HAWASSA UNIVERSITY INSTITUTE OF TECHNOLOGY

FACULTY OF INFORMATICS

DEPARTMENT OF COMPUTER SCIENCE

DETECTION AND CLASSIFICATION OF INDIGENOUS ROCK MINERALS USING DEEP LEARNING TECHNIQUES

MASTER OF SCIENCE THESIS

BY

HADIA BERKA WAKO

Hawassa, Ethiopia

May 18, 2023

DETECTION AND CLASSIFICATION OF INDIGENOUS ROCK MINERALS USING DEEP LEARNING TECHNIQUES

BY

HADIA BERKA WAKO

A THESIS SUBMITTED TO THE
DEPARTMENT OF COMPUTER SCIENCE,
FACULTY OF INFORMATICS, SCHOOL OF
GRADUATE STUDIES, HAWASSA UNIVERSITY
HAWASSA, ETHIOPIA

IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE
DEGREE OF
MASTER OF SCIENCE IN COMPUTER SCIENCE

MAY 18, 2023

SCHOOL OF GRADUATE STUDIES

HAWASSA UNIVERSITY

ADVISORS APPROVAL SHEET

This is to certify that the thesis entitled "Detection and Classification of Indigenous Rock Minerals Using Deep Learning Techniques" submitted in partial fulfillment of the requirements for the degree of Masters of Science in Computer Science, the Graduate Program of the Department of Computer Science, and has been carried out by Hadia Berka Wako Id. No PG/CSR/013/10 under my supervision. I recommend the student to submit the thesis to the department as the student has met the necessary requirements.

Dr. Dereje Teferi (Ph.D.)	[signature]	May 23, 2023
Name of advisor	Signature	Date

DECLARATION

I declare that this thesis entitled "Detection and Classification of Indigenous Rock Minerals using Deep Learning Techniques" is prepared by me, with the guidance of my advisor. The work presented here is entirely original to me, and has not been submitted, in whole or in part, for any other degree or certification.

Hadia Berka Wako

May 18, 2023

This thesis has been submitted for examination with our approval as university advisor.

Dereje Teferi (Ph.D.)

May 23, 2023

SCHOOL OF GRADUATE STUDIES

HAWASSA UNIVERSITY

EXAMINERS APPROVAL SHEET

We, the undersigned, members of the Board of Examiners of the final open defense by Hadia Berka Wako have read and evaluated his thesis entitled “Detection and Classification of Indigenous Rock Minerals Using Deep Learning Techniques” and examined the candidate. This is, therefore, to certify that the thesis has been accepted in partial fulfillment of the requirements for the degree

______________________	__________	___________
Name of the Chairperson	Signature	Date
Dr. Dereje Teferi		August 9, 2023
Name of Major Advisor	Signature	Date
______________________	__________	___________
Name of Internal Examiner	Signature	Date
Michael Melese (PhD)		July 24, 2023
Name of External examiner	Signature	Date
______________________	__________	___________
SGS Approval	Signature	Date

Final approval and acceptance of the thesis is contingent upon the submission of the final copy of the thesis to the School of Graduate Studies (SGS) through the Department/School Graduate Committee (DGC/SGC) of the candidate’s department.

ACKNOWLEDGMENT

First, I have everlasting thanks to Almighty Allah for giving me the moral, psychological, and spiritual strength to accomplish this research work. Next to him, I would like to thank my adviser from the bottom of my heart Dr. Dereje Teferi, for his encouragement and continuous guidance for writing of this thesis. Besides my advisor, I would like to extend my thanks to Mr. Wegene Shankute, who stands by my side by giving me insightful comments. Likewise, I am also grateful to all of the instructors I had at Hawassa University during my M.Sc. study, you all have an input for my study and you showed me the way through which I fulfill my dream.

I would like to thank my brother Daniel Tolera for guiding me in different ways and providing me all the necessary Mineral data and references for this study. I am very lucky to have such special brother like you.

Finally, I am also immensely grateful to all of my family members for your love, prayers and unconditional support for me as usual. Last but not least, I welcome this opportunity to thank my friends, colleagues and people who supported me directly or indirectly in the completion of this work.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

ABSTRACT

Ethiopia is undoubtedly a place of riches, with a vast and diverse landmass that is rich in resources. However, less attention has been given in utilizing computing discipline like Artificial Intelligence to solve the current problems in the area of mineral mining in Ethiopia. GUJI Zone is one of Oromia 20 administrative zones blessed with different mineral resources. Despite the fact that mineral has lions share contribution to economy of Ethiopia, little work is done in modernizing the mining industry in Ethiopia especially in empowering small-scale Artisanal community. GUJI is one of the zones following outmoded techniques to identify minerals in mining industry.

Rock mineral detection and classification employing conventional methods involves testing physical and chemical properties at both the micro- and macro-scale in the laboratory, which is expensive and time-consuming. Identifying tiny rock minerals and detecting its originality using traditional procedure and techniques takes too much time. Identification of minerals merely through visual observation is often erroneous. To address these problems, a deep learning approach for the classification and detection of Rock Minerals is proposed. The design- science research methodology is followed to achieve the objectives of the research. To conduct this study, 2000 images were collected from Guji's zone and Mindat.org website. After collecting the images, image pre-processing techniques such as image resizing, image segmentation using roboflow, and image annotation are performed. Moreover, data augmentation is applied to balance the dataset and increase the number of images. This research work focuses on classifying and detecting fifteen types of rock minerals. Based on YOLOv7 deep learning model we have used 70% of the dataset to train the model and 30 % of the dataset to test the performance of the model. Finally, the developed model is evaluated using accuracy, precision, recall, and mAP with other models. Experimental result shows that the accuracy obtained from YOLOv7 is 76%mAP for large objects comparing to other models. Consequently, the pretrained weight of yolov7 achieved a 97.3% accuracy in classifying and detecting with other images.

Keywords: Rock Minerals, Deep Neural Network, pretrained weight, training from the scratch, YOLOv7

CHAPTER ONE

1. Introduction

1.1 Background of the study

A mineral is a naturally occurring substance with distinctive chemical and physical properties, composition and atomic structure [1]. Minerals are solid substances that have formed naturally in the Earth. They often have a consistent crystal structure as well as distinctive chemical and physical characteristics [2].

Ethiopia is definitely a place of riches, with a vast and diverse landmass that is rich in resources. Geological investigations of Ethiopia's mineral resources have revealed formations that hold enormous amounts of mineral resources, including proven mines for industrial minerals, coal, precious and other metals. Ethiopia is one of a country flourished with minerals resources such as Gold, Platinum, Opal, Emerald, Oil & Gas, Niobium, Copper, Tantalum, Chrome, Nickel, Manganese, Limestone, Sandstone, Gypsum, Clay, Laterite, Iron Ore, Perlite, Halite, Potash, Oil Shale, Lignite, Diatomite and Besntonite [3].

Mineral identification is a fundamental task in geology, mining engineering, and other related fields. The cutting-edge requirements of scientific research and industry demands are reflected in the intelligent identification of minerals. The present tendency to combine computer science and earth science, and the use of artificial intelligence in connection with the history of digital earth science in deep time has attracted a lot of attention [4].

Geoscience is one of the fields in which there has also been a great deal of effort devoted to the combination of traditional problems and artificial intelligence, such as using artificial intelligence methods for the identification of rocks and minerals. For traditional methods of mineral identification, it takes a considerable amount of time and energy, as well as the use of expensive and specialized instruments to obtain the data needed to ensure the accuracy of the identification. Despite having the potential to contribute to intelligent mineral identification, it has been challenging to adopt artificial intelligence methods in the identification of minerals, and there has been little progress in that direction. In recent years, breakthroughs in artificial intelligence, including remarkable advances in deep learning methods made the development of more and easier-to-use systems possible [4]. Numerous exploratory studies have been undertaken to identify minerals. Mineral identification is the process of mineral selection, exploration, separation, and related artefact protection in archaeology [4]. These days, low-skilled, small-scale miners in Ethiopia are largely in charge of mining, using low-tech equipment and inputs [5].

Automatic identification technology could play a significant role in quality inspection of mining industry. These technologies reduce the hazardous effect of rock minerals to people by reducing physical contact with the rock minerals. The use of automatic identification technology is not questionable for its efficiency and effectiveness as compared to human based traditional way of mineral identification.

With the advancement of computing technology deep learning and image processing can automatically detect and identify Rock minerals, playing a significant role in the automatic classifying and detection of Rock minerals [6], [7], [4] , [8]. Researchers have applied image processing and deep learning to identify and categorize Rock minerals. A deep learning model and clustering algorithm are combined to create the enhanced rock mineral recognition approach suggested in [9]. The model recognized rock minerals by integrating the two algorithms deep learning and YoloV5 [8] .YOLOv5 follow the original anchor-based pipeline of YOLOv3. However, the anchor mechanism has many known problems. The researchers have used color and shape features for recognizing the rock minerals.

1.2 Motivation of the study

The researcher was inspired to conduct this study because traditional method of mineral identification is a challenging task that requires, experienced professionals, extensive resources and is time consuming. The iniquitousness of artificial intelligence discipline has been transforming how the business works by lessening the workloads on professional experts for the task that over utilize time and resources. Recently, the mining industry in Ethiopia is one of the industries that have been following legacy method of mineral identification. The huge libraries of minerals data, which are suitable for artificial intelligence works like, mindat.org, has been setup and is available online. Despite, this huge amount of data has the potential to revolutionize the mining industry; the data are under-utilized in solving legacy method of mineral identification in Ethiopia. Therefore, in this study intelligent identification of indigenous rock minerals is developed using deep learning techniques. An only a few researches are conducted on identification of minerals internationally and no research has been found yet locally.

This study will empower small scale artisanal miners' communities in GUJI Zone who are legally recognized as member of particular mining cooperatives with a greater understanding of the value a mineral.

1.3 Statement of the problem

Ethiopia is clearly a place of riches, with a huge and diverse landmass that is rich in resources [10]. However, less attention has paid in utilizing computing discipline like Artificial Intelligence to solve the current problems in the area of mineral mining in Ethiopia.

GUJI Zone is one of Oromia 20 administrative zones that is flourished with different mineral resources, that is, GUJI (Gold, Emerald, Beryl group, Chromite, Feldspar, Quartz, Kaolin, Dolomite, Tantalum, Aquamarine, Fluorite, Graphite etc.) [10]. Minerals contribute a lion's share to the country's economy. However, little work is done in modernizing the mining industry in Ethiopia especially in empowering small-scale Artisanal community. GUJI Zone is one of the zones following outmoded techniques to identify minerals in mining industry.

Rock mineral detection and classification in traditional approaches include testing physical and chemical qualities at the micro- and macro-scale in the lab, which is expensive and time-consuming. Identifying tiny rock minerals and detecting its originality using traditional procedure and techniques takes too much time. Identification of minerals merely through visual observation is erroneous and hence the procedure is carried out through intensive laboratory examination, which requires too much resources and labors. Moreover, the procedure could be affected by the problems of strong expert subjectivity, high dependence on experience, heavy workload, long identification cycle, and incapability to achieve complete and accurate quantification. Therefore, automated system helps to enhance work productivity by decreasing expert's workloads and boosting efficiency. Hence, improving decision making and lessening material dependency.

When we say strong expert subjectivity, sustainably supplying people with mineral resources is an important task for the present and future. Specialists are needed to develop efficient mining and preparation methods as well as intelligent recovery systems, in order to ensure sustainable handling of the Earth's resources. Graduates from the field of mineral resources and disposal engineering contribute greatly to making sure the following generations' need for mineral resources can be met [11].

High dependence on experience, becoming a mineralogist involves a combination of education, skills development, and experience in the field of mineralogy, which is the study of minerals and their properties, formation, and distribution. Mineralogists play critical roles in various industries, such as mining, oil and gas, environmental consulting, and research.

When we say heavy work load, traditional identification depends on experts' visual observation of rock minerals under the optical microscope, and it is imperfect in several aspects. First, the workload is diffuclt to complete with limited human resources if there are numerous rock mineral samples to be identified [12].

Mineralogists use a number of mineral properties to identify minerals. The properties most commonly used in identification of a mineral are color, streak, luster, hardness, crystal shape, cleavage, specific gravity and habit. So it takes long identification cycles to know wether it is mineral or not. The number of experienced professionals is decreasing, and there is a shortage of young successors. Identification of minerals by relying solely on image of minerals is not sufficient due to color property, which plays a significant role to identify minerals using deep learning techniques.

The identification based on naked-eye observation is person-dependent and cannot ensure accurate and quantified results. In traditional identification, a single field of view is observed under an optical microscope and then the overall situation is inferred based on the observation of multiple single fields of view. The identification results of the same sample may be different due to the differences in appraisers' experiences and perceptions. So leads to incapability to achieve complete and accurate quantification of rock minerals.

At the macroscopic level, it can be challenging to distinguish between minerals of the same species since they may have various colors and shapes, while minerals of different species may have comparable colours, shapes, and textures. For instance, the various impurities in allochromatic minerals may cause them to have varied colours. As a result, obtaining high accuracy using solely photo attributes is challenging. Studies that employ numerous attributes to increase accuracy have surfaced in a variety of other contexts. [7], [6], [9] and have shown improved effectiveness.

Therefore, having a detection model will help mineralogists and geologists to have timely detection and most importantly for fast and accurate identification of rock minerals. Accordingly, developing a classifier model that can accept a given mineral image and classifies the mineral based on the characteristics of the image through a system becomes very important. Towards solving the stated problem, some researches are conducted previously.

In [8] the researchers proposed detection of minerals using images luminance equalization that used light intensity. A total number of 220,057 images were collected with 50 mineral species from mindat. The researchers used preprocessing techniques for the quality of data such as combined histogram equalization (HE). In addition, the authors implemented the Laplace algorithm called deep learning yolov5 model. Their model achieved an accuracy of 95.6%.

There are more than 11 methods of identifying minerals such as crystals, cleavage & fracture, color, hardness, streak, luster, specific gravity, tenacity, acid test, magnetism and fluorescence [13]**.** These methods are error prone and time consuming. Therefore, we proposed a detection and classification method that uses mineral images and a deep learning technique to identify the minerals.

The task of mineral identification and classification traditionally requires highly trained experts who are equipped with material like scanning electron microscopes and optical microscopes [4]. Advances in machine learning and deep learning make it possible to envision the automation of many complex tasks in various fields of science at accuracy equal or better to that of human performance, thereby, avoiding placing human resources into tedious and repetitive tasks, improving time efficiency, and lowering costs. Therefore, in this study deep learning methods are utilized in detection and classification of minerals.

Thus, this study addresses the following research questions with consideration of the issues cited in the statement of problems:

1. Which hyperparameter best detection of rock minerals?
2. What are the significant features that can be used to identify rock minerals?
3. Which deep learning algorithm best perform to develop detection and classification Model of Indigenous Rock minerals?
4. To what extent does the developed model perform as compared to the state-of-the-art (SoTA) model?

1.4 Objectives of the study

1.4.1 General objective

The general objective of this study is to develop intelligent detection and classification model for indigenous GUJIS' Rock Minerals Using Deep Learning Techniques.

1.4.2 Specific objectives

To achieve the specified general objective, the study has the following specific objectives:

- To conduct a literature review of previous studies to understand the domain areas.
- To identify and collect the required datasets for the rock minerals using smart phone and mindat.
- Apply image preprocessing to smooth the collected images of indigenous rock minerals for further process.
- To select best performing deep learning algorithm to develop a model for Identification of Indigenous Rock minerals.
- To design a model used to detect the rock mineral images.
- To train the model by using the collected dataset.
- To test, evaluate and analyze the performance of the developed model with the **S**tate-of-**T**he-**A**rt (**S**o**TA**) model.
- To make conclusions and forward further research to improve the performance of the system.

1.5 Scope and Limitation of the study

1.5.1 Scope of the study

As detection and classification of rock mineral is a vital area of study to deal with, it is important to make some sort of boundary of task coverage for better outcomes. The study is conducted on the data obtained from Goji zone. To conduct this study, a total of 2000 images were collected. After collecting the required images, we applied image pre-processing such as image resizing, annotation and image normalization on the image datasets. The main objective of this study is to develop intelligent rock mineral detector and classifier model. To do so, a deep learning-based approach with training from scratch (i.e., setting different parameters and hyper-parameter from the beginning of building the model), and transfer learning (i.e., reusing previously pretrained weight). Hence, this method automatically extracts the representative features from the images and identifies the features of image elements that are given to a YOLOv7 detector. Finally, the model was evaluated with different metrics.

1.5.2 Limitations of the study

The following are the major limitation of this study.

- The developed model works only on 2d images
- Streak test are not included to reinforce the image dataset
- Field based mineral detection with complex environments are not included
- Rocks commonly consists of an aggregate of more than two different minerals. However, aggregate minerals are not included in this study.
- Techniques of identifying minerals such as: specific gravity, tenacity, acid test, magnetism and fluorescence will not be covered in this study.
- Graphical user interface is not developed.

1.6 Significance of the study

- **Mineralogists /Geologist Use:** The system can be used by Mineralogists, geologists and other experts for fast and accurate identification of rock minerals using smart phone. In traditional approaches, experts must equip themselves with several equipment's during their field work.
- **Small Scale Artisanal Miners Community**: the model will be usable by small scale artisanal miner's community to identify minerals originality using smart phone.
- **Field Use:** Minerals can be distinguished by a variety of characteristics. Color, streak, luster, hardness, crystal form, cleavage, specific gravity, and habit are the most frequent characteristics used to identify minerals. Even when a geologist is out in the field, most of these can be easily analyzed using this system. Identifying minerals on a field is a tedious activity and requires a lot of information and conformation.
- **Geological History:** Mineral identification is required to identify rocks and to understand the landscape as well as the geologic history of the area.
- **Originality:** the system can be used to test the originality of minerals of which identified minerals are from the original species or not

1.7 Methodology

The following methods and procedures are used in the advancement of the study to achieve the objective of this research work.

1.7.1 Study design

Design Science Research (DSR) is a paradigm for problem-solving whose goal is at improving human knowledge through creating original artefacts. Simply put, DSR aims to advance science and technology knowledge bases by developing novel products that address issues and enhance the environments in which they are used. The results of DSR form both the newly produced artefacts and design knowledge (DK), which offers a deeper comprehension of how and why the artefacts improve (or disrupt) the pertinent application contexts through design theories. Therefore, since the nature of this study encourages using Design Science Research (DSR), we used DSR as a design in this study.

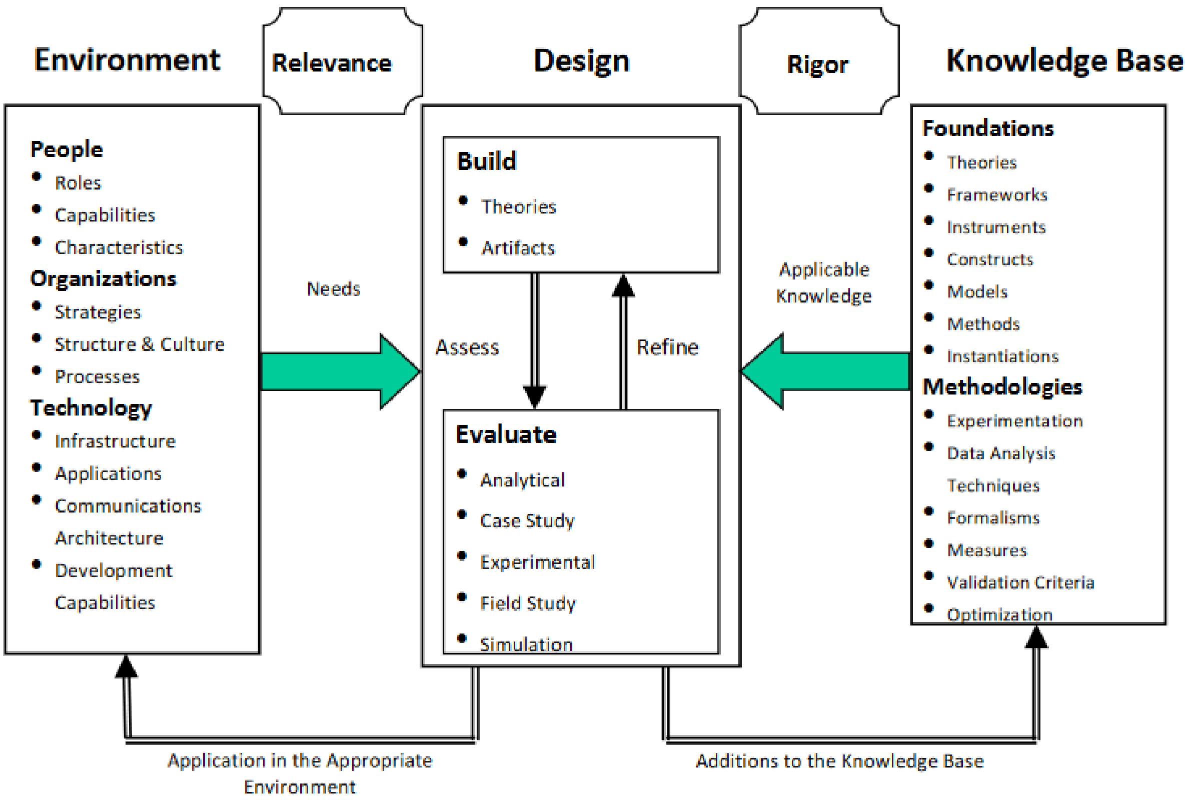


Figure 1 : Design Science Research Framework (Adapted from **[14]**).

