backbone, such as ResNeXt, together with the DDCN can potentially enhance the performance of crack identification.

### *Earthquake Crack Detection From Aerial Images Using a Deformable Convolutional Neural Network*

A Study conducted by Yu, et al. (2022) for crack detection. A significant feature of the **DCN** is the deformable convolution. Unlike traditional convolutions, a deformable convolution introduces extra 2-D offsets into the standard convolution sampling grid. This adaptability means that the sampling locations can change, which is particularly beneficial for extracting features of irregularly shaped objects, such as earthquake cracks. The deformable convolutions are embedded in various stages of the Crack-CADNet, including the feature extractor, the feature optimization stage, and the upsampling stage.

### *A Remote Sensing Image Key Target Recognition System Design Based on Faster R-CNN*

For training, two prominent models, ResNet and Inception-ResNet V2, were utilized. To enhance the model's robustness, data augmentation techniques were applied, and the training was monitored using the mean Average Precision (mAP) metric. The endpoint of the training was identified when the mAP for the evaluation data peaked. The study's results were then visualized using a detection rate-recall graph, comparing the performances of ResNet and Inception-ResNet V2, both with and without the application of data augmentation.

### *Automatic detection of diabetic retinopathy: a review on datasets, methods, and evaluation metrics*

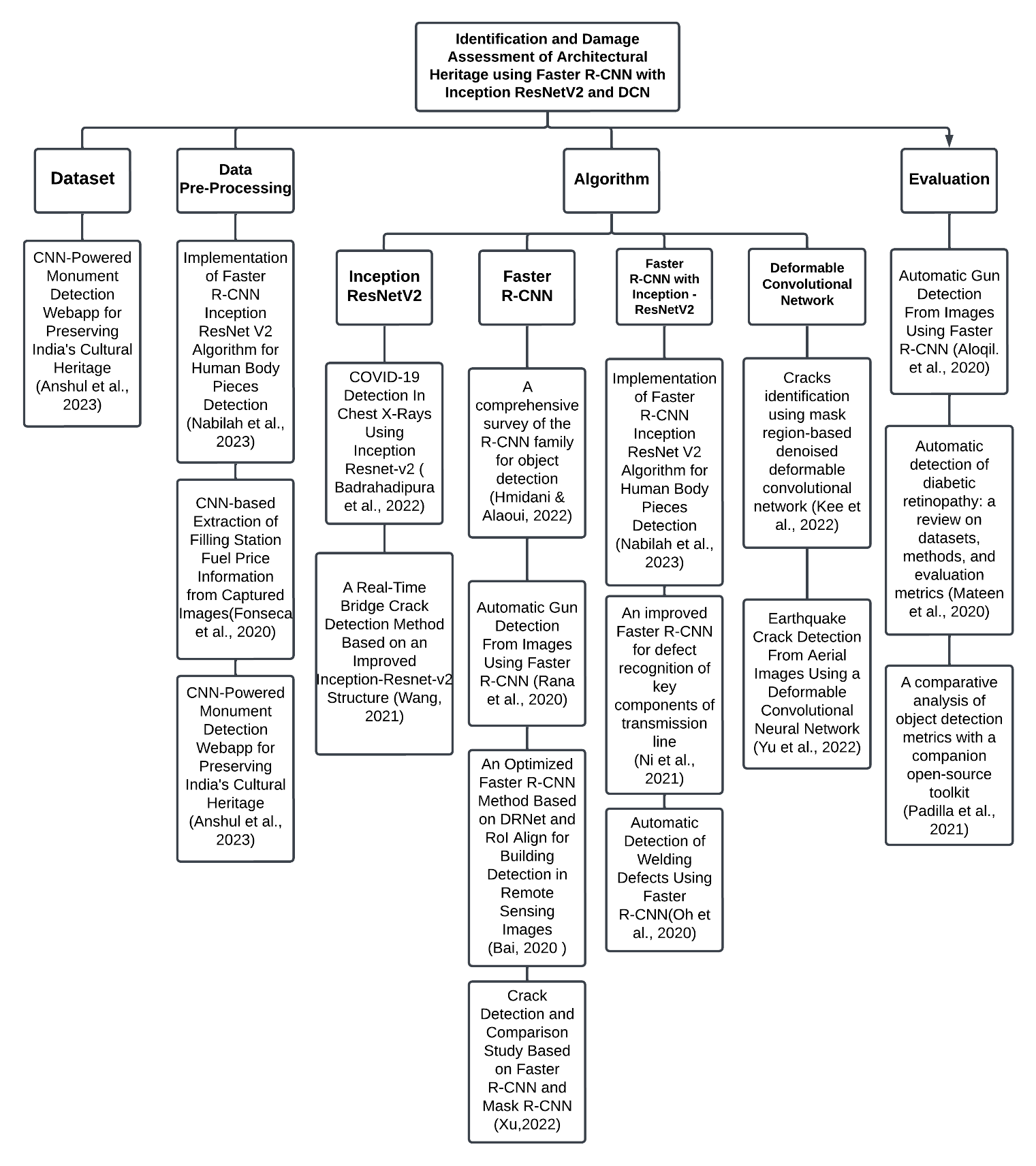
Diabetic retinopathy shows up as different signs in the eye, like small blood spots, bleeding, and other marks. The researchers will use an automatic system to find and name these signs in eye pictures. The mAP score tells us how good the system is at this job. For each sign, like small blood spots, the researchers would check how many the system finds and how many of those are correct. This gives us a score for that sign. Since diabetic retinopathy has many signs, averaging the scores for all of them to get the mAP. A higher mAP means the system is accurate in spotting and naming the signs in the eye pictures. It also means the system isn't making many mistakes, like missing a real sign or seeing something that isn't there.

1. **Literature Map**

In analyzing existing research for this study, the researchers organized the information into four main categories: Dataset, Data Pre-Processing, Algorithm, and Evaluation. To help understand all this information, the proponents used an organizing method called Literature Mapping. This is used to create a visual map that lays out the study reviewed in each category, making it easier to see how everything connects. This approach not only helps keep things organized but also provides a clear picture of how the different pieces of research fit together in the areas of Dataset, Data Pre-Processing, Algorithm, and Evaluation. By doing this, the proponents aim to build a solid understanding of the topic and present a well-structured framework that ties together insights from these important aspects of the study.

### Figure 1

### *Literature Mapping of Title*



1. **Concept of the Study**

Figure 2 represents the framework for the proposed system of Identifying Architectural Heritage and its Damage Assessment. There are three specific requirements for the input Captured Images, Architectural Heritage Dataset, and Damage Assessment Dataset. Process also includes the methodology that will be used in developing the application which is Agile methodology. The output is the expected result of the developed program, which is identifying architectural heritage using Faster R-CNN with InceptionResNetV2 and assessing its damage using DCN. The evaluation of the application will be done using Evaluation tools.