1.7.2 Data Collection

The collection of representative data for identification of mineral using deep learning techniques requires large datasets with diverse features. Hence, to get adequate data to conduct the study mineral image dataset is collected from different mining sites in different areas of the GUJI Zone using improved smartphones camera enhanced with its software capability called professional camera. Moreover, datasets that are found from the mindat.org website are also used. Besides, a sample of some mineral images, which have failings due to improper mining, are included.

1.7.3 Literature Review

To gain a deeper understanding of the topic under investigation, a number of conceptual and empirical literature on the topic of detecting rock minerals using deep learning are investigated. Several researches on image processing, and other approaches have been studied from a variety of sources, including articles, books, journals, and the Internet.

1.7.4 Evaluation Techniques

In this study, the most commonly used evaluation performance metrics used in deep learning object detection and classification model are utilized in detection and classification of of minerals. This evaluation metrics used in this study are precision, recall, mean Average Precision (mAP) and Intersection over Union (IoU).

1.8 Implementation Tools and Libraries

1.8.1 Software tools

To develop the proposed model, python 3.9.2 programming language and google colab notebook for are used. Python is a powerful programming language that has efficient high-level data structures and a simple. Python is an effective programming language that uses an Object-oriented approach. NumPy is used for computational processing of the multidimensional and single dimensional array elements. Matplotlib is also used for better representation of the data through graphs. CV2 have also been used for computer vision detection and cuda software.

1.8.2 Hardware tools

We have used tools, instruments, and materials for data collection such as Smartphones camera and laptop computer for programming.

1.9 Organization of the Thesis

This research is organized into six chapters. The first chapter mainly dealt with background of the study, motivation, and statement of the problem objective of the study, significance of the study, scope, and limitation, and research methodology of the study.

Chapter two deals with the literature review and related works. The literature review section introduces minerals in Ethiopia, computer vision; digital image processing, machine learning, and deep learning object detection concepts in general. Under the related work section, studies previously done by different authors which is relevant to this study and more related to mineral detection and classification using image processing, machine learning algorithms and deep learning algorithms are included, and finally an executive summary of the related work.

Chapter three discusses the chosen methodology for conducting this study and the process model of the methodology.

Chapter four discuss about the proposed framework development, and the process flow in detail of this study. It reveals the technique and the algorithms that the researcher used to accomplish this study.

Chapter five delivers the research findings, analysis, and discussion obtained through the proposed and pre-trained model.

The final chapter, **chapter six**, deals with conclusions, the contribution of the study, and recommendations for future work.

CHAPTER TWO

2. LITERATURE REVIEW

2.1 Introduction

This chapter is concerned with the review of conceptual and empirical literature. Overview of minerals is presented in section 2.2 and the rock minerals that found in Ethiopia are covered in sections 2.3. Then section 2.4 gives details about what computer vision is all about. Section 2.5 discusses digital image processing followed by the basic digital image processing steps in section 2.7. In line with image processing, section 2.8 discusses machine learning. Then Section 2.9 lists out some deep learning approaches. In addition, the different hyper-parameters, the architecture of YoloV7, evaluation techniques, related works are discussed.

2.2 Overview of Minerals

A mineral is a solid formation that occurs naturally in the earth while a rock is a solid combination of more than one mineral formation which also occurring naturally. A mineral's specific crystalline structure and shape, in addition to its distinct chemical makeup, serve to characterize it. On the other hand, since a rock can contain a variety of minerals, it is categorized in accordance with how it was formed. However, some rocks may include just one mineral formation. A naturally occurring inorganic element or compound with a recognizable chemical composition, crystal structure, and physical properties is known as a mineral. Common minerals include quartz, feldspar, mica, amphibole, olivine, and calcite. The commercial value of minerals is immense and rocks are mined to extract these minerals. These rocks are known as ores, and the material left behind after the mineral has been removed is known as tailing. The classification of rocks is also influenced by the texture, mineralogy, chemical makeup, and formation method. Rocks are often classified as igneous, sedimentary and metamorphic. A rock cycle defines how one rock form changes to another [15].

Minerals either alone or in a combination called compounds forms the Earth. Mineral can be of a single element or can be compound [16]. It is an inorganic substance that occurs naturally having specified chemical composition and ordered atomic structure.

2.3 Minerals in Ethiopia

Ethiopia is undoubtedly a place of riches, with a vast and diverse landmass that is rich in resources. Geological studies on Ethiopia's mineral resources have revealed formations that are thought to hold enormous amounts of mineral resources, including proven reserves of metals, coal, precious metals, and industrial minerals. Ethiopia's exportable mineral commodities were dominated by Gold, which brought in around $300 to $500 million annually. This amount is low due to tensions with neighboring nations and violent opposition organizations. Recently, massive gold resources thought to have been mined by the fabled Queen Sheba and the ancient Egyptians were discovered in Assosa, in the western region of Ethiopia. Numerous untapped gold reserves, like the Assosa gold deposit, have the potential to push Ethiopia into the top 5 gold-producing countries in the world. Opal controls the gemstone market with over 98% of the entire value of its exports due to the absence of a well-developed sector. Aquamarine, tourmaline, amethyst, emerald, garnet, peridot, sapphire, and opal are a few other gemstones with known deposits. Ethiopia has a history with oil and gas that dates back a little more than a century. Although an oil seep was found in 1860, it was not until the 1920s that the existence of endless oil seeps was confirmed [17].

While Ethiopia's royalty rates are largely comparable to those of other African nations, its equity and tax rates are exceptionally low, with only a 5% equity requirement. This makes Ethiopia a very desirable place to engage in minerals. Today, low-skilled, small-scale miners in Ethiopia are largely in charge of mining, using low-tech equipment and inputs. [17]

2.3.1 Gold

Gold is king when it comes to exportable mineral commodities in Ethiopia, bringing in about $300 to $500 million each year. Gold is mined for generations in what is now Ethiopia, and the Assosa gold mine is thought to be the world's oldest, dating back more than 6,000 years. Currently, Lega Dembi and Sakaro, which are in the Southern People's and Nations Regional State and owned by Saudi-Ethiopian billionaire Mohammed Al Amoudi through his business Midroc, are the two largest operational gold mines in Ethiopia [17]. Gold have been the main foreign exchange means of the country for the last 30years. But, recently due to several factors' gold mining contribution to the country's economy or foreign exchange gain is below average. Ethiopia has a huge potential of various mineral resources, but not yet well explored and exploited, and hence its contribution to the overall economy or GDP is low according to recent data.

2.3.2 Platinum

Ethiopia has good prospects for platinum according to the few research that has been done so far. Platinum has not yet been properly explored or exploited. The inhabitants chose to focus on gold prospecting rather than the platinum that was found in western Ethiopia in a region named Yubdo. This is where the ultramafic strati form deposits, which are the source of platinum, are distributed over an area of about 150 km. In Sirba Abay, in the most northern region of western Ethiopia, platinum reserves have also been discovered. The sole platinum mining activity in Ethiopia for a while was run by an English company called Yubdo Gold and Platinum. They generated a little over 2.5 to 2.8 kg of platinium each year from their location in Oromia Regional State [17].

2.3.3 Opal

Ethiopian opal was first discovered in 1994, and it has since become significant on the global market. Despite the Ethiopian opal's high grade, Australian opal is said to be more expensive. The term "Welo/Wollo/Wello or Ethiopian Opals" refers to the opals that are now mined in Ethiopia and include precious, fire, and black opals.

The best and most plentiful opals are found in Ethiopia's Wollo region, in the country's north, and are now referred to as "Mezezo or Shewa Opals." Opals have also been discovered in the Shewa region. Although Ethiopian opals come in a wide variety, the stable, transparent ones with a stunning internal color play are the most common. Opal production in Ethiopia is still in its infancy due to the absence of significant mining firms or jewelry brands, but with greater investment in this material, Ethiopia is certain to overtake Australia's 100-year dominance and become a key participant [17].

2.3.4 Emerald

Ethiopia's emerald mining industry is poised to expand, but more funding from the public and commercial sectors is still necessary. The emeralds are exceptionally beautiful and are typically discovered in a region close to the gold town of Shakiso in southern Ethiopia. Low tech and low skill mining is still being done in the area, but it has potential for technology introduction given how well-known emeralds are for their quality around the world. Currently, there are bright green 10-carat Ethiopian emeralds on the market. The majorities of the mined emeralds are of the lighter green hue and found below 5-carats. According to current standards, discovery and production, it is highly unlikely that Ethiopian emeralds will ever de-throne Colombian dominance in the market, but low prices and high quality will ensure that it will be a stone to contend with. The emeralds are natural and do not require the addition of oil to improve its clarity, more prospecting and better technology could bring forth more discoveries around Ethiopia. [17]

2.4 Computer Vision

A computer vision system interprets images from an electronic camera in a manner similar to how the brain processes images from the eye in a human. For electronic engineers, computer scientists, and many others, computer vision is a rich and interesting subject to study and conduct research in. [18].

Image pattern recognition is broadly applied in image data analysis, especially in earth sciences. Geological research fundamentally involves the classification and identification of minerals as well as the analysis of microscopic images of rock. Finding out the type and characteristics of a rock-mineral sample in the lab's initial analysis is another important stage. Additionally, it serves as the foundation for geochemical analysis, including major, minor, and form element tests. Engineers and academics have up till now carried out manual rock-mineral microscopic picture identification, which is dependent on the knowledge and expertise of the operators. It is time-consuming and ineffective. Additionally, the knowledge of the conductor plays a significant role on the outcome. The succeeding work is harmed by incorrect rock-mineral recognition, which can result in resource waste and financial loss. It is critical to create rock-mineral microscopic image recognition methods that are efficient, reliable, and objective. [19].

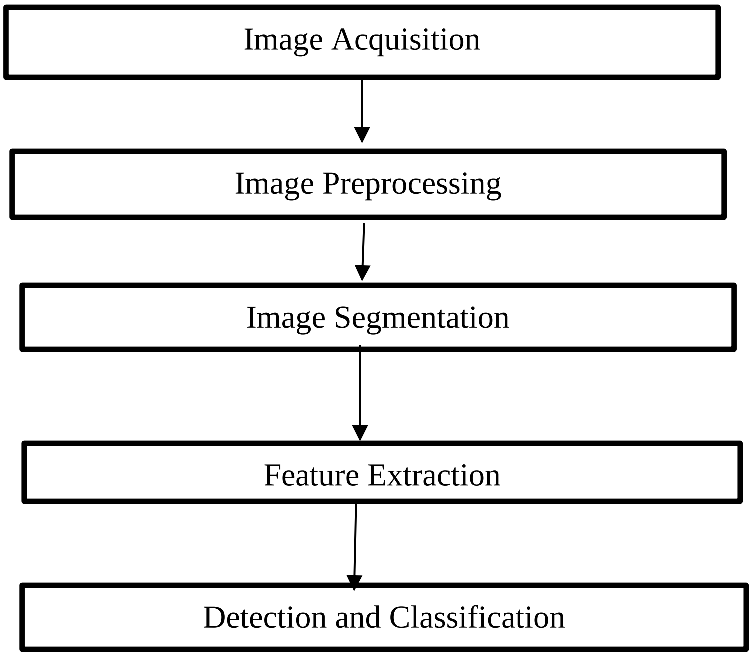
2.5 Digital image processing

For purposes of mathematical analysis, an image can be represented by the two-dimensional function f (x, y), where x and y are spatial (plane) coordinates. The intensity or grey level of the image at any given place is determined by the amplitude of f at that location. [20]. We refer to the image as a digital image when the values of f's amplitude, x, and y are all discrete, finite quantities. A digital image is a matrix representation of a two-dimensional image made up of a finite number of point cell elements, commonly known as pixels (picture elements), where each pixel is represented by a number [21]. In imaging science, image processing is the mathematical manipulation of images using any type of signal processing where the input is an image, a sequence of images, or a video, such as a picture or a frame from a video [22].

Analogue and digital image processing are the two categories into which image processing is divided. Analogue image processing can be used to process hard copies like prints and pictures. Image analysts use a range of interpretational fundamentals while using these visual tools. [23].

The term "digital image processing" describes the use of a digital computer to process digital images. You should be aware that a digital image is made up of a finite number of elements, each of which has a unique position and value. These components are referred to as pixels, picture elements, and image elements [20]. Three levels can be identified in image processing. Low-level, mid-level, and high-level processing are these. The next subsections go over the Basic Digital Image Processing procedures. Different image processing and learning methodologies can be used for the full construction of a digital image processing application. The basic procedures in the detection of rock minerals are shown in Figure 2.

Figure 2: The fundamental steps involved in the detection of the rock minerals. [20]



In digital image processing, collecting the necessary images is the first step, which is called image acquisition. Then the next step preprocesses the collected images so this phase is smoothing the collected images to get high-quality images through various methods like resizing, color conversion, noise removing, and so on. Then image segmentation comes, in this phase, the images divide into several regions or objects. The segmentation is used to increase the chance of representation of an image. Following this, the next step is extracting the relevant information from an image, different types of image features are used to represent an image for object recognition and/or identification like color, shape, and texture of an image. Finally, the last step is to predict the class labels for given images.

2.5.1 Image acquisition

Image acquisition is the first step in any image processing task, in which an original input image is acquired from the initial source. Image can be acquired differently; either capturing an image from the actual environment or use already existed image files from an electronic source [22]. These images are often unprocessed.

2.5.2 Image preprocessing

Pre-processing of an image involves the removal of distortion, which has a major effect on accuracy for identification and classification. The major task of this phase is to improve the image quality by removing unnecessary distortions or enhancing the image quality for future processing. The techniques have various methods like image resizing, image enhancement, image restoration, removing noise, and soon [24].

2.5.3 Image segmentation

After preprocessing an image, the next step is image segmentation. In this phase, an image divides into several objects or regions. The segmented image features are robust for the identification and classification of minerals. Segmentation is used to improve the chance of representation of an image, by analyzing the image data and extract useful information for further processing [25]. There are two methods for segmenting an image based on similarities and discontinuities. Similarities include the partitioning of the photos based on a set of predetermined criteria. However, in discontinuities, the images are divided according to the abrupt shifts in the intensity of values [26]. The popular techniques used for image segmentation are shown in Figure 5, the Threshold method, edge detection-based techniques, region-based techniques, clustering-based techniques, watershed-based techniques, partial differential equation-based, and artificial neural network-based techniques, etc.

There are several image segmentation techniques. Some of them are listed below:

- Region Based Method
- Edge based method
- Threshold Method
- Watershed Based Method

- Clustering Based Method
- PDE Based Method
- ANN Based Method

Some of these algorisms are briefly discussed below.

A. Region-based method

In this technique, pixels that are related to an object are assembled for segmentation. This technique is also termed similarity-based segmentation, dividing an image into segments that are comparable based on a set of parameters. In region-based segmentation, there won't be any gaps caused by missing edge pixels. [27].

B. Edge-based method

This segmentation locates the edges, if either the first derivative of intensity is greater than a particular threshold or the second derivative has zero crossings. In edge-based segmentation methods, first the edges are detected and then are connected to form the object boundaries to segment the required regions [28]. Edges are detected to identify the discontinuities in the image. Edges on the region are traced by identifying the pixel value and it is compared with the neighboring pixels.

C. Threshold-based method

These methods divide the image pixels concerning their intensity level. These methods are used over images having illumination objects than the background. The selection of these methods can be manual or automatic (i.e., can be based on prior knowledge or information of image features). The method split an image into smaller segments, by using one color or grayscale value to define their boundary in image processing. Besides, this segmentation method is generating only two classes, based on this, this technique is not applicable for complex images.

D. Watershed based method

The watershed segmentation methods are based on the topographic relief representation of an image, where the value of each image element denotes the height of the element at this location. The term element, which combines the phrases "pixel" and "voxel," is used to describe how these algorithms interpret both 2D and 3D pictures [29]. For a better outcome, watershed segmentation is often applied to the result of the distance transform of the image rather than to the original one. Accordingly, there are two main approaches of watersheds: by flooding and by rain falling [27]. If we can recognize, or "mark," foreground items and background places, segmentation using the watershed transforms works effectively. This fundamental process is used for marker-controlled watershed segmentation [30]:

1. Create a segmentation function. You are attempting to segment the objects in this image by its dark parts.

2. Calculate foreground markers. These are connected blobs of pixels within each of the objects.

3. Compute background markers. These are pixels that are not part of any object.

4. Change the segmentation function so that it only has least at the foreground and background marker locations.

5. Compute the watershed transform of the improved segmentation function.

2.5.4 Feature extraction

After segmenting the images, the next phase is extracting the necessary features. Feature extraction is a special form of dimensionality reduction. It extracts the most relevant information from the original data and represent that information in a lower dimensionality space. It is concerned about the extraction of various attributes of an object and thus associates that object with a feature vector that characterizes it [25] [31].

Feature extraction is important because the specific features made available for discrimination directly influence the efficiency of the classification task. Therefore, features can be classified into two categories: local features which are typically geometric (such as number of endpoints, joint, branches, etc), and global features which are typically for topological or statistical (e.g., number of holes, connectivity) [32]. Figure 3 illustrates different types of image features are used to represent an image for object recognition and/or identification.

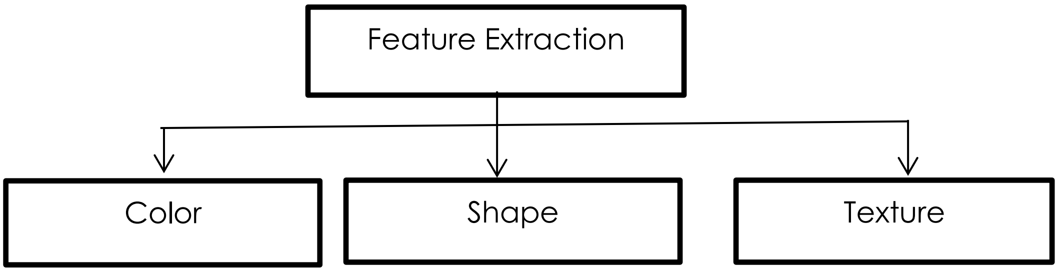


Figure 3: Classification feature extraction method

A. Color features

Color is one of the important features of an image. Several color spaces have been used such as RGB, LUV, HSI, YCbCr, HSV, and HMMD. Once the color space is identified, the color feature can be extracted from the images or regions [33]. Color features are represented using Color Histogram, Color Correlogram, Dominant Color Descriptor, and Color Co-occurrence Matrix.

B. Texture features

The texture is an important element of human visual perception and is used in many computer vision systems. There is no precise definition of texture has been adopted yet, in addition, different authors define it as a measure of roughness, contrast, directionality, line-likeness, regularity, and roughness. Texture describes how the patterns of color are scattered in the images. This feature also describes the physical composition of the surface [32]. Texture features are extracted using various methods. Grey Level Co-occurrence Matrix (GLCM), Gabor filter, Histogram of Oriented Gradients, and Haar Wavelet Decomposition are examples of popular methods to extract texture features.

C. Gabor filter

Gabor filter is one of the well-established texture descriptors introduced by Gabor in 1946 [34]. It extracts texture features by analyzing the frequency domain of the image. It is a Gaussian function modulated by complex sinusoidal frequency and orientation. It can perform both in the spatial and frequency domain and can be in any number of dimensions [32]. This filtering technique is broadly used to extract features for image classification because of the nature of the spatial locality, orientation selectivity, and frequency characteristic. Gabor filter or wavelets characterize an image by obtaining the center frequency and orientation parameter. The frequency and orientation representation, are useful for texture representation and discrimination and the same concept is used in the human visual system. The Gabor features are resilient to illumination, rotation, scale, and translation [35].

D. Shape features

The shape is one of the important features in feature extraction. When the image has been divided into distinct sections or objects, they are typically characterized. The two types of shape description are boundary-based and region-based. An effective feature for an object's shape representation should be invariant under translation, rotation, and scale [36]. The most well-known shape techniques are the Shape moment invariant and binary image algorithm.

2.5.5. Classification

Classification of an image consists of a database that holds re-defined patterns that are compared with detected objects to classify them in the right category. Classification is executed based on spectral-defined features such as density, texture, etc. Classification has a two-step process training and testing. In the training steps, a classification algorithm builds the classifier by analyzing and learning from a training dataset and its associated class labels. In the testing step, the model is used to predict class labels for given data. Machine learning and deep learning techniques are often used for classification.

2.6 Machine Learning

The type of data representation (or features) that machine learning techniques are used on has a significant impact on how well they function. As a result, building preprocessing pipelines and data transformations that give data a representation that can allow effective machine learning accounts for a sizable percentage of the actual work needed in implementing machine learning algorithms. Such feature engineering is important but labor-intensive and highlights the weakness of current learning algorithms: their inability to extract and organize the discriminative information from the data. It is possible to overcome this limitation by using feature engineering [37].

It would be highly desirable to make learning algorithms less reliant on feature engineering in order to speed up the development of new applications and, more importantly, to advance the field of artificial intelligence (AI). This would increase the range and simplicity of applications of machine learning. An AI must fundamentally understand the world around us, and we argue that this can only be achieved if it can learn to identify and disentangle the underlying explanatory factors hidden in the observed milieu of low-level sensory data. [38]

2.7 Deep Learning

In contrast to machine learning algorithms, which typically require human intervention to extract features, deep learning algorithms extract feature sets automatically and continuously improve their performance. While a deep learning algorithm needs large data sets, potentially including heterogeneous and unstructured material, a machine learning algorithm may learn from relatively modest size of data.

Deep learning can be thought of as an advancement of machine learning. Deep learning is a machine learning technique that layer's algorithms and computational units or neurons into it. The structure of the human brain serves as a model for these deep neural networks. Similar to how our brains process information, data moves through this network of interconnected algorithms in a non-linear way. Deep learning has perceptron, which a linear classifier has output single layer.

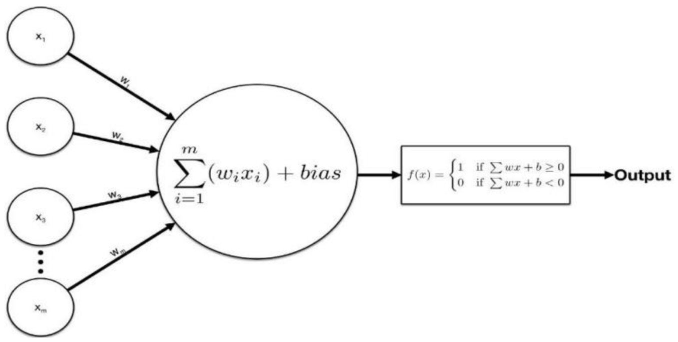


Figure 4: perceptron

Deep learning is a branch of machine learning that makes use of hierarchical architectures to learn high-level abstractions from data. Deep learning adopts a novel approach to learning representations from data that places an emphasis on learning progressively deeper and more significant representations. The deep in deep learning refers to the concept of multiple layers of representations rather than any form of deeper understanding that can be attained by the method [37]. [39]

Deep learning is significantly advancing the field of artificial intelligence by tackling issues that have long defied the greatest efforts of the field. It has proven to be very effective at identifying intricate structures in high-dimensional data and is thus applicable to many fields of science, business, and government. [40]

Tens, or even hundreds, of successive layers of representations are frequently used in modern deep learning, and each of these layers has automatically learned via exposure to training data. Other methods of machine learning, on the other hand, frequently concentrate on learning just one or two layers of data representations. Neural networks are models that are used to learn these layered representations (nearly always); the term refers to neurobiology. However, certain deep learning principles were created in part by taking inspiration from our knowledge of the brain. There is no evidence that the brain uses learning techniques similar to those utilized in contemporary deep learning models, hence deep learning models are not models of the brain [41].

The most popular deep learning methods that have been widely used are Convolutional Neural Networks (CNNs), Recurrent Neural Network (RNN), Restricted Boltzmann Machines (RBMs), Auto encoder, and Sparse Coding. [42].

2.7.1 Convolutional Neural Network

Convolutional neural network (CNN) is one of the main learning mechanisms to do images recognition, images classifications, objects detections etc. It is based on learning levels of representations. The higher lever concepts are defined from lower-lever ones and the same lower lever concepts can help to define many higher lever concepts. It learns multiple levels of representation and abstraction which helps to understand dataset such as images, audio and text. It is advantageous of simple structure, less training parameters because of shared weights and adaptability [43].Typically, a convolutional neural network has three basic structural components:Convolutional, pooling, and fully connected layers are among them.

The term "convolution" refers to a specific type of linear operation that we substitute a convolutional layer for at least one of the neural network's layers for the standard metric multiplication.

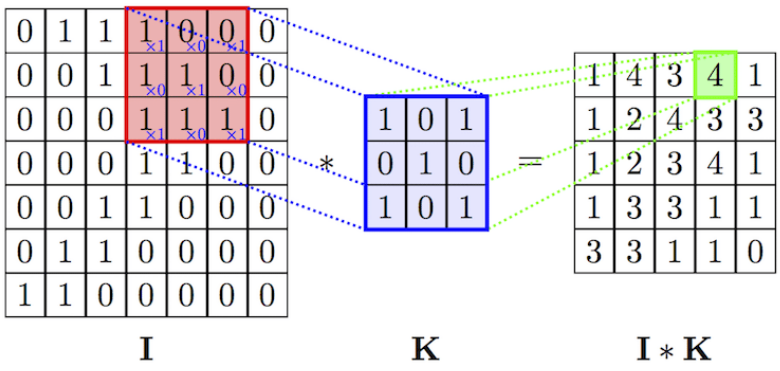


Figure 5: Illustration of the convolution operation between an image and Kernel

Simply explained, the pooling layer reduces the dimensions by replacing the result of the convolution layer with a stats summary of the surrounding outputs.Finally, a fully connected layer is placed at the end of the network to give us the final output.

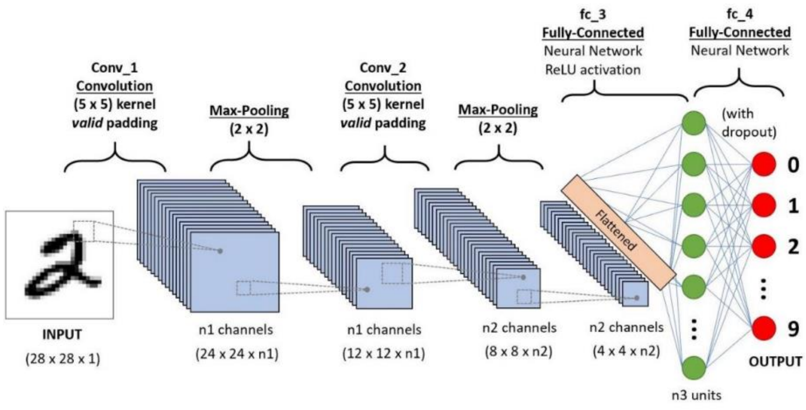


Figure 6: convolutional neural network

2.7.2 Recurrent Neural Network (RNN)

Recurrent neural networks (RNNs) are a type of artificial neural network. This structure allows the information to be circulated in the network. RNN is the ability to recognize and predict sequences of data such as text, genomes, handwriting, spoken word, or numerical time series data. They have loops that allow a constant flow of information and can work on sequences of random lengths. The RNNs attempts to address the necessity of understanding data in sequences. RNN can be used for time series prediction because it can remember previous inputs also, which is called Long-Short Term Memory (LSTM).

2.7.3 Restricted Boltzmann Machines (RBMs)

A Restricted Boltzmann Machine (RBM) is a generative stochastic neural network that comes from the Boltzmann machine and was proposed in 1986. RBM is different from the Boltzmann Machine, with the restriction that the visible units and hidden units must form a divided graph. Even though RBM has inherited the two-layer neuron structure of the Boltzmann machine, there is no connection between neurons in the same layer with only the whole connection between the visual layer and the hidden layer. Restricted Boltzmann Machines (RBM) is used to train layers of the Deep Belief Networks (DBNs) [44].

2.7.4 Auto-encoder

The deep autoencoder is a special form of the DNN (with no class labels), whose output vectors have the same dimensionality as the input vectors. It is mainly used to process complex high dimensional data, and it is often used for learning a representation or effective encoding of the original data, in the form of input vectors, at hidden layers. Note that the auto encoder is a nonlinear feature extraction method without using class labels. As such, the features extracted aim at conserving and better-representing information instead of performing classification tasks, although sometimes these two goals are correlated [45].

2.8 Object detection

Object detection has been attracting increasing amounts due to its numerous applications and current technological advancements; it has received a lot of attention lately. Research on this task is ongoing in academia and practical contexts, including security monitoring, autonomous driving, traffic surveillance, drone scene analysis, and robotic vision. Deep convolution neural networks and GPU computing capacity have made significant contributions to the rapid advancement of object detection systems, among other variables and initiatives. Currently, the deep learning approach is widely used in both generic and domain-specific object detection in the field of computer vision. In order to extract characteristics from input photos (or videos), classify objects, and determine their location, most modern object detectors use deep learning networks as their detection network and backbone [46].

Generally, there are two kinds of object detection. [46] The first one two-stage object detectors that use a preprocessing step to create an object region proposal in the first stage and then object classification for each region proposal and bounding box regression in the second stage, and the second one is single-stage detectors that perform detection without generating region proposals in advance. Single-stage detectors are highly scalable, provide faster inference, and thus are better for real-world applications than two-stage detectors such as region-based Convolutional Neural Networks (R-CNNs), the Faster-RCNN, and the Mask-RCNN [47], [48] . The one-stage detection networks are You Only Look Once (YOLO) [49] and Single Shot MultiBox Detector (SSM) [50]

2.8.1 R-CNN

To avoid the issue of having to choose so many different regions, [51] proposed a method where 2000 regions were extracted from the image using selective search, and these regions were referred to as region proposals. As a result, it can work with 2000 regions rather than attempting to classify a large number of them. The selective search algorithm mentioned below was used to generate these 2000 region ideas.

Selective Search:

1. Generate initial sub-segmentation, we generate many candidate regions
2. Use greedy algorithm to recursively combine similar regions into larger ones
3. Use the generated regions to produce the final candidate region proposals

In order to build a 4096-dimensional feature vector, the 2000 candidate region suggestions are twisted into a square and input into a convolutional neural network. The CNN performs the function of a feature extractor, and the output dense layer is made up of the features that were extracted from the picture and input into a support vector machine to classify the presence of an object within a candidate region proposal. The system additionally forecasts four offset values to improve the bounding box's accuracy in addition to whether an object will be present within the region proposals. For instance, if a region suggestion was given, the algorithm would have anticipated the presence of a person, but that person's face could have been sliced in half within that region proposal. As a result, the offset values aid in modifying the region proposal's bounding box [51]. The R-CNN uses search algorithm.

2.8.2 Fast R-CNN

The author [51] of the previous paper(R-CNN) was developed by addressing a few problems with the R-CNN approach from earlier research in order to provide a quicker object detecting method. The process is comparable to the R-CNN algorithm. CNN is fed the input picture to construct a convolutional feature map, not the region suggestions, though. The region of proposals is located using a RoI pooling layer and a convolutional feature map, and then it is distorted into squares and molded into a fixed size before being fed into a fully connected layer. It makes use of a softmax layer to predict both the class of the suggested region and the offset values for the bounding box from the RoI feature vector.

The Fast R-CNN is quicker than R-CNN since it performs the convolution process only once per image and creates a feature map from it. While the Fast-CNN algorithm does not employ the region proposal, the R-CNN algorithm does, which slows down algorithm performance. As a result, the algorithm Fast-CNN performs better than R-CNN. However, it uses a search algorithm, which makes the technique time-consuming and slow.

2.8.3 Faster R-CNN

The authors [48] developed an object detection algorithm that does away with the selective search strategy and enables the network to pick up on region proposals.

The image is fed into a convolutional network, much to Fast R-CNN, to produce a convolutional feature map. A different network is utilized to forecast the region proposals rather than utilizing a selective search algorithm on the feature map to determine the region proposals. In order to classify the picture within the proposed region and forecast the offset values for the bounding boxes, the anticipated region proposals are then reshaped using a RoI pooling layer. Faster R-CNN's detection accuracy is good, but its detection speed is slow, making it unable to satisfy real-time demands.

2.8.4 YOLO

YOLO (you only look ones) is an object detection algorithm that has gained popularity in computer vision. YOLO is a real-time object detection approach that trains a neural network in a single forward pass to process an image. Unlike traditional object detection algorithms involving multiple processing stages, YOLO performs object recognition and bounding box regression in a single step [52]. This makes it fast and efficient, with the ability to process up to 60 frames per sec. YOLO works by dividing an image into a grid of cells and predicting bounding boxes for each cell. For each bounding box, YOLO predicts the class probability. The probability that the bounding box contains a particular object and the confidence score which is the probability that the bounding box contains an object. YOLO also predicts the bounding box coordinates relative to cell [53].

Anchor boxes are preconfigured boxes with various sizes and aspect ratios that are used by YOLO to increase prediction accuracy. To determine the size and shape of the object inside a cell, each anchor box is connected to a certain cell. Anchor boxes allow YOLO to manage objects of various sizes and shapes. One of the main strengths of YOLO is its speed. YOLO can process images in real-time, which makes it suitable for applications such as autonomous vehicles, surveillance systems, and robotics [53]. YOLO is also effective since it only needs to process an image once, as opposed to conventional object detection algorithms that call for many network passes. Multiple item detection in an image is YOLO's strongest suit. Because YOLO predicts bounding boxes for each cell, it can detect multiple objects in different image parts. This makes YOLO ideal for pedestrian detection and traffic sign recognition [53].

YOLO is very quick since it only employs one feed-forward convolutional network and no longer needs a second per-region classification operation to forecast item classes and placements [49].

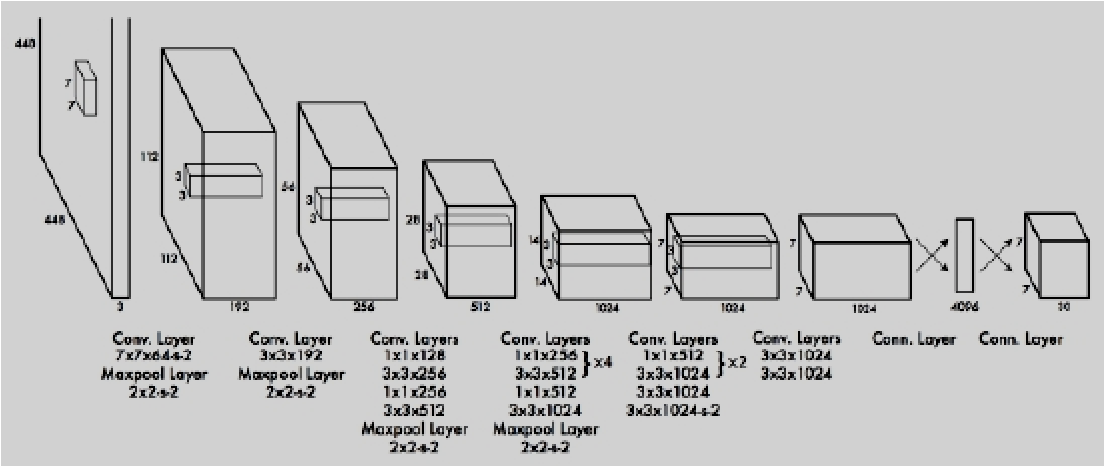


Figure 7: YOLO architecture consists of 24 convolutional layers and two fully connected layers[47].

YOLO uses trainable bag-of-freebies, which is divided into model re-parameterization and model scaling.

A. Model re-parameterization

Model re-parameterization can be divided into two categories: model-level ensemble and module-level ensemble. Both of these categories can be thought of as ensemble techniques. For model-level re-parameterization in order to produce the final inference model, there are two typical methods. The first is to train numerous, identical models using various training datasets, and then average the weights of the multiple trained models. The second method involves calculating a weighted average of model weights across various iteration counts. The more common technique is module level re-parameterization. With this approach, a module is divided during training into numerous branches that are either identical or distinct from one another, and during inference, the multiple branched modules are combined into a single, fully equivalent module. However, not every re-parameterized module that has been proposed can be used perfectly with every architecture [54].

B. Model scaling

The model scaling method typically makes use of a variety of scaling factors, including resolution (size of the input image), depth (number of layers), width (number of channels), and stage (number of feature pyramids), in order to achieve a good trade-off for the quantity of network parameters, computation, inference speed, and accuracy. One of the often employed model scaling techniques is network architecture search (NAS). Without establishing overly complex criteria, NAS may automatically look up scaling factors from the search space. The cost-prohibitive computing required by NAS to finish the search for model scaling factors is a drawback [54].

YOLOv7

YOLOv7 is the new advanced detector in the YOLO family. This network uses trainable bag-of-freebies, allowing real-time detectors to improve precision dramatically without increasing inference costs. It incorporates extend and compound scaling, enabling the target detector to efficiently minimize the amount of parameters and calculations, leading to a noticeably accelerated detection rate [55].

YOLOv7 exceeds typical object detectors in precision and speed of 5 FPS (frames per sec) to 160 FPS additionally; it makes it simple to fine-tune detection models and provides a set of freebies that are ready for use. It is simple to add further modules and create new models thanks to the YOLOv7 configuration file [56]. The paper proposes E-ELAN, which uses expand, shuffle, and merge cardinality to achieve the capability of continuously improving the network's learning ability while maintaining the original gradient path. [57]. The backbone, neck, and head components of YOLOv7's network structure can be distinguished. The CBS module in YOLOv7's network construction process consists of convolution operation, regularization, and SiLU activation. The SiLU activation function is presented in Equation 2.1. The CBS module along with maximum pooling forms the MP module. The primary computational modules of YOLOv7 are the ELAN-1 and ELAN-2 modules, which are in charge of extracting features,

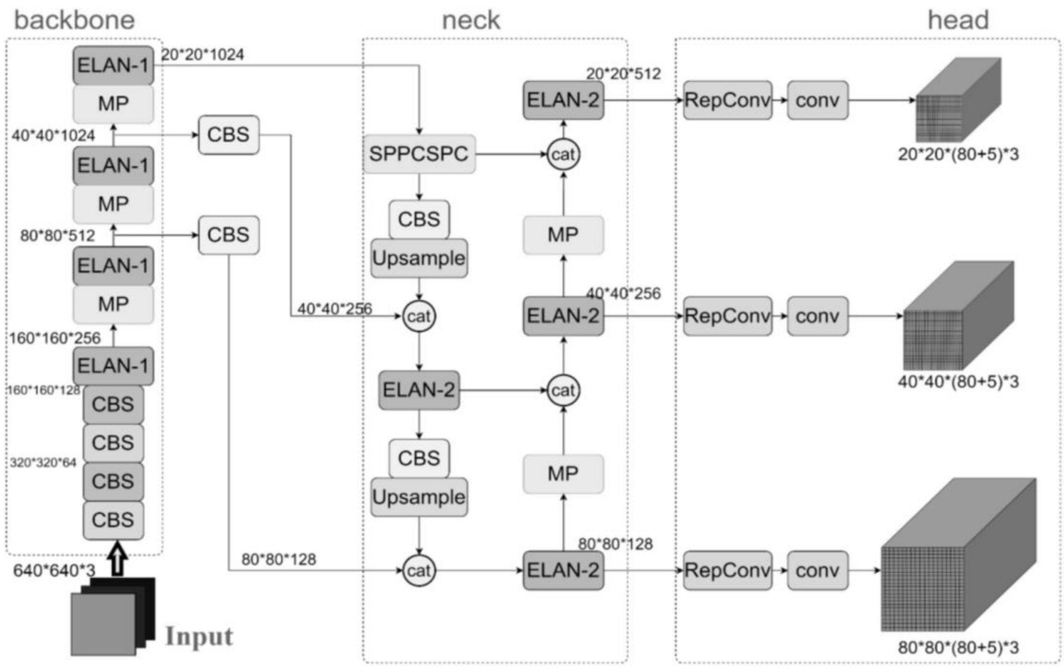


Figure 8: Original YOLOv7 network structure **[58]**

The three-channel color image will be consistently cropped to a size of 640*640 for the input section. YOLOv7 uses four ELAN-1 modules to finish the major feature extraction operation in the Backbone section, as opposed to other algorithms that directly implement the backbone network structure design that is currently in use Three ELAN-1 modules are among them and output to the Neck section after the number of channels was changed.

The Neck component of YOLOv7 uses the feature pyramid structure, and 1 SPPCSPC module and 4 ELAN-2 modules each perform the feature extraction function. Particularly, the three ELAN-2 modules on the right side are output directly to the Head section.

The network ultimately uses three detecting heads to perform feature output. Three different sizes of 20×20, 40×40, and 80×80 feature maps are output. The output feature maps become in the form of a one-dimensional vector after convolution, called the Fully Connected Layer. With the one dimensional vector output by the convolutional neural network, the targets in the image can be predicted. [58]

$$SiLU(x) = x/1+e^{-x}$$

Equation 2.1: SiLU Activation Function

The loss function of the YOLOv7 model consists of three parts: confidence loss (L_{obj}), classification loss (L_{cls}), and localization loss (L_{box}). The total loss is the sum of the three losses with different weights. The total loss function is expressed as

$$LOSS = a \times L_{obj} + b \times L_{cls} + c \times L_{box}$$

Equation 2.2: The total loss function

Where a, b, and c are the weighting factors corresponding to the three partial losses.
Binary cross-entropy loss is used for classification loss and confidence loss, and CIoU loss is used for localization loss.

$$Lp(q) = -1/N\sum^{N}_{i=1} yi \cdot \log(p(yi)) + (1 - yi) \cdot \log(1 - p(yi))$$

Equation 2.3: Cross-entropy loss

The binary cross-entropy loss can be defined as equation 2:3 Where *yi* represents the real sample label, and p(y) is the predicted probability of the point being positive for all N points.

$$CIoU = IoU - (\rho^{2\,(b,\,bgt)}/c2 + \alpha v)$$

$$v = 4/\pi 2\ (\arctan w^{gt}/h^{gt} - \arctan w/h)^2$$

Equation 2.4: Class loss

CIoU can be expressed as equation 2.4 Where $\rho 2(b, bgt)$ represents the Euclidean distance between the centre points of the prediction frame and the real frame, and c represents the diagonal distance of the smallest area that can contain the predicted and real boxes. The IoU represents the intersection area of the prediction box and the real box.