### Figure 2

### *Conceptual Framework*

A diagram of process and output

Description automatically generated

## Definition of Terms

## Term Definition

| Algorithm | In the context of computer science and machine learning, an "algorithm" is a set of predefined rules or instructions used to perform a specific task. In the system, the algorithm is a computational process that analyzes high-resolution images to identify architectural heritage structures and assess their damage. |
| --- | --- |
| Annotated Images | Images that have been labeled or annotated to emphasize specific items or features within the image, and which are frequently used to train machine learning models. |
| Architectural Heritage | Historical buildings, monuments, or landmarks with cultural, historical, or societal significance, frequently symbolizing the ideals and customs of numerous cultures throughout history. |
| Computer Vision Techniques | A collection of methods and algorithms that allow computers to analyze and grasp visual information from photos or videos and are commonly used for tasks such as object detection and categorization. |
| Convolutional Neural Networks | It is utilized for their ability to process and analyze images of architectural heritage, aiding in the identification and assessment of structural conditions and damages. |
| Cultural Heritage Conservation | It encompasses the efforts and methodologies for maintaining and preserving the physical and cultural integrity of historical buildings. |
| Dataset | a collection of high-resolution images that serve as the input data for the algorithms. These images, which capture various architectural heritage structures, are crucial for training, testing, and validating the machine learning model. |
| Damage Severity | The extent or level of damage or deterioration found in architectural heritage structures, which is often classified as slight, moderate, or severe. |
| Deep Learning Models | These models are integral to the system and used for complex tasks like image recognition and object detection in the context of architectural heritage conservation. |
| Deformable Convolutional Networks (DCN) | Represents a sort of convolutional neural network (CNN) that can alter its receptive field adaptively, improving its ability to recognize objects with complicated shapes or deformations. |
| Faster R-CNN | A deep learning system employed for efficient and accurate recognition of architectural heritage structures within images. |
| Fine-Grained Damage Assessment | Involves the careful evaluation and categorization of structural historical damage or deterioration, allowing for exact identification and prioritizing of restoration operations. |
| Heritage Conservation | This term underlines the goal of the system to aid in the conservation of cultural and historical heritage by providing a tool for assessing and managing the condition of architectural structures. |
| Historical Research Methods | This refers to providing a comprehensive background and context for the heritage sites being examined, aiding in understanding their significance and the implications of their preservation. |
| Inception-ResNetV2 | This architecture is used as a foundation for object detection models, aiding in the precise identification of heritage structures in the images processed by the application. |
| Machine Learning | System uses machine learning algorithms to process and analyze high-resolution images of architectural heritage structures. By learning from this data, the system can identify these structures and assess any damage, improving its accuracy and efficiency over time through continued learning and adaptation |
| Object Detection: | A major component in the automated identification of architectural heritage is the process of detecting and locating objects of interest within pictures or films. |
| Policy Analysis in Heritage Conservation | The use of policy analysis to understand the existing legislative and regulatory framework governing heritage conservation, identify gaps, and suggest improvements. |
| Preservation Strategies | Various approaches and techniques proposed or analyzed for maintaining the structural and historical integrity of heritage sites, especially in the face of challenges posed by modernization and environmental factors. |
| Restoration Efforts | The application's role in facilitating the restoration of damaged heritage structures by providing accurate damage assessments, which inform and prioritize restoration work. |
| Socio-Economic Advantages | Used to highlight the broader benefits of preserving architectural heritage, such as boosting tourism and local economies, which indirectly supports by facilitating damage assessment and restoration planning. |
| Structural Health Monitoring | Process of evaluating the condition of heritage structures, employing advanced technologies like machine learning to detect and assess damage or deterioration. |

# 3 OPERATIONAL FRAMEWORK

## Project Design

### *System Architecture*

**Figure 3**

*System Architecture*

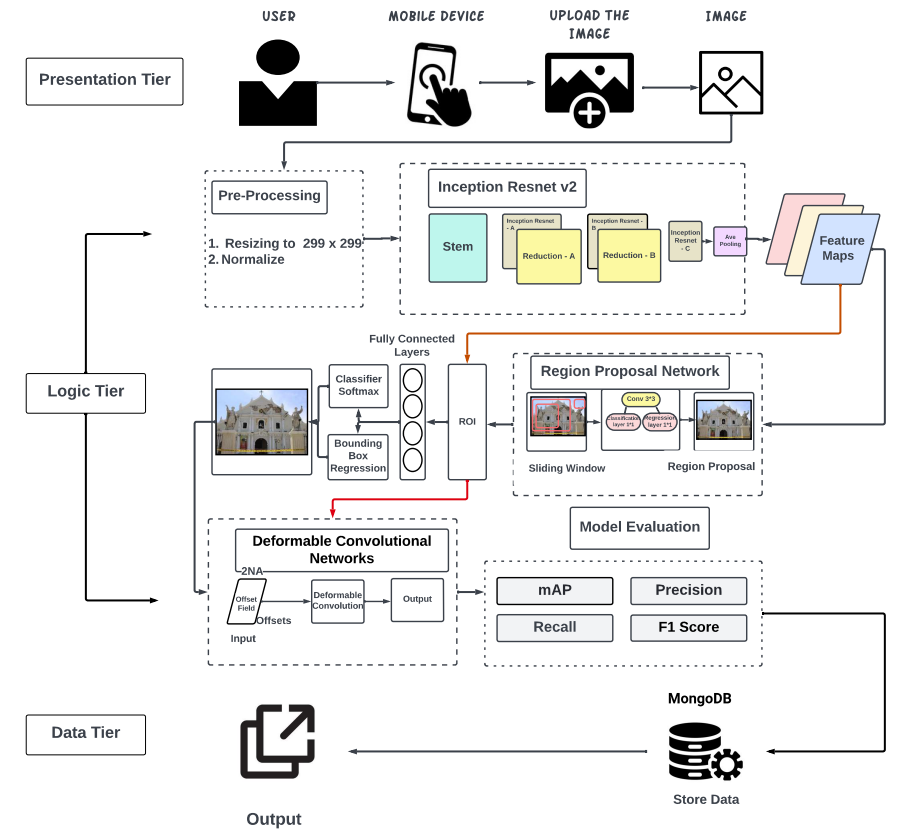


Figure 3 illustrates the system architecture, which begins with the uploading of an image. This image is pre-processed through resizing and normalization. Inception-ResNetV2 serves as the feature extractor, generating feature maps. These maps are then directed either to the Region Proposal Network (RPN) or directly to the Region of Interest (ROI). If the map is directed to the ROI, it is processed by the Deformable Convolutional Network (DCN), which is utilized to assess the damage. Data passed through the RPN, it subsequently feeds into the ROI and then moves to fully connected layers for object classification and bounding box regression. After classifying the input image, the data undergoes further processing by the DCN for damage assessment. Once processed by the DCN, the data is evaluated and information stored in MongoDB will be retrieved, and then output as the result.