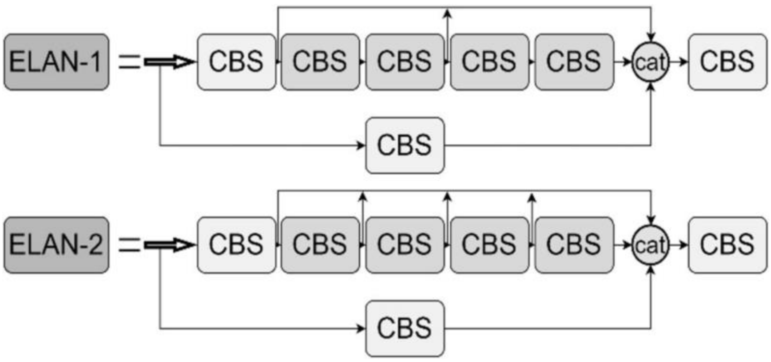


Figure 9: Original ELAN module structure.

2.9 Related Works

Various researches have been conducted and proposed for classification and identification of rock mineral using deep learning. In this section, related works in the area of intelligent identification and classification rock minerals using deep learning systems are reviewed.

In paper [7], the researchers proposed a deep learning approach for mineral identification that combines image and Mohs hardness. Detecting minerals that uses both mineral photo images and Mohs hardness are the main objective of proposed model. The researchers were collected important data and the data source was from mindat. The researchers categorized 36 minerals and they used 15,291 numbers of images out 183,688 images samples. Therefore, the researchers used deep neural network algorithm to detect a model. The developed model achieves an accuracy of 90.6% top-1 and 99.6 top-5 accuracy for 36 common minerals. The test metrics were top-1 and top-5. However, in this study the transparency, color and density were not included. In addition, the researchers used limited number of minerals and the main objective of the study was not developing classifier model rather than identifying.

In [8] the researchers proposed detection of minerals. The proposed model identifies minerals using images luminance equalization that used light intensity. A total number of 220,057 images were collected with 50 mineral species from mindat. The researchers used preprocessing techniques for the quality of data such as combined histogram equalization (HE). In addition, Laplace algorithm, which is called deep learning yolov5 model, is implemented. The developed model achieves an accuracy of 95.6%. However, in the study transparency, density, streak of minerals was not included. In addition, the researchers gave low attention for tiny rock and single rock mineral images.

Automatic gemstone classification using computer vision is proposed by [6]. A total number of 2326 images were collected from Kaggle database. The researchers used color and texture features for classifying gemstones. In addition, random forest algorithm was implemented. From the collected samples, 90% are assigned for training and the remaining 10% for testing and the gemstone was categorized in to 68 classes. The developed classifier model achieves an accuracy of 66.9% of human expert and 69.4% of computer vision. The human expert and computer vision testing techniques are implemented for comparison. However, the researchers have not investigated the high quality and low-quality gemstones separately. The number of gemstones samples used was small and machine learning algorithm was employed.

In [59], the Authors proposed deep learning on automatic mineral grain segmentation and recognition. The model classifies the minerals into five classes. The required data were collected manually. A total number of 21,091 images were collected by optical microscope. While developing the classifier model, deep learning techniques such as ResNet1 and ResNet2 algorithm were implemented. The developed classifier achieves an accuracy of 90.5%. High performance GPU is used for training and classification into only five classes.

In paper [60], the researchers developed a model for the identification of minerals types. The model identifies minerals and grain using hyper spectral imaging and deep learning for mineral processing. The required data were collected manually by the researchers. A total number of 256 RGB images of five minerals are used. The researchers have used hyperspectral imaging feature for the identifying the minerals. A deep convolutional neural network (CNN) - based architecture called VGG techniques was used for identifying minerals. The developed model achieves an accuracy of 91.9% with a low precision of 39.52% for the RGB images. However, they used limited types of minerals (only five).

An Enhanced rock mineral recognition method integrating a deep learning model and clustering algorithm is proposed by [9]. The proposed model is recognizing rock minerals by integrating the two algorithms. The necessary data were collected manually by the researchers. For their experimentations 4178 images of 12 types of elaborated rock minerals have used. The the researchers used color and textures features for recognizing the rock minerals. Towards creating the model, k-means algorithm and deep learning coupled with color model was implemented. For the experimentations top-1 and top-3, test option was implemented. Finally the proposed model achieves accuracies of 74.2% and 99.0% of top-1 and top-2 respectively. However, cluster and cleavage are not included in the study. Moreover, the study is not applicable for field study.

In paper [61], the researchers proposed a model for intelligent identification of rock types based on faster R-CNN. For their experimentations, 1034 images were collected from 8 types of rock. The necessary data were collected manually by the researchers. In this study, different algorithms such as faster R-CNN and VGG16 convolutional neural network are implemented. The color feature was used for the identification of rocks. After all, the developed model achieves accuracies of single type 96% and multi type 80%. The developed model identifies rocks in general and works in underground mines.

A deep learning model for quick and accurate rock recognition with smartphones was developed by [62]. A total number of 3795 data were collected from 30 types of rock. The data were collected from china geological survey. Convolutional neural network algorithm was implemented to develop the model. Color is used as a feature by the researchers. The developed model achieved an accuracy of 97.65%. However, the researchers used limited number of instances for training.

In [63], the researchers proposed deep learning based mineral image classification combined with visual attention mechanism. The proposed model classifies minerals based on deep learning and combined with visual attention mechanism. The necessary data have been collected manually by the researchers. A total number of 1400 images of minerals were collected. In this study, four CNN model was used namely, ResNet18_MA, ResNet34_MA, ResNet50_MA and ResNet101_MA. The confusion matrix was used for evaluation. The evaluation demonstrates that the classification model achieves accuracies of ResNet18 79.29%, ResNet34 81.95%, ResNet50 84.89% and ResNet101 86.41%. However, the researchers used limited number of minerals data types.

In paper [64], the authors proposed mineral photos recognition based on feature fusion and online hard sample mining. For the experimentation, 14,986 images of 22 common mineral were collected from the national mineral rock and fossil specimen's resource center of China mindat.org. Different low-level features namely color and shape of minerals was extracted. A deep learning ResNet-50 was used for the recognition. In addition, the multi resolution features were extracted from ResNet-50. For the experimentation, 10-fold cross-validation, top1 and mean average precision testing option were implemented. The proposed model achieved an accuracy of 88.01% with Top1 and 87.15% with mAP. However, the researchers used only the images of minerals.

In [65], the researchers proposed intelligent identification for the rock-mineral microscopic images using ensemble machine learning algorithms. The necessary data have been collected manually with microscope. A total number of 481 images were collected for their experimentation. In this study, different algorithms such as LR, SVM, and MLP are implemented. 10-fold cross validation testing option was implemented. Finally, the proposed model achieves an accuracy of 90.9%. However, the researchers used small dataset and only microscopic mineral images were used.

The summary of the related works is presented in Table 1 below.

Table 1: Summary of Related Works

Author	Title	Data Size and Data Source	Properties	Techniques	Results	Limitation
[7]	Mineral Identification Based on Deep Learning That Combines Image and Mohs Hardness	36 minerals categories 15,291 out of 183,688 images samples Mindat	uses both mineral photo images and the Mohs hardness	Deep Neural Network	90.6% top-1 accuracy and 99.6% top-5 accuracy for 36 common minerals	✓ transparency, color and density are not included ✓ limited number of minerals
[8]	Mineral Identification Based on Deep Learning Using Image Luminance Equalization	220,057 samples With 50 mineral species mindata	light intensity	(preprocessing) Combined histogram equalization (HE) and the Laplace algorithm DeepLearning Yolo V5 Model	95.6%	i. Transparency, density ,streak of minerals are not included ii. Single rock mineral image identification ii. Attention for Tiny rock is very low
[6]	Automatic Gemstone Classification Using Computer Vision	2042 training images and 284 unseen (test) images 2,326 68 categories of gemstones Data are available from Kaggle	color & texture	Machine Learning(Random Forest)	human expert provided an accuracy of :66.9% & computer visiion :accuracy of 69.4%	i. small number of samples of gemstones ii. machine learning algorithms were employed iii. high quality gemstones and low quality gemstones are not separately investigated

Author	Title	Data Size and Data Source	Properties	Techniques	Results	Limitation
[59]	Deep-Learning-Based Automatic Mineral Grain Segmentation and Recognition	final data set consisted of 21,091 images divided into five classes Manually collected	grains captured from optical microscopes	Deep learning(ResNet 1 and ResNet 2)	accuracy of 90.5%	• The scarcity of mineral data sets • Unbalanced data for different classes. • High-performance GPUs are required for training • Only five categories
[60]	Automated Identification of Mineral Types and Grain Size Using Hyper spectral Imaging and Deep Learning for Mineral Processing	RGB images of five types of minerals 256 of data Manually collected by Researchers	hyperspectral imaging	Combined hyperspectral imaging and deep learning techniques (CNN called VGG-16)were utilized	Deeplearning (accuracy was 91.9%) Low precision (accuracy was 39.52% for the RGB images)	Limited types of minerals(only five) • only visual data were Employed

Author	Title	Data Size and Data Source	Properties	Techniques	Results	Limitation
[9]	An Enhanced Rock Mineral Recognition Method Integrating a Deep Learning Model and Clustering Algorithm	4178 images of 12 kinds of elaborate rock minerals Manually collected by Researchers	Textures& Color	K-means algorithm deep learning model, coupled with color model	top-1 and top-3 accuracies of 74.2% and 99.0%	• Cleavage and luster not included • Not applicable for field study • Dataset is small
[61]	Research on Intelligent Identification of Rock Types Based on Faster R-CNN Method	8 types of rock images of 1034 rock images of rocks Manually collected by Researchers	Rock images &color	framework of Faster R-CNN VGG16 convolutional neural network	Single type: 96% Multi type:80%	• Its rock identification(very general) • It works for underground mines
[62]	A Deep Learning Model for Quick and Accurate Rock Recognition with Smartphones	30 types of rocks & 3,795 images were take Geological Survey Projects of China Geological Survey	Color & rock photos	Convolutional Neural network	97.65%	• Its rock identification(very general) • rock training samples is limited

Author	Title	Data Size and Data Source	Properties	Techniques	Results	Limitation
[63]	Deep Learning Based Mineral Image Classification Combined with Visual Attention Mechanism	14,000 three types of ore particles in China Manually collected by Researchers	Mineral images	NNs combined with visual attention mechanism	ResNet18_MA, ResNet34_MA, ResNet50_MA, and ResNet101_MA is 79.29%, 81.95%, 84.89%, and 86.41%,	• Not rock minerals • Only three types of minerals which are limited
[64]	Mineral Photos Recognition Based on Feature Fusion and Online Hard Sample Mining	a dataset consisting of 14,986 images of 22 common mineral National Mineral Rock and Fossil Specimens Resource Center of China Mindat.org	Low level features (Color &Shape)	multi-resolution features extracted from ResNet-50	Top1 :88.01% mAP: 87.15%,	• This work used only photos of minerals
[65]	Intelligent Identification for Rock-Mineral Microscopic Images Using Ensemble Machine Learning Algorithms	a dataset consisting of 481 images Manually collected with microscope	microscopic images only	R, SVM, and M LP	accuracy of about **90.9%**	• mineral microscopic images only used • dataset is very small

2.9.1 **Research Gaps**

Limitation of **S**tate-of- **T**he- **A**rt (**S**o**T**A) works (Xiang Zeng, 2021)**,** (Bona Hiu, 2022) and (Teng Long, 2022)**,** (Junyu Zhang, 2022):

- There are some minerals which are not included in the existing works; especially local minerals.
- Works good under poor lighting conditions (too bright or too dark) but not enough.
- The method achieves 95.6% accuracy, still require better method to increase accuracy.
- Image processing techniques are not sufficient.
- Transparency, density, streak of minerals was not included in the existing works.
- Attention is not given for tiny minerals.

Deep Learning (Yolo V5) (Junyu Zhang, 2022) and (Bona Hiu, 2022) Machine Learning (Random Forest) Model was utilized which is outmoded recently. And Deep Learning (Yolo V5) (Junyu Zhang, 2022): **YOLOv5** follow the original anchor-based pipeline of YOLOv3. However, the anchor mechanism has many known problems.

- First, to achieve optimal detection performance, one needs to conduct clustering analysis to determine a set of optimal anchors before training. Those clustered anchors are domain-specific and less generalized.
- Second, anchor mechanism increases the complexity of detection heads, as well as the number of predictions for each image. On some edge AI systems, moving such large number of predictions between devices (e.g., from NPU to CPU) may become a potential bottleneck in terms of the overall latency.

The study in (Teerapong Panboonyuen S. T., 2022)**,** proposed Object Detection of Road Assets Using **Transformer-Based YOLOX** with **Feature Pyramid Decoder** on Thai Highway Panorama. The proposed method (**Transformer-Based YOLOX with FPN**) in (Teerapong Panboonyuen S. T., 2022)**,** learns very general representations of objects. It significantly outperforms other state-of-the-art (SOTA) detectors, including **YOLOv5S, YOLOv5M**, and **YOLOv5L** on the Thailand highway corpus, surpassing the current best practice (**YOLOv5L**) by 2.56% AP. The most important focus points in object detection are anchor-free detectors, advanced label assignment strategies, and end-to-end detectors. These new focal points are still not integrated into the **YOLO** algorithm, and **YOLOv4** and **YOLOv5** are still anchor-based detectors and use hand-crafted assigning rules for training. It is a fundamental reason for the development of the **YOLOX** algorithm (Teerapong Panboonyuen S. T., 2022).

2.10 Chapter Summary

As observed in the above studies in [6], [7], [4] , [8] several researches were carried on mineral identification and classification using deep learning. However, the researchers did not include some minerals which weren't there in the existing works. Image processing techniques are not sufficient and attention was not given for tiny minerals. In addition to this, the models proposed achieves low accuracy that still require better method to increase performance. Transparency, density, streak of minerals was not included in the existing works.

Deep learning (Yolo V5) [8]: YOLOv5 follow the original anchor-based pipeline of YOLOv3. However, the anchor mechanism has many known problems. First, before training, clustering analysis must be done to identify the best anchors in order to attain the best detection performance. These clustered anchors are more specialized and less all-encompassing. Second, the addition of the anchor mechanism makes the detection heads more complex and increases the number of predictions made for each image. Moving such a high number of predictions between devices (such as from NPU to CPU) may constitute a possible bottleneck in terms of overall latency on some edge AI systems.

CHAPTER THREE

3. METHODOLOGY

3.1. Introduction

This chapter gives a detailed discussion about the selected research methodology, the proposed framework and its process, the developed detection model with transfer learning and training from the scratch, and finally the performance evaluation metrics are discussed.

3.2. Research Methodology

3.2.1. Design science research

In this research work, the researcher follows design science research methodology. This approach aims to create and evaluate the artifacts to solve the identified problems by enabling the transformation of the current state into the desired state. It is increasingly recognized as an equal companion to Information System (IS) behavioral science research and the novel paradigm that can tie all the other Information System paradigms together [66].

Hence, design science research is so far another lens or set of analytical techniques and perspectives for performing research. Engineering and the sciences of the artificial are the ancestors of the design-science paradigm. It aims to develop novel concepts, methods, technical advancements, and products that describe how to analyze, design, execute, manage, and use information systems in an effective and efficient manner [67].

DSR is chosen to conduct this study due to the following:

- first, this methodology mainly emphasizes developing knowledge and/or solution, in the building and application of the designed artifact that supports practically the specialist can use to solve the real-world problem. [68]
- In addition, DSR, which is focused on solving practical problems, incorporates prescriptive or solution-oriented knowledge, where the results of scientific justification (such as prediction) can be used to provide answers to challenging field challenges [66] [69].
- Finally, DSR targets attaining knowledge and understanding of a problem and develop a solution for the needs of the business and its environment in the field of information systems (IS) [68]

This study follows DSR methodology. There are several DSR process models that have been suggested by various scholars; however, we will be using the DSRM process model in this work proposed by Hevner et. al. This process model is chosen because other process models such as Archer, Takeda, Nunamaker miss the important phase like objectives of a solution, demonstration, evaluation, and communication. However, Hevner et al. process model is complete and it improves the suggestion of different process models by different scholars. Figure 10 shows the conceptual framework of the Hevner et al design science research methodology.

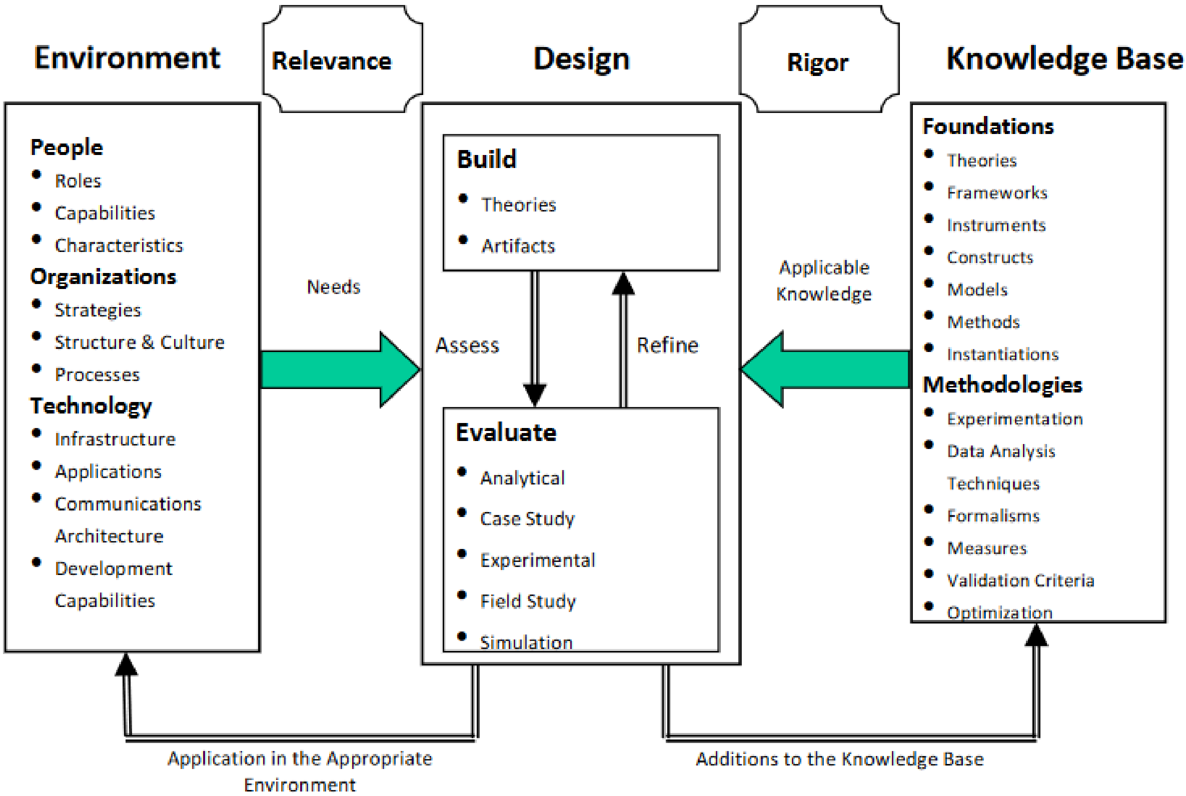


Figure 10: Design Science Research Methodology adopted from Hevner et al. 2004

As shown in Figure 10, this study is grown from relevance by identifying the problem that exists in the environment, which consists of people, organization, and technology. On the other hand, rigor provides appropriate applicable knowledge. The provided applicable knowledge allows building and evaluating the artifact activities of the design itself. The architecture described above is iterative until the generated artifact is added to the knowledge base, providing the foundation for further study, and the designed artifact is capable of meeting the needs of the business. Six steps make up this framework: problem identification and motivation, solution objective, design and development, demonstration, evaluation, and communication. Each design step will discuss in the next section.