### Algorithm Design

**Inception-ResNetV2.** It is a combination of the improved Inception structure and residual networks to prevent gradient divergence. This backbone is used for extracting features and is renowned for its deep and efficient architecture. It has been implemented in the Faster R-CNN Inception-ResNet V2 Algorithm for human body piece detection by Hanun, N. et al (2023). The synergy of Inception-ResNet V2 with Faster R-CNN offers a robust and precise solution, as evidenced by impressive accuracy rates in experiments.

**Faster R-CNN.** An advanced object detection algorithm. It integrates a Region Proposal Network (RPN) into its architecture, enhancing the detection process and speed. The RPN generates rectangular region proposals from the feature map. Faster R-CNN with Inception-ResNetV2 has been used to automatically identify architectural heritage in captured images (Hanun, 2023).

**Deformable Convolutional Network.** DCN introduces deformable convolutions. Unlike traditional convolutions, a deformable convolution introduces extra 2-D offsets into the standard convolution sampling grid. This adaptability means that the sampling locations can change, which is particularly beneficial for extracting features of irregularly shaped objects, such as earthquake cracks. The deformable convolutions are embedded in various stages of certain architectures, including the feature extractor, the feature optimization stage, and the upsampling stage. DCN has been used for damage assessment, especially in scenarios where the shapes and scales of damages can vary significantly (Yu, et al., 2022).

## Project Development

## *Datasets and Sources*

The dataset for Philippine Architectural Heritage Identification will be gathered from online resources using an extension from Chrome called ‘Download All Images’. This extension enables downloading images from Google Image Search by using specific keywords related to the Philippine Architectural Heritages. To verify the images gathered, the proponents will manually filter and select high-quality and relevant images for the specific heritage.

The collection of data for damage assessment will involve utilizing the 'Download All Images' extension, consisting of pictures that depict damaged architectural heritage. The process of collecting will be the same for the Architectural Heritage dataset, using Google Search Engine-related keywords to the damaged heritages will be used as a prompt. Gathered images will be verified by the proponents and proceed to annotating and labeling individual images.

## *Data Processing Techniques*

**Data Splitting.** The dataset will be divided into training and testing for the development of the model. The ratio for the dataset division will be 80:20, 80% of the dataset will be for training the model and 20% will be for testing the model. ([Badrahadipura](https://ieeexplore.ieee.org/author/37089279877) et a., 2022)

**Data Pre-Processing.** In the pursuit of developing a robust machine learning model for recognizing architectural heritage structures and assessing damage severity, the researchers have implemented a comprehensive data pre-processing procedure. This procedure begins with the standardization of all collected images, where the researchers would resize them to a uniform scale, specifically 299x299 pixels ([Badrahadipura](https://ieeexplore.ieee.org/author/37089279877) et a., 2022). This scaling step is crucial as it ensures consistency across the dataset, forming a solid foundation for reliable and effective model training.

To further enrich the dataset and enhance the model's ability to generalize across various scenarios, the researchers would employ a series of data augmentation techniques. These include rotations, which introduce varied angular perspectives to the images, and flipping the images horizontally or vertically, creating mirror-like reflections. It's important to note that these augmentation methods are not mere adjustments; they are strategic enhancements. By implementing these techniques, the researchers would significantly improve the model's performance, ensuring it can operate with high accuracy and adaptability in diverse real-world scenarios. (Fonseca et al., 2020)

## *Algorithm*

**Inception-Resnet v2**. These are known for their unique approach to convolutional neural network design. They use 'modules' that allow the network to choose from different kernel sizes (filters) within the same layer. This design enables the network to capture information at various scales, making it highly effective for image recognition tasks where the relevant features can vary significantly in size and complexity (Hanun, 2023).

**Region Proposal Network**. RPN Uses the processed feature map from the extractor to produce a set of rectangular region suggestions. The RPN uses a 3x3 kernel, which then branches into two 1x1 layers. One layer classifies regions to identify objects, while the other predicts the object's boundaries. As the RPN employs a sliding window approach, it generates k anchor boxes at each point on the previously obtained feature map. Each point signifies the center of the sliding window. These anchor boxes vary in size and shape, leading to challenges in classifying regions as containing objects or not (Hanun, 2023).

**Region of Interest**. Region of Interest or ROI produces a set of rectangular region suggestions based on the processed feature map from the feature extractor. ROI Pooling is applied, using a 2x2 kernel for max pooling on the proposed regions. This ensures consistent feature map sizes for each region, aiding in their classification. The component consists of two distinct fully connected layers. One classifies the proposed regions into specific categories, while the other adjusts the object's central coordinates, predicting the optimal fit. This results in bounding boxes that highlight the object's position and size (Hanun, 2023).

**Bounding Box Regression**. Employs a fully connected layer, split into two separate layers. The first layer categorizes each suggested region into specific categories, while the second focuses on regression for the center coordinates of the object. This regression refines the bounding box's dimensions, like its width and height, to better fit the target object. Essentially, the bounding box pinpoints the object's position and size (Hanun, 2023).

**Deformable convolution network**. DCN introduces modifiable convolutional layers, allowing for more flexible and adaptable feature extraction. This is especially crucial in damage assessment where the shapes and scales of damage can vary significantly. The DCN's ability to adapt its filters makes it particularly effective in identifying and categorizing various types of damage in images (Yu, et al., 2022).

## *Pseudocode*

The process involved in the identification and damage assessment of Architectural Heritages is outlined together with pseudocode for each step.

| Step 1: Initialize Faster R-CNN with Inception ResNet V2 Module.  faster\_rcnn = Initialize\_Faster\_RCNN\_Inception\_ResNetV2\_module()  Step 2: Initialize Deformable Convolutional Module.  dcn = Initialize\_Deformable\_Convolutional\_Network\_Module()  Step 3: Define a function to perform object detection using Faster R-CNN.  def Object\_Detection(image):  Step 4: Use Faster R-CNN with Inception ResNet V2 module to Identify structures in the image.  detected\_structure = faster\_rcnn.detect(image)  return detected\_structure  Step 5: Define a function to perform crack detection using the Deformable Convolutional module.  def Crack\_Detection(image, detected\_structures):  Step 6: Extract the regions of Interest around the detected structure for structure in detected\_structure.  roi = extract\_roi(image, structure)  Step 7: Apply DCN to adaptively detect cracks within the ROI structure.  detected\_cracks = dcn.detect\_cracks(roi)  Step 8: Combine the detected cracks to the structure object.  structure.add\_detected\_cracks(detected\_cracks)  return detected\_structures  Step 9: Main function to process an image and detect structures and cracks.  def Main(image):  Step 10: Perform object detection using Faster R-CNN.  detected\_structure = Object\_Detection(image)  Step 11: Perform crack detection using a Deformable Convolutional Network  detected\_structures = Crack\_Detection(image, detected\_structures)  Step 12: Output the results, including the identified structures and cracks  output\_results(detected\_structures) |
| --- |

## 

## Application Design

***Flowchart*** The flowchart functions as a visual guide, breaking down the systematic approach for heritage image analysis and damage assessment. Its primary purpose is to provide a straightforward overview of the analysis process; it serves as a formal tool to facilitate comprehension of the methodology employed for thorough heritage assessment.