3.2.1.1 Design science process model

The research paradigm followed in this subject study is design science research. A set of activities called processes and a proof of concept called a model are defined and achieved receptively. The design science research methodology (DSRM) process model is shown in figure 10 This DSR process model followed in this subject study includes six steps: problem identification and motivation, the definition of the objectives for a solution, design and development, demonstration, evaluation, and communication [70];

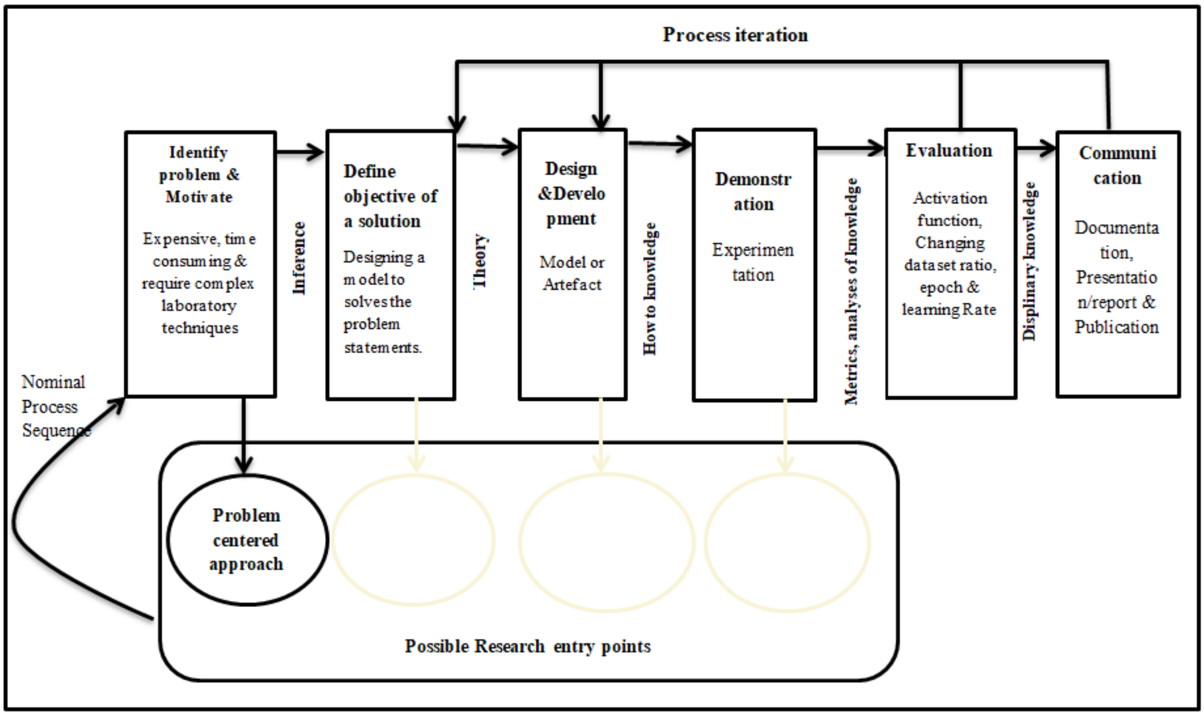


Figure 11: Design science research methodology process models [48]

The entire process of research is broken down into three phases. In phase one, the problem is defined by reviewing literature and observation which leads to an objective formulation that is divided into general and specific. In the second phase, the data were prepared and partitioned into three sets that are training, validation, testing set. After the data partition, different deep learning models were selected. In the third phase, different models have experimented. Then, the models are evaluated and thesis report is generated.

- **Problem Identification and Motivation**

In the problem formulation phase, the researcher has been reviewed different literature and related works previously done related to this research work, to identify practical problems in an actual application environment and also the researcher informally interviewed mineralogist and mineral extension workers to know the gap between existing mineral identification and the state-of-the-art. After identifying and formulated the gaps, the researcher was inspired to fulfill the gap by designing an artifact.

- **Literature Review**

For conducting the research, the researcher reviewed different literatures (Internet and journals) to understand the concept of image processing and different deep learning models, how they applied to solve the related problems in the prior research. During the revision of this literature, the gap of the previous solution will be identified.

- **The Objective of the Solution**

The objectives of a solution are inferred from the problem definition. The objectives of the study that are inferred from the problem specification are explained. Various resources have been reviewed to know the state of the problem, the state of current solutions, and their efficacy. The study is iterative until it meets the users' needs in terms of competence, and the created artifact is primarily used by a domain expert.

- **Design and Development**

In this section, designing the actual artifact using a deep-learning approach, among deep learning approaches yolov7 has been chosen for different reasons. Yolov7, Colab and Jupiter notebook is used for training the model. Python is used for writing the required source codes.

- **Demonstration**

The developed system is demonstrated by simulating how the developed system detects and classifies rock minerals. Inter Core i7-7800X CPU Window, 10pro 64-bit Operating system, Python development environment, is used to implement the model of the system. The window is an open-source cross-platform IDE for scientific programming in Python.

- **Evaluation**

The developed system is evaluated to measure how well it supports a solution to the problem. To evaluate the system in a rational method, testing datasets were fed into the developed model.

Subsequently, the developed artifacts evaluate through different performance evaluation metrics such as precision, recall, F-score, and accuracy, mean Average Precision (mAP) and Intersection over Union (IoU).

- **Communication**

In this section, the researcher presents a research work report for the defense to get feedback and to evaluate the importance and effectiveness of the designed artifact.

3.2.2 Data collection methodology

In this study, datasets were collected from two main sources of datasets; local (i.e. indigenous) image dataset and online publicly available images datasets which only found our region. Publicly available online mineral images dataset will be downloaded from mindat. These datasets are available online and comprises labeled public images of more than 500,000 datasets and that are obtained from mindat. The mineral images which are found in our area were employed in this study from Mindat (A mineral database) and obtained using web spider. Mindat is a mineral database that consists of large samples that was previously collected from all mining industry found around the world.

To collect local data samples of minerals we have been used purposive sampling techniques to select the sites of mining. These data are found in: SHAKISO, DAWA, ADOLA KENTICHA. The datasets were collected from mining sites found in Ethiopia (Lega Dembi& Sakaro Midroc) the largest gold mine in GUJI Zone, Oromia Region of Ethiopia. The researcher uses augmentation techniques to balance the images inside the class and to increase the number of rock minerals images. We have used the following data collection methods.

- **Observation**

We had observed physically the mining site found in Ethiopia (Lega Dembi and Sakaro Midroc) the largest gold mine in Guji Zone, Oromia Region and gather data with the help of minerologist experts in mining sites.

- **Interview**

To get the basic information about the rock minerals the researcher interviewed different experts. During data collection, some experts in guji zone told as there are deferent rock minerals occur in the field but my datasets contain images with fifteen classes.

- **Document Analysis**

The researcher reviewed and analysis different research articles, research repository and websites. We collected datasets from website called mindat.org.

3.2.3 Data Processing

Data processing is the method to perform some operations on an image, in order to get an enhanced image and to extract some useful information from it. It also analysis and manipulate the digitized image, especially in order to improve its quality. It is basically signal processing in which input is an image and out is image or characteristics according to requirement associated with that image. To make the data ready for process the researcher use the following steps: -

Step 1: Image Acquisition is the first step that requires capturing an image with the help of smart phone. Using the augmentation techniques to increasing the number of images in datasets.

Step 2: Preprocessing of the input image to improve the quality of the image and remove the undesired distortion from the image.

Step 3: Image Segmentation is performed to segment the fifteen rock minerals.

Step 4: Feature Extraction after the segmentation, the mineral image extracted for the images. The mineral area is treated as region of interest for image processing. Then the color and shape features are extracted that are used to detect the rock minerals.

Step 5: Rock minerals Detection and Classification, classifiers are used for training and testing of the dataset.

3.2.4 Evaluation Metrics

In this study, the most common performance evaluation metrics used in deep learning object detection and classification model were utilized. These are: Precision, Recall, balanced F-1 score, mean Average Precision (mAP) and Intersection over Union (IoU) which are defined in equations (1), (2), (3), (4) and (5) respectively.

Equation 3.1: Precision Formula

$$\mathrm{P} = \frac{TP}{TP + FP}$$

Equation 3.2: Recall Formula

$$\mathbf{R} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FN}}$$

Equation 3.3: F1-Score

$$\mathbf{F1 - Score} = \mathbf{2} * \frac{\boldsymbol{P * R}}{\boldsymbol{P + R}}$$

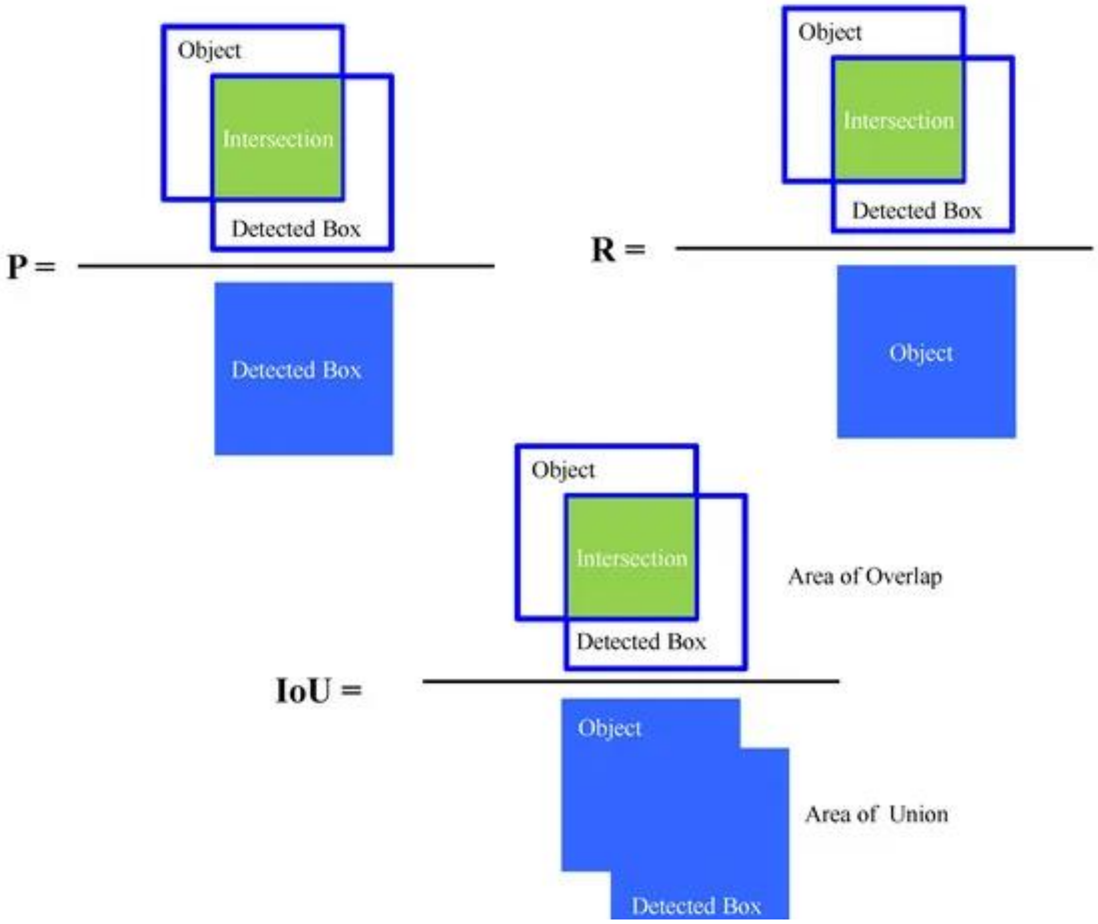


Figure 12: Schematic definition of precision (P), recall (R), and intersection over union (IoU) during object detection. [71]

A. Mean Average Precision Score

mAP score is the mean average precision (map) that is also used as the validation metric for mineral detection. mAP is calculated by dividing the true positives plus false positives by the true positives' average across all classes, as shown in equation 3.4. We employed this technique since it is suggested by [72] as the suitable metrics to evaluate classification having more than one class.

Equation 3.4: Mean Average Precision

$$\boldsymbol{mAP = \frac{1}{\#classes} \sum_{1}^{\#classes} \frac{\#\mathrm{TP}}{\#TP + \#FP}}$$

B. Intersection over Union (IoU)

Intersection over union works by calculating the efficiency and performance of particular model via measuring the overlap area ratio between bounding box predicted by the model and the true bounding area of the object as expressed in equation 3.5.

Area of Intersection (AoI) is defined as the intersection area between the bounding box prediction from the model and true bounding box of the object as shown in Figure 13, and is the union area between the bounding box prediction from the model and true bounding box of the object. For binary classification, if IoU is greater than 0.5, the classified object class can be defined as true positive (TP). For IoU below 0.5, corresponding class can be labeled as false positive (FP).

Equation 3.5: Intersection over Union

$$\mathbf{IoU(MG, MP) = \frac{MG \cap MP}{MG\ U\ MP}}$$

MG= Mineral Ground Truth Bounding Box

MP= Mineral Predicted Bounding Box

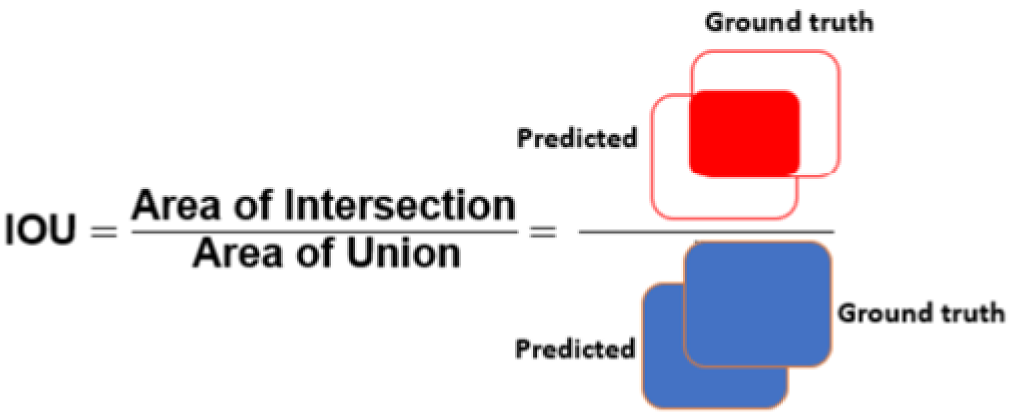


Figure 13: Pictorial Representation of Intersection of Union (IoU)

- There is no overlap between the boxes if the IoU is 0. When the union of the boxes equals their overlap
- The IoU = 1, this signifies that the boxes are entirely overlapping.

3.3 Implementation tools

3.3.1 Software tools

A review of different software tools and their libraries is done in order to choose the best one for YOLOv7 algorithm implementation for Rock minerals picture categorization and identification. We found that there are tools which are generic for both deep learning and machine learning algorithms and tools which are exclusive to only one of them. Prior to selecting the tools, we took into account a number of factors that will help in the selection of the best software tools and libraries. The programming language that will be used to put the algorithm into practice is the first crucial factor. Other considerations include choosing tools with adequate learning resources, such free video lessons and prior experience, as well as the tools' capacity to be used in computers with minimal resources (like CPU or GPU).As software tools to develop a YOLOv7-based model in this thesis, different machine learning libraries(CV2), data science libraries (Pandas, NumPy), visualization libraries (Matplotlib), Python programming language, and coolab Jupyter notebook were used. In addition to being built in a language we are familiar with, Python, these tools satisfy all the requirements for consideration. Additionally, roboflow was used for data preprocessing (an image re-sizing,annotation) and data splitting. The details of these fundamental software program used in this investigation are shown below.

A. Spyder 13.3.6[1]

Spyder is free open-source scientific development environments, that use to write a python program designed by and for scientists, engineers, and data analysts. We used this IDE to write and run the python code. Specifically, when we designed the user interface by using a flask we used this IDE.

B. Anaconda Navigator 24.9.2[2]

Anaconda Navigator is a desktop application that used to navigate, manages conda package and environments without using the command-line command. We used this navigator to launch the jupyter notebook, to manage frameworks that are necessary for developing the proposed system.

C. Jupyter Notebook3 6.0.1[3]

Jupyter notebook is a web application that used to perform different simulations and data visualization. The advantage of using this application is it is easy to write the code, run code, and manage the file, and data processing and visualization. We used this application to write, run a python program and for data processing and visualization while we developed the proposed system.

D. Colab[4]

Colab also known as co-laboratory is a free jupyter notebook that runs on the cloud and stores data on google drive. It allows the users to run deep learning code that takes a long time within a short time by allowing users to use GPU and TPU. We used this web application for the visualization of data, processing data, writing the code, and run the code.

E. Tensorflow 7 1.15.0[5]

Tensorflow is an open-source library developed by Google it uses in deep learning application development for classification and prediction. We used in proposed system development to applying deep learning algorithms for intent classification.

---

1 Home — Spyder IDE (spyder-ide.org)
2 Anaconda Navigator — Anaconda documentation
3 Project Jupyter | Home
4 Welcome To Colaboratory - Colaboratory (google.com)
5 Install TensorFlow with pip

F. Streamlit[6]

An open-source Python framework called Streamlit is used to create web applications for machine learning and data science. Using Streamlit, we can quickly design and launch web applications. You can use Streamlit to create apps in the same manner that you create Python code. Working on the interactive cycle of coding and watching outcomes on the web application is made simple by Streamlit

G. Python

The entire work of the Thesis is coded in Python programming language. Python is one of the most popular programming languages that offer open-source, interpreted, interactive, powerful, fast, high-level, dynamic language and provides a great approach for object-oriented scripting language. Its design emphasizes being highly readable, widely used in data science, and for producing deep learning algorithms because of its ease of use and simple syntax; this makes it easy to adapt for people who do not have an engineering background and allows programmers to express concepts in fewer lines of code than possible in other high-level languages. It has also a large number of standard libraries and community support. Usually, Python supports various programming paradigms at the same time, mainly technical, procedural, object-oriented, functional, and efficient programming. It is mainly used for data managing manipulation and forecasting, making it an excellent tool to use for managing massive volumes of data for training the deep learning model, inserting input, or even making sense of its output. Because of this Python feature, we were able to successfully implement the algorithm in smaller code lines. Thus, the main reasons why Python programming language is used for this thesis are Software quality, Developer productivity, Program mobility, Support Library, Community support, and Component integration to mention a few[i]

[6] https://docs.streamlit.io/library/get-started

CHAPTER FOUR

4. SYSTEM DESIGN

4.1 Introduction

In this chapter, detailed description of the proposed system for the classification and detection of rock minerals is discussed. The model requires passing via a series of steps starting from preprocessing of images, feature extraction and learning to detect into predefine classes. Detection and classification mainly encompass two major phases; these are training phase and testing phase.

4.2 Proposed Model

The proposed system architecture for Rock minerals detection and classification consists of four components, namely: image preprocessing, feature extraction, image segmentation, loading the weight model and data augmentation.

The first component of the proposed model is image preprocessing. This component performs tasks including image resizing, image annotation and removing noise. In addition, this component does the initial task of making the input image ready for the feature extraction component. In feature extraction, the researcher used E-LAN feature extractor to extract features.

The third component of the proposed architecture is loading the yolov7 pre-trained model. In this study, two yolov7 models namely: training from scratch and transfer learning is implemented. Following this, the next component is data augmentation, this technique is applied during training to make the system see a given image from different perspectives and to overcome model over-fitting.

Finally, the yolov7 classifier and detector model are trained to create a model so that the developed model can classify and detect the rock minerals and compare to other models.