**Figure 4**

*System Flowchart*

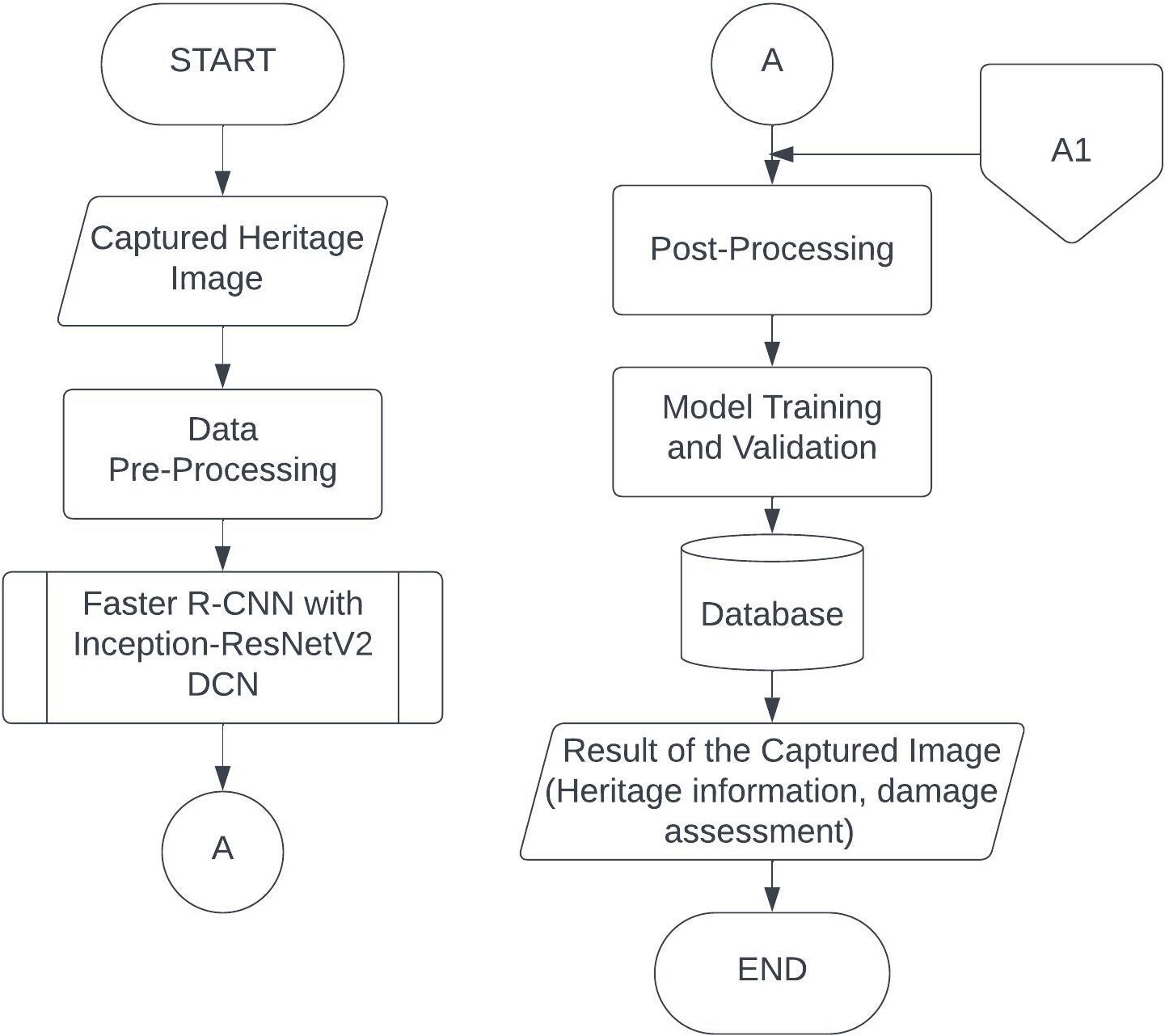


Figure 4 displays the system's flowchart, the process begins with the input of a captured heritage image, which serves as the basis for heritage information extraction and damage assessment. The captured heritage image undergoes data preprocessing to enhance its suitability for subsequent analysis. This step may include normalization and resizing to prepare the image for feature extraction. The preprocessed data is then fed into a dual processing pipeline involving Faster R-CNN with Inception-ResNetV2 and Deformable Convolutional Network (DCN). This combination allows for comprehensive analysis, including architectural identification and damage assessment. Post-processing refines raw predictions, eliminating duplicates and enhancing interpretability, ensuring a polished and accurate result. In the training and testing process, the system ensures the model's proficiency in correctly identifying and categorizing trees. Training involves exposing the model to data, allowing it to learn patterns. The model's performance is subsequently assessed on new, untested data, confirming its ability to generalize and make accurate predictions in real-world scenarios. Next step is to retrieve information from a database. This step involves retrieving detailed information from a database, encompassing historical or supplementary data related to the heritage site. The final output of the process encompasses a detailed result of the captured heritage image. This includes valuable heritage information such as architectural identification and a damage assessment.

### 

### Figure 5

### *Faster R-CNN with Inception-ResNetV2 and DCN Flowchart*

A screenshot of a computer screen

Description automatically generated

Figure 5 represents the flowchart, detailing a dual processing model that includes the initialization and utilization of both Faster R-CNN with Inception-ResNetV2 and DCN. The process initiates with the preparation phase, initializing the Inception-ResNetV2 model for subsequent feature extraction. The captured image serves as the input, ready to undergo processing for object detection. Inception-ResNetV2 is employed to extract rich and informative features from the input image, capturing hierarchical representations. The extracted features are then utilized by a Region Proposal Network (RPN) to propose potential regions of interest (ROIs) where objects might be present. The proposed ROIs undergo ROI pooling, creating fixed-size feature vectors for each ROI. Simultaneously, the ROIs are directed to two paths: one leading to a connector (B) and the other to the process of Classifier Softmax. The ROIs continue through the classification process, utilizing softmax to assign class probabilities to the detected objects. Bounding box regression is applied to refine the proposed ROIs, improving the localization accuracy. The final output is an image processed by the Faster R-CNN model with Inception-ResNetV2, incorporating object detections and bounding box predictions.

The second path of the flowchart begins with the initialization of DCN, the ROIs obtained from the initial Faster R-CNN process are used as input to the DCN, focusing on regions of interest identified during object detection. The DCN performs feature extraction on the ROIs, emphasizing adaptability in convolutional sampling. Deformable convolution is applied within the DCN, enabling flexible and adaptive convolutions based on learned offsets. The final output from this path is an image processed by the Deformable Convolutional Network, providing enhanced information for detailed damage assessment.

### 

### 

### *Data*

In the training of the Inception-ResNetV2 model for the detection of architectural heritage, the researchers will harness online resources with a Chrome extension known as ‘Download All Images’. This tool will process a diverse array of images necessary for a comprehensive training dataset. Once collected, these images will be verified to ensure they meet the quality and relevance criteria for the study. The selected images will then be used to train the model, with MongoDB serving as the database solution for storing the resulting data on damage detection.

The evaluation of the system will include the use of a validation dataset, separate from the training set, to impartially assess the model's accuracy and efficiency in identifying architectural heritage and evaluating damage. The validation dataset will carefully represent a range of scenarios, ensuring a robust test of the model's capabilities. Additionally, the evaluation of the model’s performance will involve quantitative metrics such as precision, recall, and F1-score, alongside qualitative assessments from domain experts to validate the practical applicability of the model in real-world conditions. Through iterative testing and evaluation, the proponents aim to ensure that the performance of the model will show significant results that indicate reliable output for detection models.

## *User Interface Design*

Figure 6 shows the main function of the program wherein the user would be able to input the image and click whether the user agrees to continue with the image. Wireframe Loading Displays a loading bar for the heritage damage analysis which allows the user to have a clear visualization of the current progress of damage assessment from the inserted heritage image. Wireframe Completed shows the damage assessment of the program and displays it within the bottom text section; the user also has the option to redo the image input or confirm the damage assessment reading.

### Figure 6

### *Wireframe*

A screenshot of a computer

Description automatically generated

## *Integration*

**Figure 7**

*Integration*

A diagram of a computer system

Description automatically generated

The system will have an analyzing page that holds three algorithms. When the user has successfully prompted an image, the user would then be able to decide whether to confirm the analysis or to resubmit an image. The system analyzes the image through a series of steps. The first step is the user must input an image and then the system would analyze the image by using the Faster R-CNN model with Inception ResNet v2 algorithms to classify it as a specific architectural heritage. When successfully classified, the system would then display the heritage information from the database and assess the damage severity using a Deformable Convolutional Network. The assessment result is shown on the user's screen. If the image is not recognized or classified, a message will be displayed to indicate that the image is not recognized by the system as a heritage.

## Testing and Operating Procedure

### *Hardware and Software*

**Hardware.** Table 8 presents the computers that will be used to train the model. These computers will have powerful RAM, CPU, and GPU to speed up the training process. While table 9 shows the specification of the mobile device that will be used in training this model.

***Table 8***

*Desktop Hardware Specifications*

| Hardware | Specification |
| --- | --- |
| OS | Windows 10 |
| CPU | intel i7 |
| RAM | 8GB RAM |
| GPU | GTX 1050 Gigabyte |

### Table 9

### *Mobile Phone Hardware Specifications*

| OS | Android V9 |
| --- | --- |
| CPU | 2.2 GHz Snapdragon 626 Octa-core |
| RAM | 4GB RAM |

**Software.** For the development and execution of the proposed system, certain software applications are essential. The researchers will be using Python, Visual Studio Code, and TensorFlow as software tools.

***Visual Studio Code*** is the code editor selected by the researchers to implement the system, incorporating the Inception-ResNet v2 Algorithm. With its fast startup time, powerful features, and user-friendly interface, Visual Studio Code stands out as the top choice.

***GitHub Desktop*** will be used to ensure smooth collaboration among the researchers by synchronizing the project files. The primary branch of the project will be consistently stored in the cloud and periodically updated with every significant modification.

***Python*** is a renowned general-purpose programming language that researchers will employ to transform the system into functional software. This language will serve as the foundation for writing the system, leveraging its diverse native components and libraries.

### *Comparative Analysis*

Assessing machine learning models like Faster R-CNN with Inception-ResNetV2 is essential for identifying their strengths, such as high accuracy and object detection efficiency, and weaknesses, like computational demands or limited performance under certain conditions. Sensitivity analysis is also crucial, as it determines how varying inputs or parameters affect performance, ensuring robustness and reliability. This evaluation helps optimize the model for real-world applications.