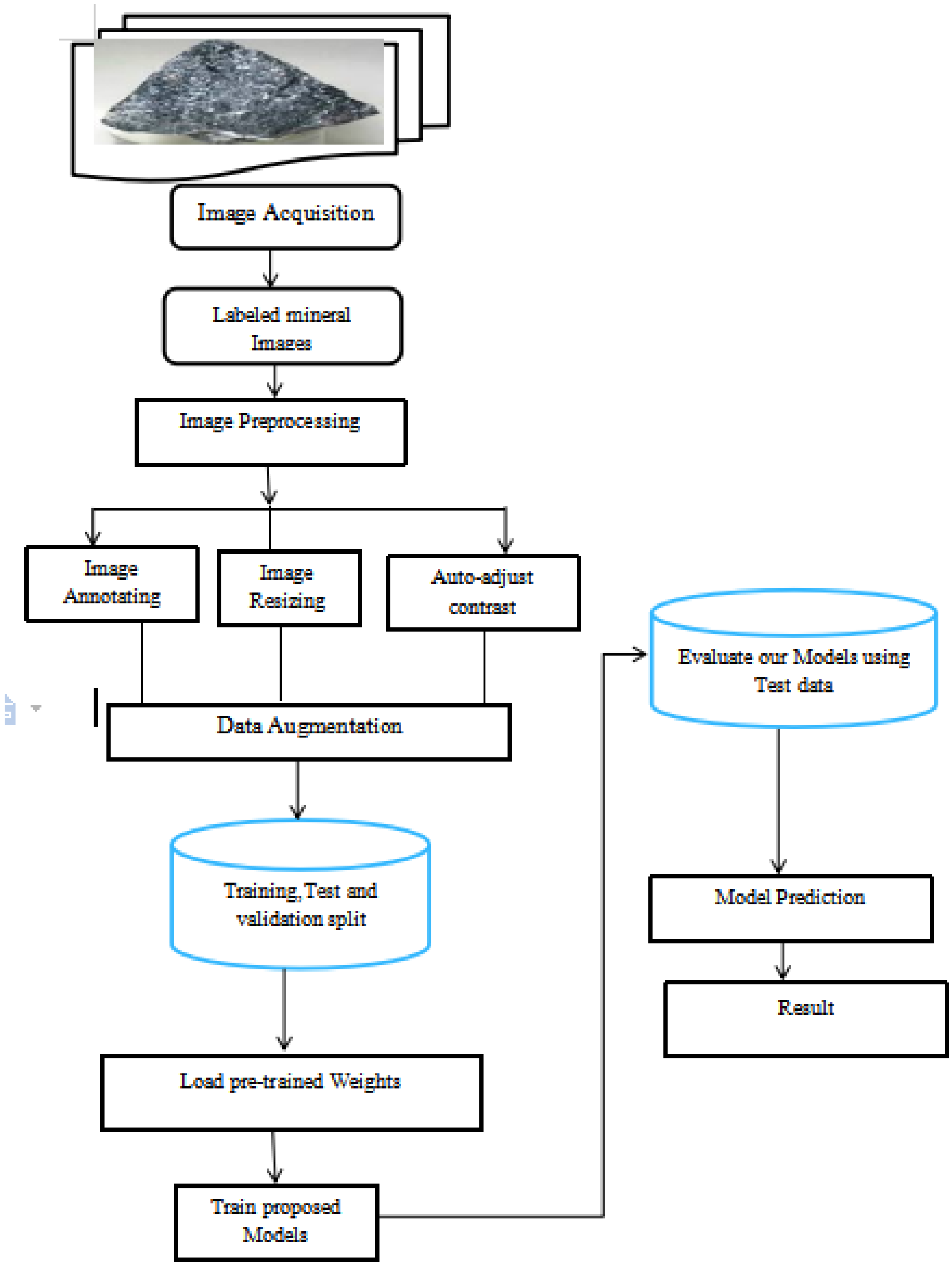


Figure 14: Proposed Model

In general, this model shows the overall process followed to classify and detect a particular input image in either of 15 classes. As shown in figure 14, there are two separate phases in the proposed model. The training phase and the testing phase with a minor difference in method. The training phase starts from importing the input images. That is, several images are processed at a given time in one after the other manner. After image is imported the training phases pass through a different process like preprocessing, feature extraction, loading the pre-trained weight and data augmentation techniques then it outputs the trained model.

In the testing phase, a single image is imported and passed through the preprocessing phases before being submitted to the model for classification.

4.3. Image Acquisition

Image acquisition in image processing can be generally defined as the act of extracting an image into digital form using particular hardware, usually a camera or scanner. The next step is to capture or scan the images into a digital format that can be easily understood by a computer.

The image that is obtained is entirely unprocessed and is the result of whatever technology was used to make it, which can be very significant, in particular fields to have a constant baseline from which to work.

In this study, the data is mainly collected from two main sources of datasets; local (i.e., indigenous) image dataset and online publicly available images datasets. Publicly available online mineral images dataset was downloaded from github[7]. These datasets were available online and comprises labeled public images of more than, 500,000 datasets that were available online[8] and they were obtained from mindat.org[9]. The mineral images employed in this study were obtained from Mindat (A mineral database) using web spider.

From the database, we extracted only mineral images which were found in Ethiopia. The dataset[10] is in two CVS files. The first file img_url_list.csv which contains a list of the image URLs and the raw labels corresponding to the title of the image as seen on mindat.org. The second file img_url_list_converted.csv contains the same image URLs with cleaned up labels and images with no labels removed. All scripts to generate these CVS files are provided alone

[7] https://github.com
[8] https://github.com/loliverhennigh/MinDat-Mineral-Image-Dataset
[9] https://www.mindat.org/loc-21874.html
[10] https://github.com/loliverhennigh/MinDat-Mineral-Image-Dataset

with a script to download all the images. Generating the CVS file of images took around 10 hours.

A total of 2000 images were collected in this study, from this 1300 images were captured with smart mobile phone and the rest are from mindat. To capture the images, Samsung J8 camera is used in automatic mode with autofocus and the camera was approximated to 40 cm above the surface of the minerals. Moreover, all the captured images are in JPEG (Joint Photographer Expert Group) file format. The images are labeled manually to its specific classes by using roboflow.

The dataset is divided into three parts which are training, validation, and testing before the actual training begins. The training split is used to train the model, where object, class and box loss values are calculated and learnable parameters are updated. the validation split is used to assess the performance of the model which is built during the training and it is also used model parameters in order to select the best performing model. Finally, the test split is used only once at the end of the project in order to evaluate the performance of the final model that was fine-tuned and selected on the training process with training and validation sets. The number of Rock mineral images collected for this thesis was 2000. In this thesis, experiments are conducted by using three different ratios which are 07:01:02(70% training, 10% validation, and 20% testing), This means that 70% of the dataset which is 7k images is used to train the model, 10% of the dataset which is 303 images is used to validate the model, and 20% of the dataset which is 2730 images used to test the trained model.

4.4. Training and Testing phase

The training phase is to develop a model that can correctly categorize and detect mineral images. The model builds by the yolov7 algorithm by examining or “learning” from the training dataset made up of data tuples and their linked class labels also called labeled datasets. In the training phase, there are five components namely: image pre-processing, feature extraction, augmentation classification, detection and model training (building) components.

The testing phase is to check the accuracy of the model. Testing is used to evaluate the performance of the final trained model that was fine-tuned and selected on the training process with training and validation sets. In the testing phase, the test image passes through the same image pre-processing components as that of the training.

4.5 Image Preprocessing

The input image passes through different preprocessing such as filtering techniques to get a quality image. The preprocessing task is the fundamental operation in an image processing program that enables a more insightful interpretation of an image prior to use. Additionally, a classifier and detector model's overall performance can be significantly affected by the preprocessing operations. Consequently, in this classifier model, the sub-tasks are used one or more times during the process of implementation. The pre-processing component receives the training and validation Mineral images as input.

4.5.1 Load images

In this task, we simply use roboflow to upload images and store the path to our image dataset in a variable and used to load image-containing folders. The uploaded images are exported to the dataset folders.

4.5.2 Image annotation

Image annotation is the process of image labelling. This implies drawing the bounding box and labelling the class for each image. The image labeling was done manually by using the image data annotation software 'Labeling' from roboflow that was used to create the outer smart polygon of the mineral's portions in all training set images using the 'labeling' package in python. After the successful installation, image labelling is done for each image. After successfully labelling the image, the output is stored as a text file and a class file. The rock mineral images were labeled based on the smart polygon surrounding of the mineral in images.

4.5.3 Image resizing

Image resizing is the first image pre-processing task that defines preserving an important region of an image and improving efficiency. The models take an input image having resolution of 640 X 640 requiring all images to be of this size. In addition to this, image resizing helps to reduce the training time of the neural network. As there are more pixels in a picture, there are also more input nodes, which improves the complexity of the model. This image sizing is applied on all imported images because the accuracy of the feature extraction process is affected if the images are in different sizes.

4.5.4 Image normalization

Normalization is applied to pre-process an input image before further processing begins. In our case, it is used to scale down the feature values in the range between 0 and 1. Image pixel values are an integer between the ranges of 0 to 255. Although these pixel values can be presented directly to the model, they can result in slower training time and overflow, hence normalization. Overflow happens when numbers get too big and the machine fails to compute correctly. We normalized our data values down to a decimal between 0 and 1 by dividing the pixel values by 255. Sample code for normalizing images into decimal values between 0 & 1:

```
x_train = x_train / 255.
x_valid = x_valid / 255.
x_test = x_test / 255.
```

4.6 Feature extraction

Chapter 2 in Section 2.5.4, briefly state that images have different features like shape, texture, and color features. These features are used as a representative of an image. For this study, color and shape features are used to represent an image for the identification and classification of the rock minerals. The researcher used these features because the color and shape feature of an image can differentiate the different categories of the rock minerals. Color and shape features within images can be used for the accurate detection and classification of mineral images.

4.7 Data augmentation

For this study, the data augmentation technique has been applied on the collected images to generate a larger dataset that is more varied for training. To do this on roboflow Keras, deep learning neural network library has been used to transform, randomly shift, randomly rotate, and randomly scale and it delivers the ability to fit models via the ImageDataGenerator class. The ImageDataGenerator built-in function in the python library is used for data augmentation.

This technique will solve the overfitting problem and increase training performance. In addition, data augmentation is a regularization method used to generate additional training data and improve the model's ability to classify. Therefore, using this method, the model can see new, to some extent and modified forms of the input image and the network can learn more robust features. In this study, data augmentation such as flip, rotation, shear and mosaic are used.

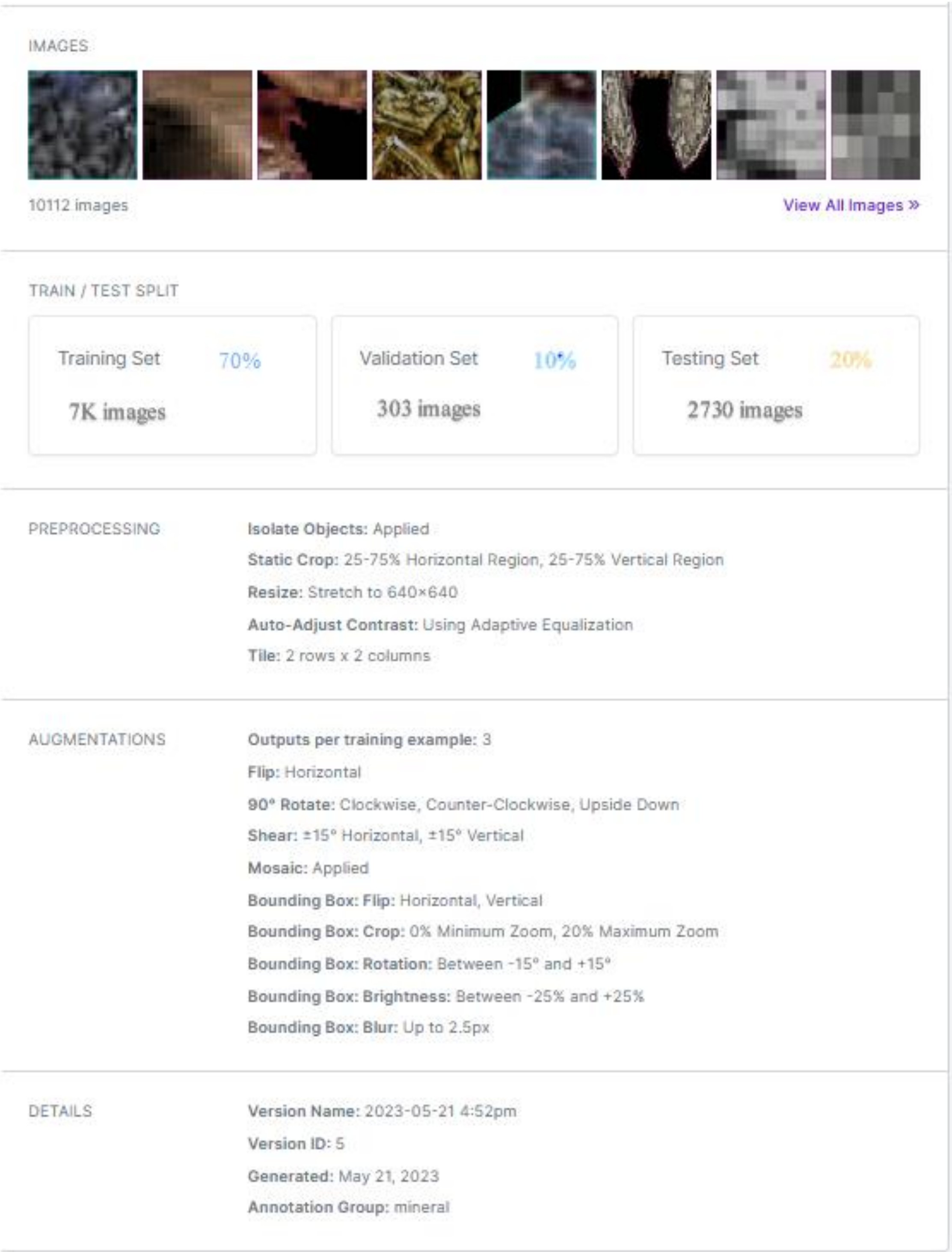


Figure 15: pre-processing and augmentation of datasets

4.8 Loading the pre-trained weights

In this study, we used deep learning algorithms to develop models that are made up of several layers of neurons in a neural network. There are different methods used to develop deep learning models such as convolutional neural networks, Autoencoder, ResNet, Yolo etc. A deep learning algorithm which is Yolov7 is chosen for this study based on different literature that was conducted in computer vision, especially in image classification and detection. Yolov7 is deliberated by matching human's understanding of vision for image-related detection tasks, Yolo is better than other deep learning models for detection and classification for tiny objects.

The main advantage of using the Yolov7 algorithm for the detection of rock mineral is that it is more robust and automated than other deep learning algorithms. Some of the main reasons why Yolov7 is used in this study are:

- It is faster than other state of the art object detectors.
- It has optimized architecture and loss function
- It provides multiple variants of YOLOv7 models for tiny objects and accuracy requirements

The YOLO package is installed by downloading the YOLOv7 code from GitHub and cloning it. The newest version of YOLO v7 is supported by Torch and can be easily implemented with the help of Google colab. The system will then create a new folder with the name "YOLOv7." The pre-trained weights for the model and the unique YOLO directory structure are kept in this new folder. As soon as training is finished, a new subfolder is created in YOLOv7.

The notation 'YOLOv7/run/training/experiment/weights/last.pt' is used to add the path of the subfolder's location. Depending on the yaml document that was used, the document's size and weight will change.

4.9 Training the model from the scratch

The proposed model (i.e., training from the scratch) consists of three layers namely: Head, Backbone and Neck. To train the model, the input images size must be 640 x 640. The number of images in each batch and feature extraction of the data must be in data.yaml. In addition, measurement of an image’s dimensions in both height and width and batch size which is the number of images fed in at once during iteration. The epochs are amount of training repetitions and the data structure, including the location of the training and validation data, the total number of classes and the names of each class was characterized in this YAML file. The input to the first neck is 640 x 640 x 3 images. The Extended Efficient Layer Aggregation Network (E-ELAN) stands for the computing block in the YOLOv7 backbone. Expand, shuffle, and merge cardinality techniques are used in the YOLOv7 E-ELAN architecture to continuously improve the learning capacity of the network without ending the initial gradient route.

The lead head, which is in charge of producing the final output, is the one that contains the projected model outputs. In addition, the head known as the auxiliary head is utilized to help training in the middle layers. While two more connections to the block were produced by convolving the input with 2 and 4 blocks of 3x3 convolutions using the same channel multiplier, input is fed straight into the bottom block with 1x1 convolutions. At the bottom block, all the features are concatenated and 1x1 convolutions is applied to learn rich information.

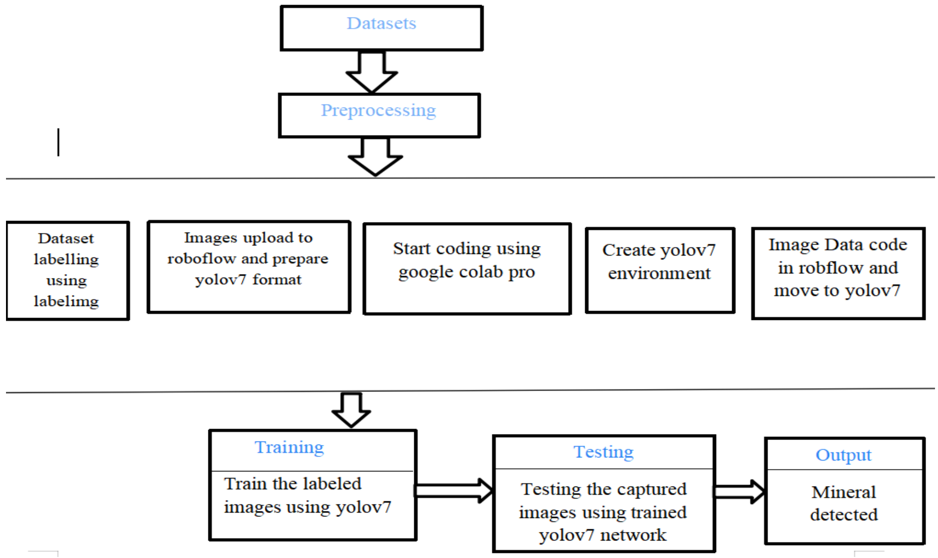


Figure 16: The proposed YOLOv7 model.

4.10 YOLOv7 architecture

YOLOv7 is derived from the Scaled YOLOv4, and YOLO-R model architectures. The YOLOv7 model preprocessing approach is combination of data augmentation and YOLOv5 model preparation. As a result, the model works well at recognizing little things. The Yolov7 architecture used expanded ELAN (E-ELAN) that is proposed as an extension of ELAN. The YOLOv7 backbone is known as E-ELAN. E-ELAN uses Expand, shuffle, and merge cardinality to continuously enhance the network's capacity. It is a group of convolution to:

- Expand the channels and cardinality of the computational block. Apply the same channel multiplier and group parameter to all the computational blocks in a computation layer.
- The feature map from each computational block will be shuffled into group size and then concatenated together.
- Finally, a shuffled group feature map is added to perform merge cardinality.

Group convolution is utilized to increase the channel and cardinality of the computing block in the architecture of the computing block. Different sets of computational blocks are instructed to acquire various features. YOLOv7 also introduces compound model scaling for concatenation-based models. The method of compound scaling allows for the preservation of the model's starting attributes and the best structure. Then the model concentrates on several trainable optimization modules and techniques known as bag-of-freebies (BoF) [54] [73] . The Yolov7 used the planned re parameterized Convolution and Coarse for auxiliary and fine for lead loss. The following figure 17 Network architecture diagram of YOLOv7 along with 5 basic components: CBS, MP, ELAN, ELAN-H.

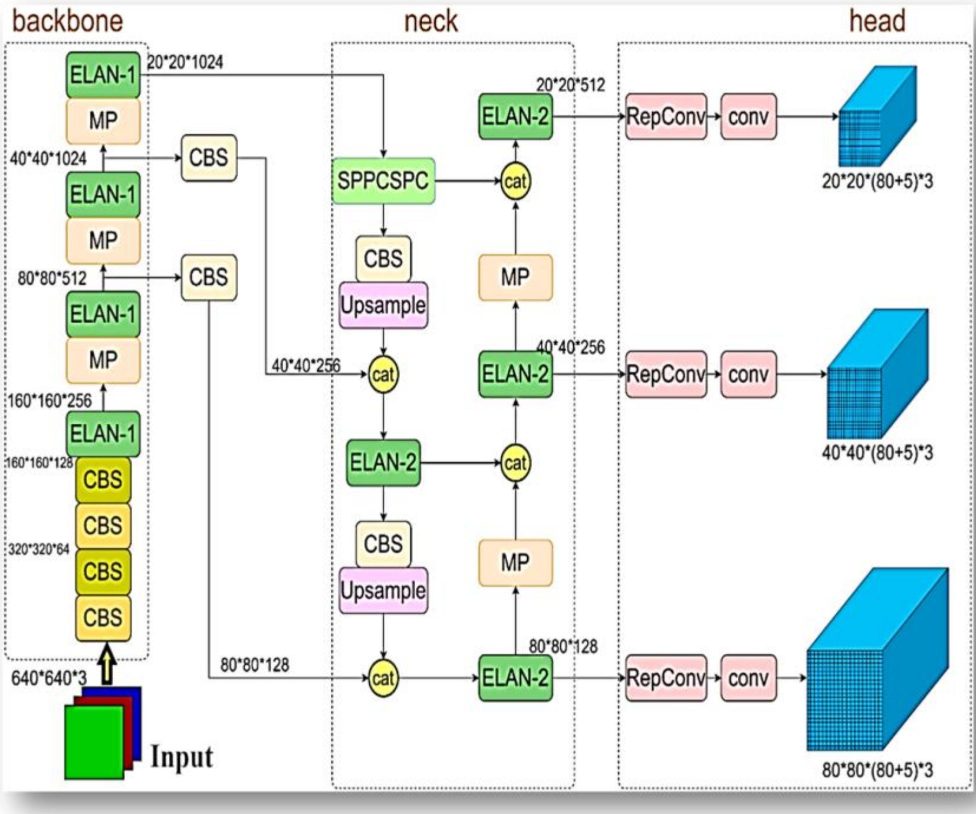


Figure 17: Basic Component of YOLOv7

A. **Planned re-parameterized Convolution.**

Re-parameterization is a technique for enhancing a model following training. It lengthens the training duration but improves inference results. There are two methods of re-parameterization to complete models: model-level ensemble and module-level ensemble. Consequently, Module level re -parameterization has garnered significant interest in the scientific community. In this method, the process of model training is divided into various modules. The outputs are aggregated to produce the fall model.

YOLOv7 use gradient flow propagation channels to recognize the model segments (modules) that require re-parameterization. The architecture's head module is based on the concept of multiple heads. Accordingly, the lead head is responsible for the final categorization [74].