### Table 10

### *Mobile Phone Hardware Specifications*

| **Tools and Technologies, Computing processes, systems, and theory used** | **Strength/s** *(uniqueness and novelty)* | **Opportunities**  *(areas for improvement/possible enhancement/innovation/adaptation of technology/computing theory)* | **Title and Author/s** |
| --- | --- | --- | --- |
| **Inception-ResNetV2** | The combination of Inception and ResNet architectures is an efficient and effective approach in deep neural network design, offering a versatile solution that can be tailored to meet different needs with ease. This architectural fusion brings together the strengths of both models, creating a scalable and adaptable system capable of real-time inference, which is crucial for applications requiring rapid data processing. Its versatility extends across a wide range of domains, making it a popular choice for tasks like image classification and object detection. In a specific performance evaluation, Inception-ResNetV2 outperformed other models, achieving the highest accuracy rate of 89.56%, followed by Support Vector Machines (SVM) at 80.14% and Convolutional Neural Networks (CNN) at 88.23%, demonstrating the effectiveness of this architecture in delivering superior results. | Exploring the use of deep learning techniques, such as recurrent neural networks or attention mechanisms, may lead to a better understanding and representation of patterns in image recognition. The development of more robust and efficient algorithms for classification and regression models can improve the performance of recognition systems.  The development of this system will combine Faster R-CNN with Inception ResNet V2. Faster R-CNN shares convolutional layers between the RPN and the object detection network which allows efficient computation by reducing redundant processing. | Rishu Chhabra, Saravjeet Singh, Aditi Moudgil (2023) A Transfer Learning  Approach using Inception ResnetV2  https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10127302  Wan, X., Ren, F. & Yong, D., (2020). Using Inception-Resnet V2 for Face-based Age Recognition in Scenic Spots.  https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9073696 |
| **Faster R-CNN** | Faster R-CNN is a deep learning-based object detection framework. It is developed from the previous frameworks R-CNN and Fast R-CNN and now has significantly improved the accuracy and speed of detection. | Investigate the use of other network architectures apart from ResNet-50, for feature extraction to improve the performance of the model.  In developing this system, Faster R-CNN will be utilized with Inception ResNet V2 as the backbone. Inception ResNet V2 has high-level feature extraction and integrating it with Faster R-CNN could lead to better object detection accuracy. | Jianghong Tang, Yingchi Mao, Longbao Wang, Jing Wang (2019) Multi-task Enhanced Dam Crack Image Detection Based on Faster R-CNNhttps://ieeexplore.ieee.org/document/8981093 |
| **Deformable Convolutional Networks** | The introduction of DCNs (Deformable Convolutional Networks) to deep neural network architectures is a significant advancement that contributes to improved localization in images. This enhanced localization is particularly valuable when precise object detection or the separation of various elements within an image is essential. DCNs stand out due to their ability to adaptively adjust the sampling locations, allowing the network to concentrate on more informative regions within the input data, which results in more efficient and accurate feature extraction. Furthermore, DCNs are adept at capturing fine-grained details in the data, making them a valuable asset for tasks that demand high-resolution features. In summary, the integration of Deformable Convolutional Networks into neural network structures not only refines the accuracy of object localization but also enhances the network's capacity to focus on critical areas within the data and extract intricate details, making them particularly well-suited for applications that require precise and detailed feature analysis. | Deformable convolution, while beneficial for certain tasks, can pose computational challenges due to its inherent computational intensity. Therefore, it's imperative to focus on further research aimed at optimizing the computational cost of deformable convolution or exploring alternative methods that offer similar advantages while demanding fewer computations. Additionally, an essential area of improvement lies in enhancing the interpretability of the algorithm's results, particularly in cases where it generates false positives. Gaining insights into the decision-making process of the model can be invaluable for debugging and establishing trustworthiness in the system, ensuring that it not only performs effectively but also provides meaningful and actionable output for end-users and applications.  The integration of Deformable Convolution in crack detection, following the structure identification by Faster R-CNN with Inception-ResNet-v2, can significantly improve the model's ability to accurately and robustly detect cracks within structural environments. | Yao Wang, Ganggang Dong, and Hongwei Liu  SMALL SHIP DETECTION VIA DEFORMABLE CONVOLUTIONAL NETWORK  https://ieeexplore.ieee.org/ document/9553367 |

### *Experimental Design*

In the experimental design for architectural and damage assessment using Faster R-CNN with Inception ResNet v2 and DCN, the researchers will primarily adopt an iterative testing strategy. This approach will enable the researchers to continually improve the system's performance over time. The researchers will carry out a sequence of iterative tests, with each round building upon insights and improvements from the previous one. These iterative tests will facilitate the refinement of the algorithm, evaluation of its effectiveness, and reinforcement of its robustness. The researchers will closely monitor performance metrics like accuracy, precision, recall, and F1 score at each iteration to assess the system's progress. This method ensures that the architecture and damage detection are accurate and capable of adapting to the imputed image.

## *Iterative Testing*

1. **Data Collection.** The researchers will conduct search queries regarding specific architectural heritage structures. For damage detection, the researchers will collect images of damaged heritage structures.
2. **Testing and Evaluation (Iteration 1).** The researchers will use the acquired dataset to test the initial system, evaluating its performance using clear criteria such as mAP, precision, recall, and the F1 score. This evaluation will help determine the system's effectiveness in architecture detection and damage assessment. The outcomes of this initial testing will provide the researchers with a foundational understanding of the system's performance.
3. **Iterative Improvement.**  Upon reviewing the assessment findings, the researchers will pinpoint weaknesses and places where improvements can be made in the system. These problems may involve instances where the system wrongly identifies images or fails to detect an architecture or any other areas where the performance falls short. In this stage, the researchers will make adjustments, fine-tune parameters, optimize algorithms, or modify the warning system to rectify these issues.
4. **Re-Testing (Iteration 2).** After implementing improvements, the modified system will be re-tested on the same dataset used in the initial testing by the researchers. This is a critical step to determine whether the adjustments have resulted in better performance and to provide a basis for comparison with the initial version.
5. **Iterative Testing Cycles.** The researchers will perform multiple iterations of steps 2 and 4. In each iteration, they will analyze the outcomes, adjust the system, and conduct a re-evaluation. The number of iterations might differ based on the proximity of the system to reaching the predetermined performance objectives. If needed, they will conduct extra iterations.
6. **Validation.** Once the system has undergone iterative refinement and attains a satisfactory performance level, the researchers will subject the final version to validation using a separate dataset. This dataset will be distinct from the one utilized during iterative testing and will consist of previously unobserved data. The validation process will confirm that the technology effectively extends its performance to real-world architectures.

## 3.4 Project Evaluation

For a system that is used in identifying and assessing the damage to Architectural Heritage, it is important to use evaluation metrics that could test the performance of the system. The evaluation metrics that will be used to analyze the system’s performance are Precision, Recall, mAP, F1-Score, and ISO/IEC 25010. This is important to ensure that the system will perform following the objectives of this study.

### *3.4.1 Evaluation Metrics*

The evaluation metrics that will be used to assess the performance of the model are Mean Average Precision, Precision, Recall, F-1 Score, and ISO/IEC 25010 for the system. Precision is an evaluation metric in assessing the accuracy of object detection systems. The system aims to classify architectural heritage structures and assess the damage within an image and F1-Score is an evaluation metric that takes into consideration both precision and recall.

**Precision.** A measure of accuracy, revealing how many of the model's positive predictions are genuinely correct. High precision indicates that the model rarely makes mistakes when identifying something as positive.

(1)

**Recall.***It* measures how many of the actual positive items the model finds. High recall means the model rarely misses positive items.

(2)

**Confusion Matrix.** A table employed in the fields of machine learning and statistics to assess how well a classification algorithm is performing. (Ni et al., 2021)

**Table 11**

Confusion Matrix

|  | Prediction Positive | Prediction Negative |
| --- | --- | --- |
| Actually Positive | TP | FN |
| Actually Negative | FP | TN |

**Table 12**

*Definition of TP, FP, TN, FN*

| True Positive (TP) | indicates the number that the model predicted as positive, and it is correct. |
| --- | --- |
| False Positive (FP) | indicates the number that the model predicted as positive, and it is incorrect. |
| True Negative (TN) | indicates the number that the model predicted as negative, and it is correct. |
| False Negative (FN) | indicates the number that the model predicted as negative, and it is incorrect. |

***F1-Score****.* An evaluation metric that combines both Precision and Recall into a single value, resulting in a balanced value between these two factors.

(3)

**Mean Average Precision.** Itquantifies how well-ranked lists or object detection accuracy is in tasks such as image analysis.

(4)

**ISO 25010.** A questionnaire for an online survey will be developed. It will be employed to assess the recently proposed application. The system will be evaluated by two distinct groups: the technical group (Domain experts) and the non-technical group (if applicable). To gather additional data, pre-test and post-test measures will be used alongside the focused group discussion. For the decision criteria, the researchers will be utilizing the five-point Likert Scale of Interpretation. The lowest is the value of one (1) while the highest is the value of five (5). The Likert Scale of Interpretation will be used in interpreting the data to be gathered from the survey questionnaire.

***Functionality***. This will be responsible for evaluating if the prototype effectively fulfills its intended objectives in the context of architectural heritage identification and damage detection. Researchers can use this metric to assess how well the system matches predefined criteria or user expectations when identifying heritage structures and detecting damage.

***Usability***. This is utilized to evaluate the ease of interaction and user satisfaction with the software designed for architectural heritage identification and damage detection. It focuses on assessing how user-friendly the user interface is and how easily users can engage with its features. This metric includes the clarity of visual representations and the overall user experience in the context of architectural heritage assessment.

***Performance Efficiency***. This metric assesses the software's ability to perform effectively under specific conditions when identifying architectural heritage and detecting damage. Researchers can employ this metric to determine whether the prototype is optimized for swift and accurate heritage recognition and damage assessment within specified timeframes and with limited resources. This optimization ensures a smooth and responsive process for heritage identification and damage detection.

The Likert scale, ranging from 1 to 5, will be employed to evaluate system performance within the framework of ISO 25010. This scale encompasses five levels of agreement, from "Strongly Agree" to "Strongly Disagree." Users are presented with statements related to diverse aspects of system quality, and their responses are classified into these categories.

**Table 13**

Likert Scale

| Interpretation | Rating |
| --- | --- |
| Strongly Agree | 5 |
| Agree | 4 |
| Neither Agree nor Disagree | 3 |
| Disagree | 2 |
| Strongly Disagree | 1 |

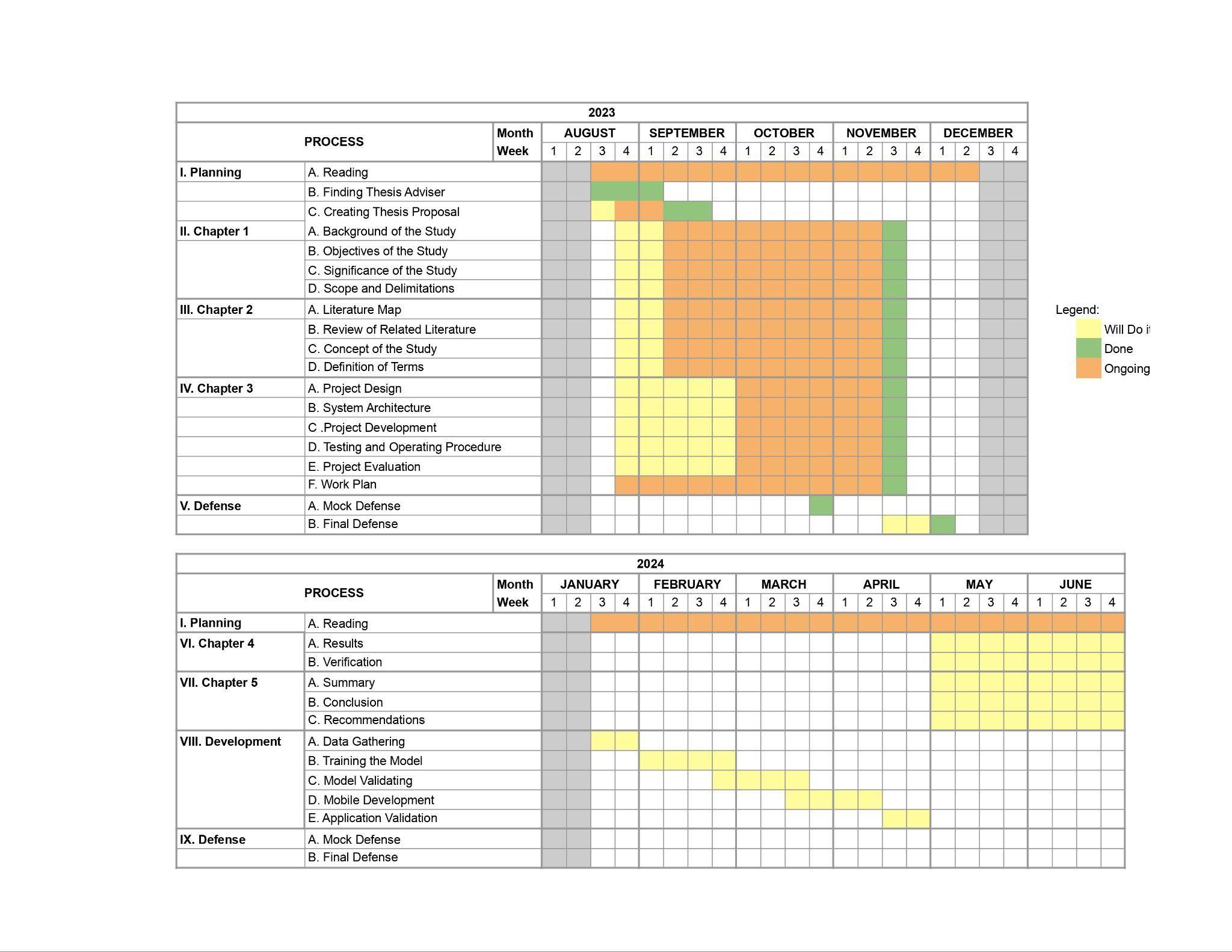
## 3.5 Work Plan

Figure 3.6 is the work plan of this paper, it outlines the process involving planning, writing of chapters 1-5, and preparing for the defense of this project. The work plan spans 10 months, starting in August 2023 and ending in June 2024. The development of this project is divided into 2 terms, the first term is from August to December 2023 which is for the writing of Chapters 1-3. The second term is from January to June 2024, for the development of the system which includes data gathering, model training, model validation, and mobile application development together with finalizing the entire Documentation for deployment.