B. **Course for auxiliary and fine for lead loss.**

YOLO's top section contains the anticipated model outputs. Because it was influenced by deep supervision, a popular training method for deep neural networks, YOLOv7 is not limited to a single head. It can accomplish anything it wants thanks to its many heads. The head used to support middle layer training is known as the auxiliary head, whereas the lead head is in charge of the final output.

A Label Assigner mechanism that assigns soft labels based on network prediction outcomes and the ground truth was developed to enhance the training of deep neural networks. Traditional label assignment directly creates hard labels based on predetermined criteria using the ground truth. Reliable soft labels employ computation and optimization techniques that take into account the accuracy and distribution of the prediction output in addition to the ground truth [73] [54].

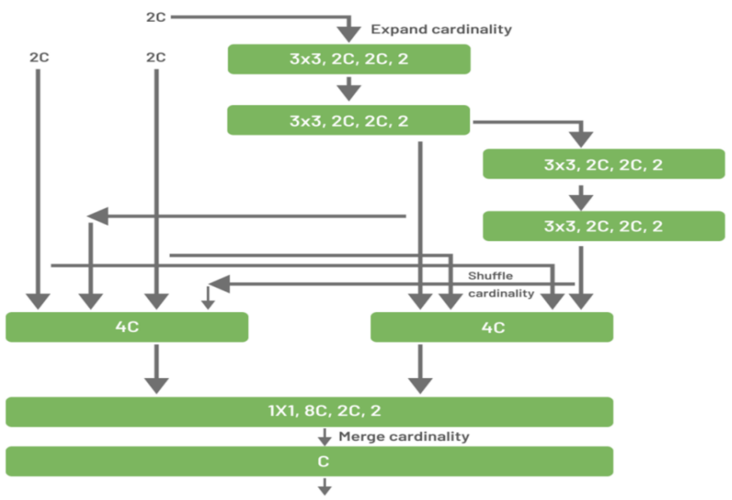


Figure 18: Extended Efficient Layer Aggregation Network (E-ELAN) [75]

4.11Designing Yolov7 models using Transfer Learning

The Yolov7 pre-trained neural networks are more effective for image detection and classification. It has high classification accuracy than other pre-trained models. These are the way to retrain the networks, which are previously trained on more than a million images and able to categorize images into 1000 object categories then classify new images. The different hyper-parameters and their values used for the model are as follows.

A. Optimization Algorithm

To train the deep neural network an optimization algorithm plays a major role because it allows a model to learn faster and achieve better performance. There are many types optimazation algorithms from this adam was used. We exiperiment in the learning rate of 0.001 and 0.01. Learning rate is the amount of weight update during training. It is configurable and ranges between 0 and 1. Learning rate is the most important hyper parameter, it controlls how fast a model adapts to a given problem.

B. Loss Fuction

Loss function is the method of evaluating how well the algorithm is on parameters. If the prediction are totally off, the loss fuction will output a higher number. If the model is pretty good, it will output a lower number. As we tune our algorithm to try and improve the model, loss function will tell us if we are improving or not. Loss helps us to understand how much the predicted value differ from actual value. We are using binary class entropy which is a prediction algorithm where the output can be either one of two values, indicated by 0 or 1.

C. Epoch

It describe the number of the items the training data shown to the deep learning network. During exiperimentation as baseline, the number of epoch are taken.

D. Batch Size

The batch size is a hyperparameter that defines the number of samples to work through before updating the internal model parameters. At the end of the batch, the predictions are compared to expected output variables and error calculated.

E. Activation Function

After each convolution layer the trained models used non linear activation function called RELU.

• Image size: 640 × 640
• Number of images in each batch: 10
• Feature extraction: coco.yaml
• Developed Yolo: YOLOv7.yaml
• Image: measurement of an image’s dimensions (in both height and width).
• Batch: 16
• epochs: 100

4.11 Detection and classification

Detection and classification perform identifying the sample images of rock minerals into their types of classes based on the features of the image. It is done after the distinctive features are extracted and learned.

4.12 Performance evaluation

After building the detector, model the upcoming process is to evaluate the effectiveness of the models. The performance of the models is evaluated using accuracy, precision, recall, and mAP.

CHAPTER FIVE

5. EXPERIMENT AND RESULT DISCUSSION

5.1. Introduction

In this chapter, the experimental evaluation of the proposed model for automatic classification and detection of rock minerals is described in detail. The dataset used and the implantation of the proposed model are thoroughly described. The effect of the YOLOv7 is evaluated and compared with other YOLO versions. In addition, the test results are presented and compared with other research results.

5.2 Description and Preparation of Dataset

The process used for getting data ready for the classification and detection model can be summarized in: collect dataset, preprocess data, extracting features and transform data. The labeling tool in roboflow was used to label the ground truth box of the images. The number and distribution of minerals dataset tags were counted. The data augmentation improves the training dataset's information, preserves data diversity, and modifies the distribution direction in the original images to increase the trained model's generalization. Labels are counted along the ordinate axis, while their names are represented along the abscissa axis. Rock mineral samples in sufficient numbers are found in the dataset. Label center's ordinate 'y' and abscissa 'x' are its abscissa ratios to the image's height and width, respectively. The center of the image is where the data is equally and precisely distributed.

The mineral dataset contains labels for ground-truth boxes. The size statistics of all image borders are shown. A clustering algorithm generates anchor boxes of varied sizes based on all ground-truth boxes in the dataset, ensuring that the initial anchor box size of the algorithm matches the intended size of the rock minerals. In this case, the ground truth box refers to the boxes annotated around each instance of minerals in the training dataset. These ground truth boxes served as training data for the YOLOv7 algorithm, which used them to determine how to recognize objects from related classes in current photos. The majority of boundary boxes were centered. When establishing the anchor boxes for the YOLOv7 algorithm, it is possible that the system can identify objects more frequently that are closer to the center of the image.

This process is iterative with many loops to prepare the dataset required. The datasets were partitioned into training and testing sets. In the training, the dataset is repeatedly presented to the pattern recognizer, while weights are updated to obtain the desired response, and the testing dataset is then provided a standard used to evaluate the model performance. For this experiment, the researcher used a percentage split method for splitting the dataset into training and testing. Accordingly, 70% of the data is assigned for training the model, and the remaining 30% is assigned for testing. A total of 15 classes.

```
train: data/train
val: data/val
test: data/test

# number of classes
nc: 15

# class names
names: [ 'benitoite','calcite','copper','cuprite','erythrite','gold',
'gypsum','halite','limonite','magnetite','opal','prehnite','pyrite','silver','tantalite' ]
```

5.3 Building the Model

The google colab coding tool was used for writing, testing, and debugging the code. The programming is done using Python 3.9 language. CV2 (Computer Vision 2), a library of programming functions mainly aimed at real-time computer vision, is also used. In this study, the classification and detection of mineral image is developed using one of the deep learning-based approaches with YOLOv7. Moreover, YOLOv7_training.pt was applied to develop classification and detection models namely training from scratch and transfer learning. The experimentation is done on Intel Core ™ i7-7800x CPU, 8 GB of RAM, and a 64- bit window operating system, Deep learning libraries of CV2, PyTorch 1.8.2, and CUDA.

5.4 Experimental Analysis and Result

The models are evaluated in terms of accuracy and speed. Additionally, we have tested the models on different hardware to test the speed. In the following sub section charts and figures are presented to compare the results. All the input images are re-sized to 640×640 pixels in all the experiments.

5.4.1 Experiment I

Let's compare the accuracy results for all models. When we say accuracy, we mean the mean average precision (mAP). Here, we compare the mean average precision used in MSCOCO competition (mAP[0.5:0.05:0.95]) for all models.

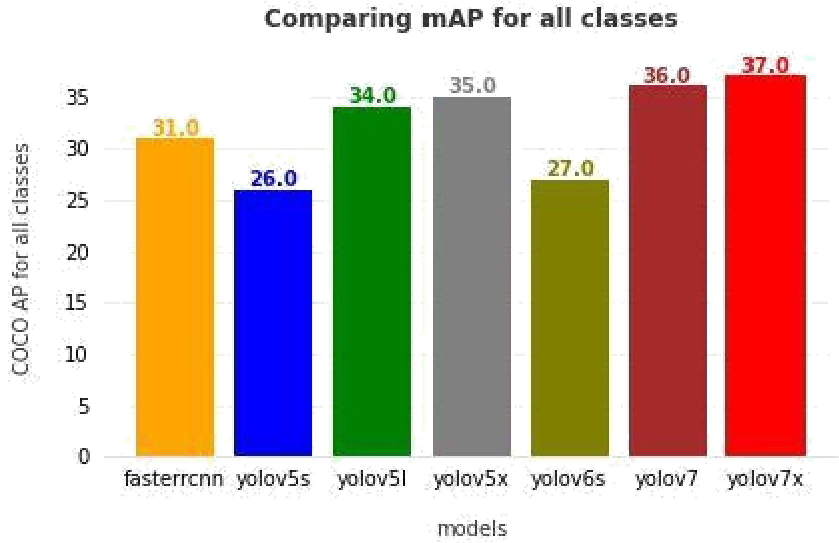


Figure 19: COCO AP for all classes

Figure 19 illustrates a comparison between all models using the COCO AP for all classes. On the other hand, Figure 20 and Figure 21 use 0.5 and 0.75 the AP at fixed IoU threshold respectively to compare the models.

As we can see from the plots, yolov7 models (yolov7 and yolov7x) overcome all the other versions of YOLO family, as well as Faster RCNN model in mentioned two metrics (COCO AP, mAP75).

On the other hand, yolov5s and yolov6s were the worst models among all, with 26 and 27 COCO AP, 21 and 23 for mAP75 respectively.

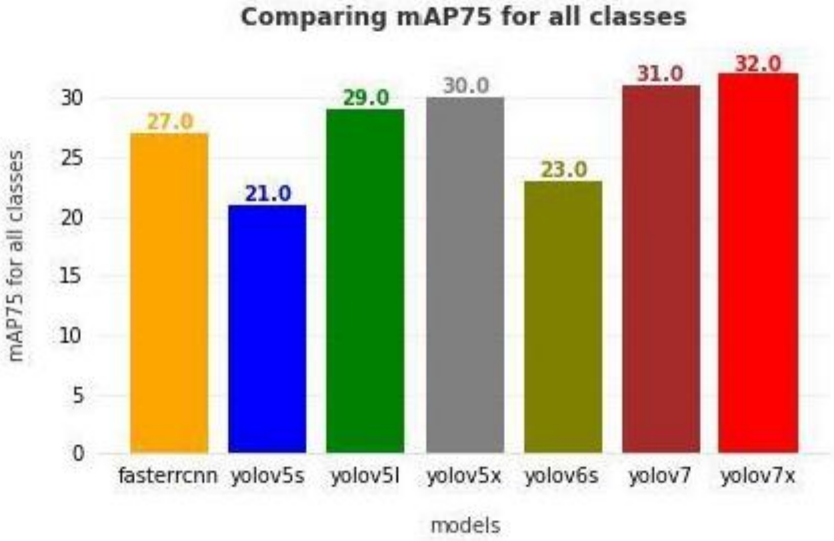


Figure 20: mAP with 0.75 IoU threshold for all classes

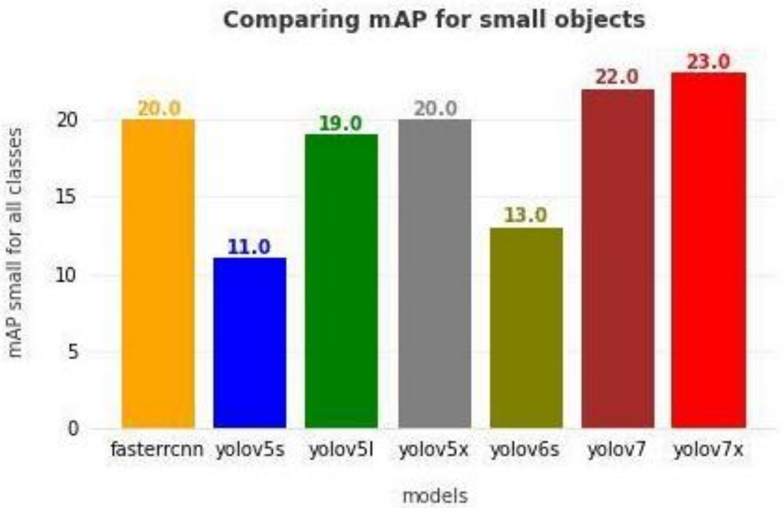


Figure 21: mAP for small size objects

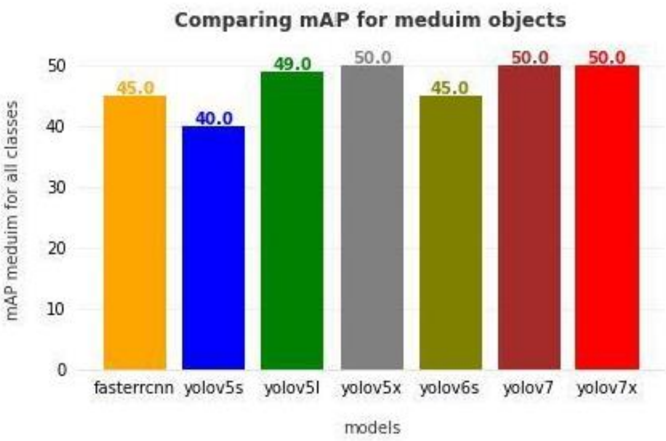


Figure 22: mAP for medium size objects

Faster RCNN model gave a good result and it is comparable with other models like yolov5l and yolov5x, although yolov7 models still better than Faster RCNN.

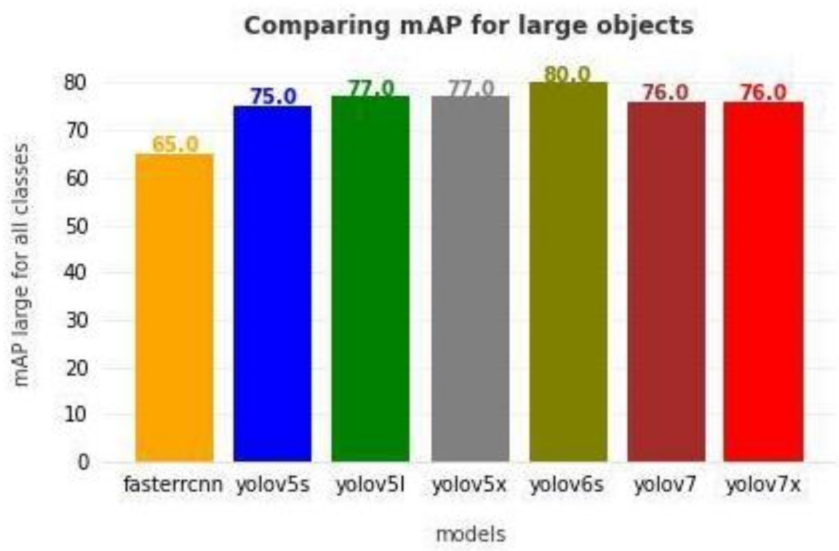


Figure 23: mAP for large objects

Metrics	Detector models						
	Faster RCNN	**YoloV5s**	**YoloV5l**	**YoloV5x**	**YoloV6s**	**YoloV7**	**YoloV7x**
Precision	31	26	34	38	27	39	40
Recall	29.0	23.0	31	33	22	35	37
mAP@0.5	27.0	21.0	29.0	30.0	23.0	31.0	32.0
mAP@0.5:0.95	15	11	17	19	13	22	25

Table 4: Precision, Recall, mAP@0.5, mAP@0.5:0.95 for all models

To understand the models better, we have decided to compare them in more details. Figure 21, Figure 22 and Figure 23 show the result of comparing the seven models using the COCO AP for different objects sizes, in particular small, medium and large objects.

For small objects (Figure 21), as in previous comparison, yolov7 models are the best, yolov7x is the best one to be specific. Although, models like yolov5l, yolov5x and FasterRCNN gave a decent result as well in detecting small objects. On the other hand, yolov6s and yolov5s struggled in detecting small objects.

For medium sized objects, almost all models performed well as we can see in Figure 22. In this case, yolov7, yolov7x, yolov5x and yolov5l had almost the same results. Faster RCNN and yolov6s, on the other hand, had the same mAP as well. However, yolov5s seems to struggle in detecting medium sized objects with the worst mAP among all.

On contrast, nearly all models achieved a good result in detecting large objects as we can see in Figure 23. The interesting thing here is the result of yolo6s, it overcomes all other models even yolov7 and yolov7x. In addition, all YOLO models are better than Faster RCNN in this case.

All the data that we have used to plot the accuracy comparison is available in Table 2. The cells with orange background represent the best result in each column.

Table 2: Model Comparison

Faster RCNN	31	27	20	45	65
YOLOv5s	26	21	11	40	75
YOLOv5l	34	29	19	49	77
YOLOv5x	35	30	20	**50**	77
YOLOv6s	27	23	13	45	**80**
YOLOv7	36	31	22	**50**	76
YOLOv7x	**37**	**32**	**23**	**50**	76

The developed model's efficacy is shown in graphs, which show different metrics of the performance of training and validation sets. Three separate types of loss are box loss, objectness loss, and categorization loss. The box loss assesses an algorithm's ability to precisely locate an object's center and estimate its bounding box. As a metric, objectness quantifies how likely an object can be found in a given area. High objectivity suggests that it is likely that an object lies inside the visible region of an image. Classification loss indicates the accuracy with which an algorithm can determine the proper class of an object. In the course of 0–100 iterations, the model's parameters vary considerably. When the number of iterations increased the model's performance was continuously optimized. The abjectness loss is negligible, as the figure shows; YOLO v7 gives higher precision and recall values than others.

Model	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Pretrained yolov7	99.5	98.7	97.3	88.9

Table 3: Precision, Recall, mAP@0.5,mAP@0.5:0.95 for pretrained yolov7

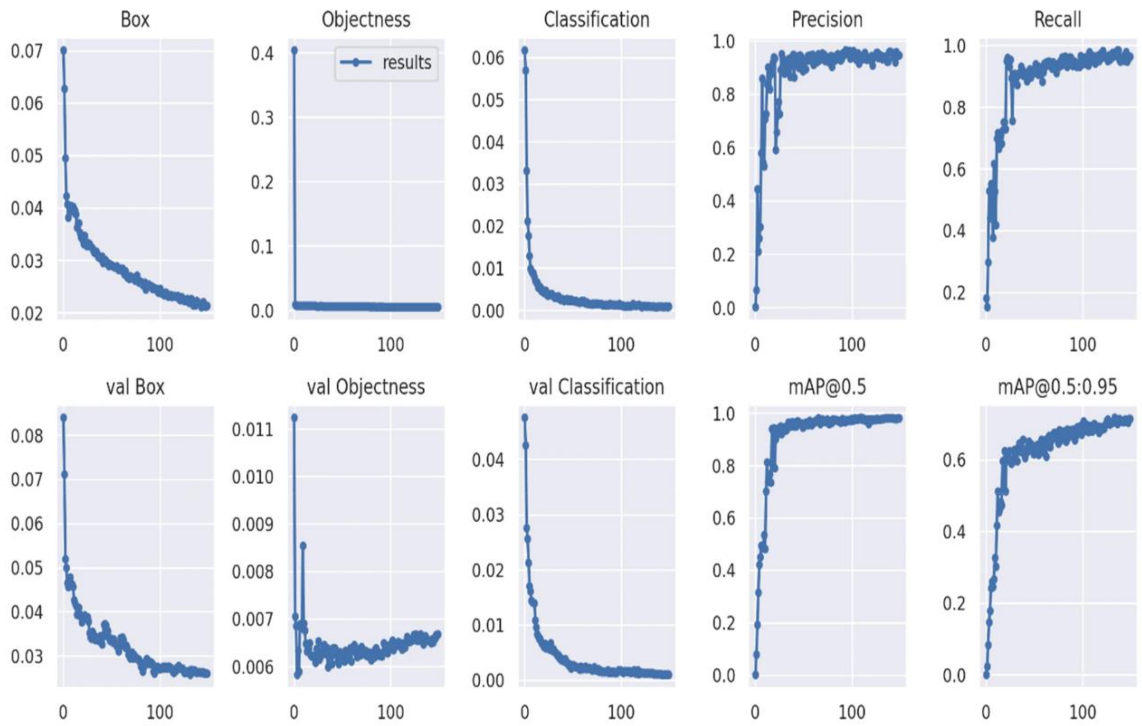


Figure 24: Visual analysis of model evaluation indicators (Precision, recall, and mAP@0.5 for the pretrained YOLOv7) during training.

5.5 Results Demo

The following images show sample results from the proposed model.