***3.5.1 Gantt Chart***

**Figure 13**

*Work Plan*



REFERENCES

Alaqil, R. M., Alsuhaibani, J. A., Alhumaidi, B. A., Alnasser, R. A., Alotaibi, R. A. & Benhidour, H. (2020, November 1). Automatic Gun Detection From Images Using Faster R-CNN. Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9283780

Anshul, Bahuguna, A., S, G. P. M., Bhagnal, Y., Yeole, A. N., & Prabhu, B. (2023). CNN-Powered Monument Detection Webapp for Preserving India’s Cultural Heritage. https://doi.org/10.1109/icaisc58445.2023.10200440

Ãzgenel, Ã., & SorguÃ, A. G. (2018). Performance comparison of pretrained convolutional neural networks on crack detection in buildings. *Proceedings of the 35th ISARC*. https://doi.org/10.22260/isarc2018/0094

Badrahadipura, R., Septi, S. Q. N., Fachrel, J., Yulita, I. N., Pravitasari, A. A., & Agustian, D. (2021). COVID-19 Detection In Chest X-Rays Using Inception Resnet-v2. A. https://doi.org/10.1109/icaibda53487.2021.9689723

Bai, T., Pang, Y., Wang, J., Han, K., Luo, J., Wang, H., Lin, J., Wu, J., & Zhang, H. (2020). An optimized faster R-CNN method based on DRNET and ROI Align for building detection in remote sensing images. *Remote Sensing, 12*(5), 762. https://doi.org/10.3390/rs12050762

Copeland, B. (2024, January 26). *Artificial intelligence (AI) | Definition, Examples, Types, Applications, Companies, & Facts*. Encyclopedia Britannica. https://www.britannica.com/technology/artificial-intelligence

Cruz, G. R. (2019, February 22). A Review of How Philippine Colonial Experience Influenced the Country’s Approaches to Conservation of Cultural Heritage. Docsity. https://www.docsity.com/en/a-review-of-how-philippine-colonial-experience-influenced/8799826/

Dong, R., Yang, R., Tang, Y., & Shi, Z. (2018). A remote sensing image key target recognition system design based on faster R-CNN. *2018 3rd International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*. https://doi.org/10.1109/icmcce.2018.00040

Estonanto, C. D. L., Javier, J. C. V., Centeno, P. E. L., & Laquindanum, E. S. (2019). The Role of the National Government in the Preservation of the World Heritage Site of Crisologo Street Vigan Ilocos Sur Philippines. *Proceedings of the International Symposium on Social Sciences, Education, and Humanities (ISSEH 2018)*. https://doi.org/10.2991/isseh-18.2019.57

Hanun, N., Sarosa, M., & Asmara, R. A. (2023). Implementation of Faster R-CNN Inception ResNet V2 Algorithm for Human Body Pieces Detection. *2023 International Seminar on Intelligent Technology and Its Applications (ISITIA)*. https://doi.org/10.1109/isitia59021.2023.10220446

Hasanah, N. A., Atikah, L., & Rochimah, S. (2020). Functional Suitability Measurement Based on ISO/IEC 25010 for e-Commerce Website. *2020 7th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*. https://doi.org/10.1109/icitacee50144.2020.9239194

Hmidani, O., & Alaoui, E. M. I. (2022). A comprehensive survey of the R-CNN family for object detection. *2022 5th International Conference on Advanced Communication Technologies and Networking (CommNet)*. https://doi.org/10.1109/commnet56067.2022.9993862

Jiang, Y., Pang, D., & Li, C. (2021). A deep learning approach for fast detection and classification of concrete damage. *Automation in Construction, 128*, 103785. https://doi.org/10.1016/j.autcon.2021.103785

Kee, K. W., Lim, K. H., Lim, C. H., Lim, W. L., & Yap, H. E. (2022). Cracks identification using mask region-based denoised deformable convolutional network. *Multimedia Tools and Applications, 82*(3), 4387–4404. https://doi.org/10.1007/s11042-022-13422-w

Li, Z., Tian, X., Liu, X., Liu, Y., & Shi, X. (2022). A Two-Stage industrial defect detection framework based on Improved-YOLOV5 and Optimized-Inception-ResnetV2 models. *Applied Sciences, 12*(2), 834. https://doi.org/10.3390/app12020834

Mateen, M., Wen, J., Hassan, M., Nasrullah, N., Sun, S., & Hayat, S. (2020). Automatic Detection of Diabetic Retinopathy: A review on datasets, methods and evaluation metrics. *IEEE Access, 8*, 48784–48811. https://doi.org/10.1109/access.2020.2980055

McNeely-White, D. G., Beveridge, J. R., & Draper, B. A. (2020). Inception and ResNet features are (almost) equivalent. *Cognitive Systems Research, 59*, 312–318. https://doi.org/10.1016/j.cogsys.2019.10.004

Mishra, M. (2021). Machine learning techniques for structural health monitoring of heritage buildings: A state-of-the-art review and case studies. *Journal of Cultural Heritage, 47*, 227–245. https://doi.org/10.1016/j.culher.2020.09.005

Ni, H., Wang, M., & Zhao, L. (2021). An improved Faster R-CNN for defect recognition of key components of transmission line. *Mathematical Biosciences and Engineering, 18*(4), 4679–4695. https://doi.org/10.3934/mbe.2021237

Oh, S., Jung, M., Lim, C., & Shin, S. (2020). Automatic detection of welding defects using faster R-CNN. *Applied Sciences, 10*(23), 8629. https://doi.org/10.3390/app10238629

Oppermann, A. (2023, December 12). *What is deep learning and how does it work?* Built In. https://builtin.com/machine-learning/deep-learning

Padilla, R., Passos, W. L., Dias, T. L. B., Netto, S. L., & Da Silva, E. a. B. (2021). A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit. *Electronics, 10*(3), 279. https://doi.org/10.3390/electronics10030279

Sarraf, A., Azhdari, M., & Sarraf, S. (2021). A comprehensive review of deep learning architectures for computer vision applications. *ResearchGate*. https://www.researchgate.net/publication/349702934\_A\_Comprehensive\_Review\_of\_Deep\_Learning\_Architectures\_for\_Computer\_Vision\_Applications

Schroer, A. (2024, January 3). *What is artificial intelligence (AI)? How does AI work?* https://builtin.com/artificial-intelligence

Xu, X., Zhao, M., Shi, P., Ren, R., He, X., Wei, X., & Yang, H. (2022). Crack Detection and Comparison Study based on faster R-CNN and Mask R-CNN. *Sensors, 22*(3), 1215. https://doi.org/10.3390/s22031215

Wang, J., He, X., Shao, F., Lu, G., Hu, C., & Jiang, Q. (2021). A Real-Time bridge crack detection method based on an improved Inception-Resnet-V2 structure. *IEEE Access, 9,* 93209–93223. https://doi.org/10.1109/access.2021.3093210

Wan, X., Ren, F., & Deng, Y. (2019). Using Inception-Resnet V2 for Face-based Age Recognition in Scenic Spots. *2019 IEEE 6th International Conference on Cloud Computing and Intelligence Systems (CCIS)*. https://doi.org/10.1109/ccis48116.2019.9073696

Yu, D., Ji, S., Li, X., Yuan, Z., & Shen, C. (2022). Earthquake crack detection from aerial images using a deformable convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing, 60,* 1–12. https://doi.org/10.1109/tgrs.2022.3183157