Figure 25: Detected and classified validation images (Example 1)

Figure 26: Detected and classified validation images (Example 2)

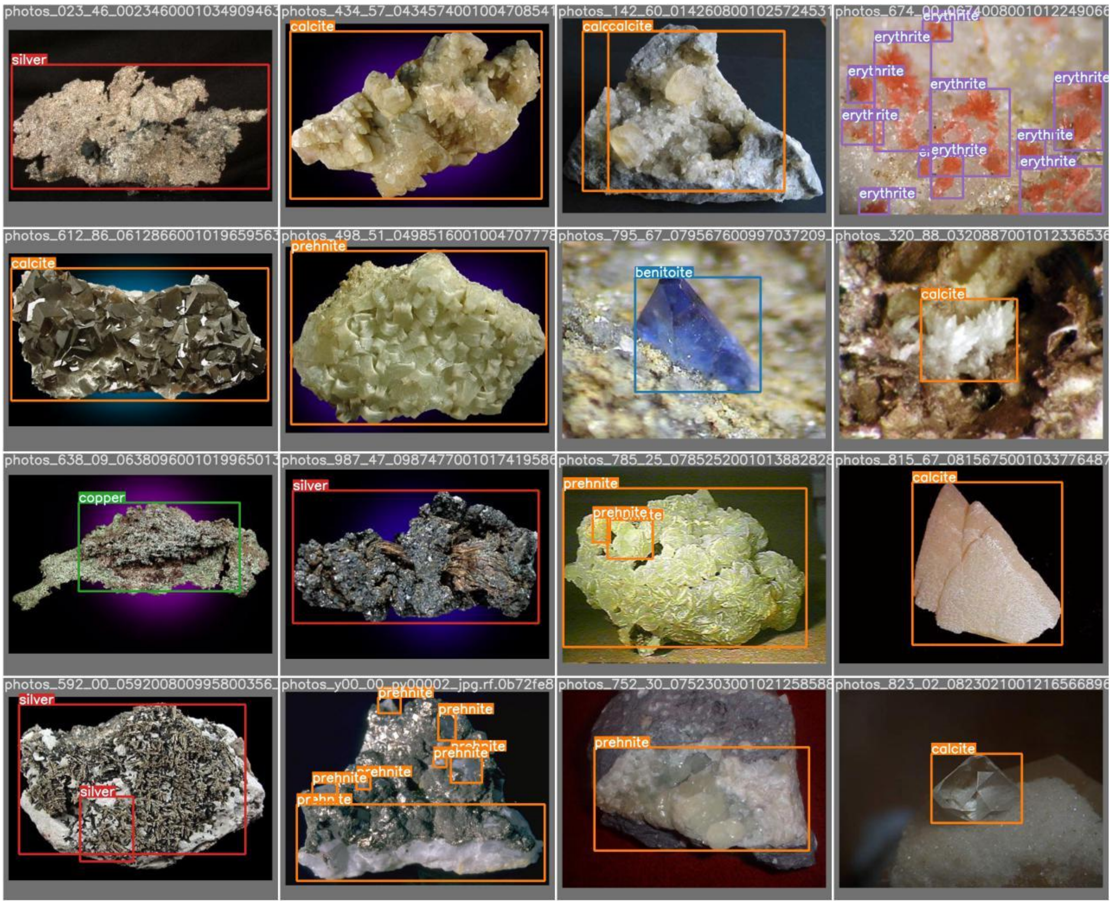


Figure 27: Detected and classified validation images (Example 3)

Figure 28: Detected and classified validation images (Example 4)

5.6 Summary of experimental result

In this chapter, we have seen the experimental results of this research from dataset collection, dataset preparation, building the model, and the performance of each classifier model. The challenge that was encountered in the experiments is overfitting and oscillation in the training and validation loss or accuracy. It is due to a random sample from our dataset: the dataset at each evaluation step is different, so is the validation loss. In addition, at each epoch, there are 18 (70% of the training dataset, 1200 divided by the batch size, 64) iterations. At each iteration, different samples are taken, trained, and tested thereby creating oscillation or overfitting. The other challenge in this experiment is a computational resource. As stated each experiment has to take two and half days. So, when the research changes each hyper-parameter value and re-runs it takes a lot of time to execute.

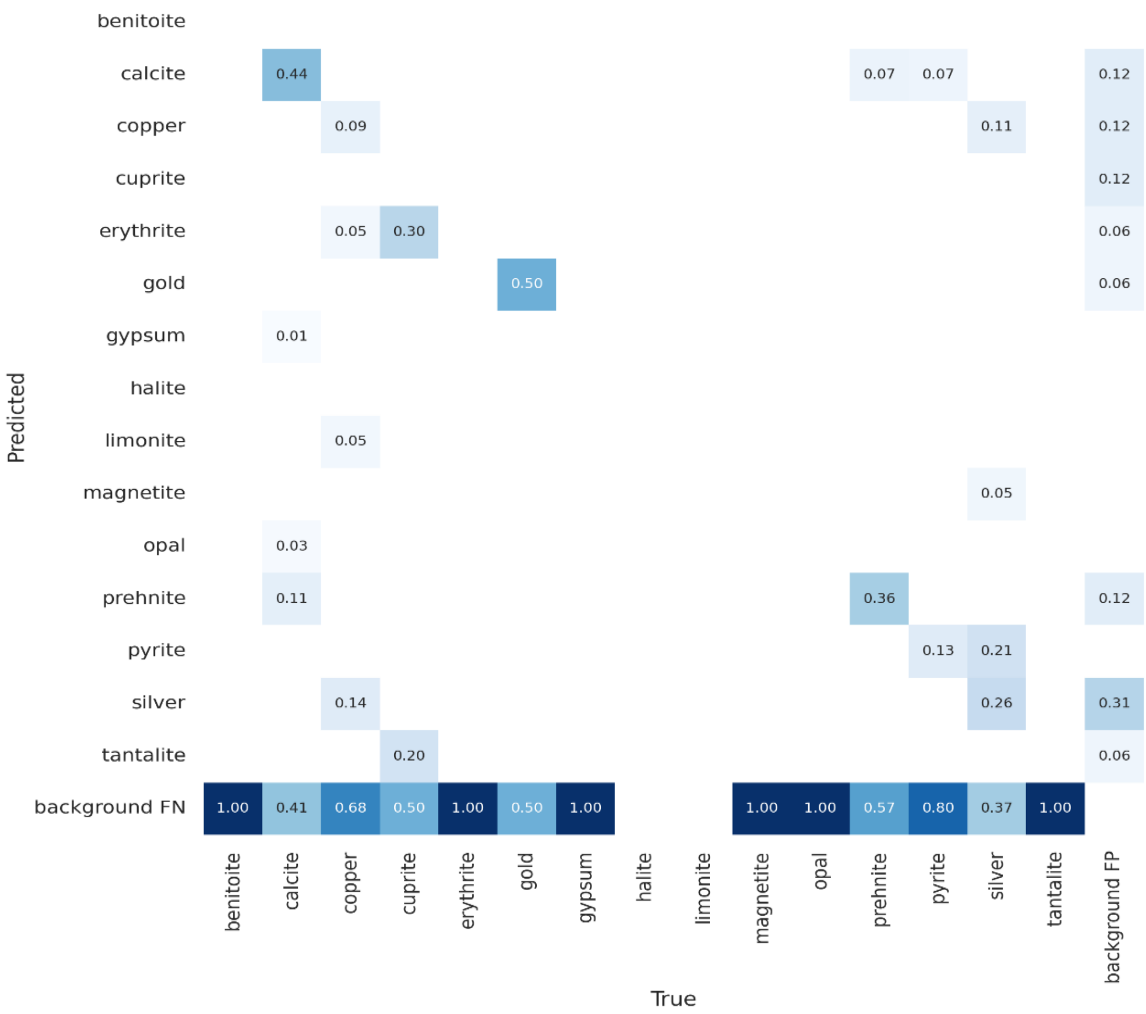


Figure 29: Confusion Matrics for all classes

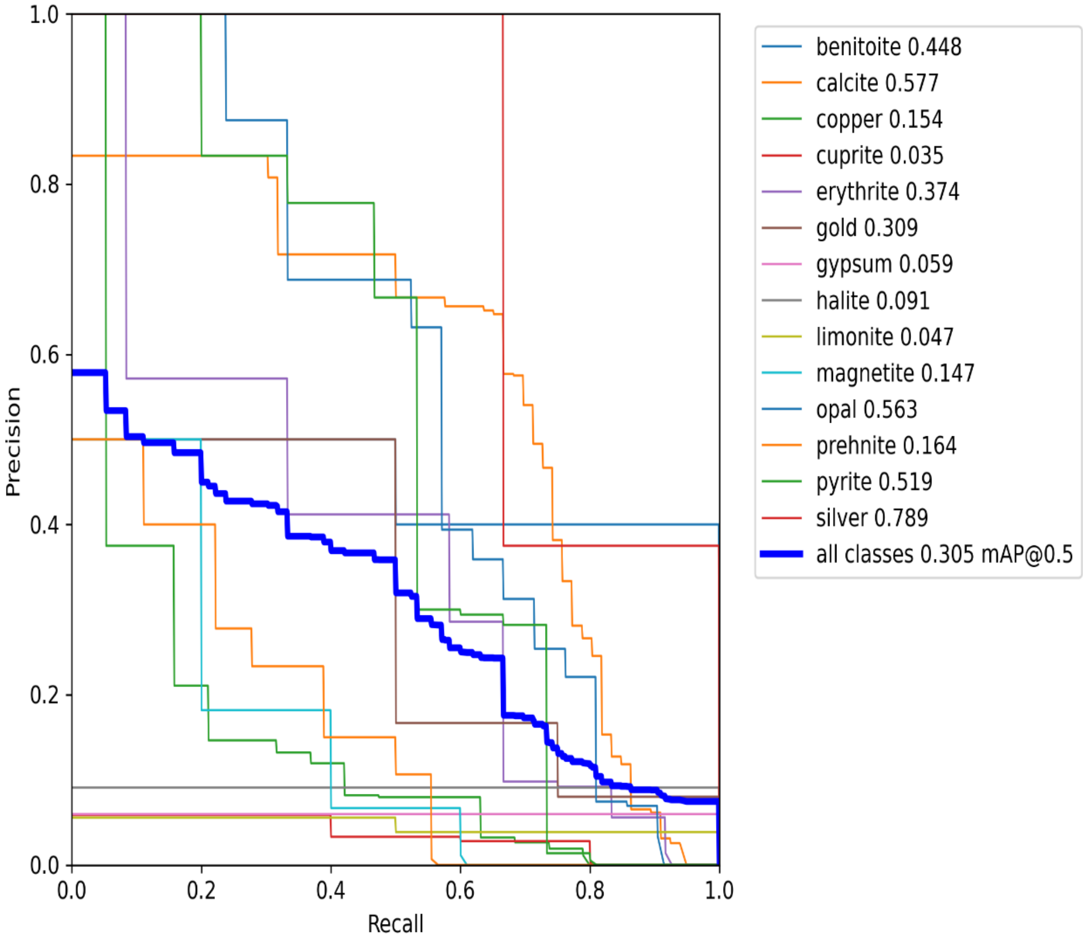


Figure 30: precision, recall for all classes

CHAPTER SIX

6. CONCLUSION AND RECOMMENDATION

6.1 Introduction

In this chapter, the study conclusions based on the findings on the experimentation of the thesis work is reported. Similarly, a suggestion for future research and contribution of the study is forwarded.

6.1 Conclusion

Ethiopia is definitely a place of riches, with a vast and diverse landmass that is rich in resources. Geological studies on Ethiopia's mineral resources have revealed formations that are thought to hold enormous amounts of mineral resources, including proven reserves of metals, coal, precious metals, and industrial minerals.

Geoscience is one of the fields in which there has also been a great deal of effort devoted to the combination of traditional problems and artificial intelligence, such as using artificial intelligence methods for the identification of rocks and minerals. Automatic identification technology could play a significant role in quality inspection of mining industry. Hence, it reduces the hazardous effect of rock minerals thereby, reducing physical contact with the rock minerals. Therefore, artificial intelligence with Image processing techniques plays important role in the field of mineral research.

In this research, the study aimed to develop detector model for rock mineral using a deep-learning approach. To do so, the researcher followed design science research methodology. In this research, 2000 mineral images were collected from Guji zone. After collecting the images, the researcher applied different image pre-processing techniques such as image resizing, image annotation, and augmentation.

YOLOv7 deep architecture was used to develop the model in this study. Specifically, training from scratch (i.e. the researcher settings different hyper-parameters and trained the model from the scratch) and transfer learning on YOLOv7 were used to build a model that can detect and classify rock mineral. During training, image augmentation technique is applied to reduce overfitting and increase the dataset size. Additionally, 70%:30% percentage split is utilized to train and test the model and a confusion matrix and detection report are applied to visualize the model performance. The proposed model automatically detected 15 different types of rock minerals and differentiated each mineral to their classes. Overall classification accuracy is

97.30%. The suggested model outperforms the most recent models reported in the discussion section regarding overall precision, accuracy, and recall. Moreover, the performance of YOLOv7 is compared with the previous version, YOLOv5, and it was observed that YOLOv7 outperforms the previous version. Overall, the system can successfully detect and classify rock minerals while saving time and effort.

6.2 Contribution of the Study

As a contribution to the new knowledge, this research work has contributed the following.

- The researcher designed a classifier and detector using YOLOv7 that can successfully classify and detect rock minerals.
- the proposed model achieves better results in terms of accuracy, the classifier has gotten 97.3% testing accuracy that are above state-of the-art models.

Lastly the research contributions are:

- Dataset:Rock minerals that have a heterogenenous background were collected and prepared by capturing an image from different site. The images were collected with the help of experts. Annotating each image manually and available for the other researchers which found on roboflow.
- There were no research done on rock minerals locally and few internationally. So it used for researches as baseline.
- This research help mineral worker, users, and mineralogist to use and apply the non-destructive way of detecting and classifying rock minerals to avoid much loss in time and equipment.

6.3 Recommendation

The mineral industry is the backbone of the Ethiopian economy no matter how well it is developed. Therefore, there is the need to pay much attention to it so that the right output will come from it. In Ethiopia where there is limited application of technology in classification of rock minerals, mineralogist find it very difficult to identify minerals easily. It is therefore recommended that much attention should be given to the detection of minerals to avoid much loss in time and equipment. In addition, the government should have to focus on training mineral worker, users, and mineralogist to use and apply the non-destructive way of detecting and classifying rock minerals. Moreover, the government also need to implement this kind of research for easy detection and classification of rock minerals and reducing time loss and equipment.

6.4 Future work

- Future researchers can employ batch normalization to accelerate the training process and improve precision.
- The expansion of the dataset is one of the focuses for future development.
- The proposed algorithm can be implemented on a mobile application to facilitate the use of the technology anywhere at any time.

Reference

[1] [Online]. Available: https://www.bgs.ac.uk/discovering-geolo/gy/rocks-and-minerals.

[2] [Online]. Available: https://www.gsi.ie/en-ie/geoscience-topics/geology/Pages/Minerals-Precious-Metals-and-Gems.aspx.

[3] [Online]. Available: https://allaboutethio.com/23-minerals-in-ethiopia-details-gold-gemstones.html.

[4] Z. Z. G. H. Y. B. a. Q. G. Teng Long, "A Review of Artificial Intelligence Technologies in Mineral Identification: Classification and Visualization," *Journal of Sensor and Actuator Networks,* vol. 11, no. 50, pp. 1-24, 2022.

[5] [Online]. Available: https://allaboutethio.com/23-minerals-in-ethiopia-details-gold-gemstones.html.

[6] Y. C. a. C. C. R.-A. Bona Hiu, "Automatic Gemstone Classification Using Computer Vision," *Minerals,* vol. 12 , no. 60, pp. 1-21, 2022.

[7] Y. X. ,. X. J. a. G. W. Xiang Zeng, "Mineral Identification Based on Deep Learning That Combines Image and Mohs Hardness," *Minerals,* vol. 11, no. 506, pp. 1-11, 2021.

[8] Q. G. H. L. a. T. L. Junyu Zhang, "Mineral Identification Based on Deep Learning Using Image Luminance Equalization," *applied sciences,* vol. 12, no. 7055, pp. 1-12, 2022.

[9] M. L. Y. Z. S. H. a. Y. Z. Chengzhao Liu, "An Enhanced Rock Mineral Recognition Method Integrating a Deep Learning Model and Clustering Algorithm," *Minerals,* vol. 9, no. 516, pp. 1-17, 2019.

[10] [Online]. Available: https://oromda.com/.

[11] "Rector of RWTH Aachen University," Aachen University, [Online]. Available: https://www.rwth-aachen.de/rectorate. [Accessed 10 june 2023].

[12] "Rock thin-section analysis and identification based on artificial intelligent technique," *science direct,* vol. 19, no. 4, pp. 1605-1621, 2022.

[13] "outreach," isgs.illinois.edu, [Online]. Available: https://isgs.illinois.edu/outreach/geology-resources/using-characteristics-minerals-identify-them. [Accessed 12 january 2023].

[14] Hevner et al., "DESIGN SCIENCE IN INFORMATION SYSTEMS RESEARCH," *MISQUARTERLY,* vol. 18, no. 01, pp. 75-105, 2004.

[15] "Appalachian state university," [Online]. Available: https://ci.appstate.edu. [Accessed 10 01 2023].

[16] A. Magnus, The Book of Minerals., dalca: Dalcassian Publishing Company, 1967.

[17] "allaboutEHIO," ethioSEO, 2023. [Online]. Available: https://allaboutethio.com/23-minerals-in-ethiopia-details-gold-gemstones. [Accessed may 2023].

[18] M. Nixon and A. Aguado, Feature Extraction and Image Processing for Computer Vision, west reno, USA: Elsevier Science & Technology, , 2012.

[19] m. l. ye zhang, 2019.

[20] R. C. Gonzale, "Digital image processing third edition.," in *Upper Saddle River,* New Jersey: Pearson Prentice Hall, 2008.

[21] V. Tyagi, "Understanding digital image processing," in *India CRC Press,*, 2018.

[22] A. W. P. A. R. K. I. a. A. A. Abdullah, "Digital image processing analysis using Matlab.," *American Journal of Engineering Research (AJER),* pp. Pp. 143-147, 2016.

[23] R. M. a. M. K. A. Rao, "Overview of image processing.," *In Advanced image processing techniques for remotely sensed hyperspectral data,* pp. pp. 51-85., 2018.

[24] G. a. R. R. Bhadur, ""Agricultural Crops Disease Identification and Classification through Leaf Images using Machine Learning and Deep Learning Technique:"," in *A Review in International Conference on Intelligent Communication and Computational Research,*, Jalandhar,, 2020.

[25] K. P. H. D. A. K. S. a. S. C. m. Panigrahi, ""Maize leaf disease detection and classification using machine learning algorithms."," *In Progress in Computing, Analytics, and Networking Advances in Intelligent Systems and Computing,*, 2020.

[26] A. A. K. D. a. M. S. Das, ""Leaf disease detection,quantification and classification using digital image processing."," *International Journal of Innovative Research in Electrical, Electronics, Instrumentation, and Control Engineering,*, Vols. vol. 5, no. 11,, pp. pp. 45-50,, 2017..

[27] Y. R. a. J. K. Anjali, ""A Comparative Study of Image Segmentation Methods,"," *International Journal of All Research Education and Scientific Methods,*, pp. Pp. 232-235,, 2017.

[28] K. Dilpreet and K. Yadwinder, ""Various Image Segmentation Techniques,"," *International Journal of Computer Science and Mobile Computing,,* pp. Pp. 809-814, , 2014.

[29] A. S. a. I. V. S. Kornilov, ""An overview of watershed algorithm implementations in open source libraries."," *Journal of Imaging,,* pp. Pp. 1-15,, 2018.

[30] A. Bala, ""An improved watershed image segmentation technique using MATLAB."," *International Journal of Scientific & Engineering Research,,* pp. Pp 1-4,, 2012.

[31] G. a. R. R. Bhadur, ""Agricultural Crops Disease Identification and Classification through Leaf Images using Machine Learning and Deep Learning Technique:," in *A Review." in International Conference on Intelligent Communication and Computational Research,,* Jalandhar,, 2020.

[32] R. S. Choras, ""Image feature extraction techniques and their applications for CBIR and biometrics systems."," *International journal of biology and biomedical engineering,,* Vols. vol1, No 1, pp. Pp. 6-16,, 2007.

[33] S. P. a. S. E. Umapathy, ""Various approaches of color feature extraction in leaf diseases under image processing,"," *International Journal of Engineering & Technology,,* Vols. Vol. 7, No. 2,, pp. Pp. 712-717,, 2018.

[34] D. Gabor, " "Theory of communication. Part 1: The analysis of information."," *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering,,* Vols. Vol. 93, , no. 26, pp. Pp. 429-441, , 1946.

[35] U. Marmol, "Marmol, Urszula. "Use of Gabor filters for texture classification of airborne images andLIDAR data."," *Archives of Photogrammetry, Cartography and Remote Sensing, ,* vol. 22, pp. 325-336,, 2011.

[36] R. S. T. R. M. M. a. H. M. De Menezes, ""Object recognition using convolutional neural networks."," pp. Pp. 1-10,, 2019.

[37] Y. Y. L. A. O. S. L. S. W. a. M. S. L. Guo, ""Deep learning for visual understanding."," *Neurocomputing,,* pp. 27-48, 2016.

[38] C. A. V. P. Bengio Y., "Unsupervised Feature Learning and Deep Learning: A review and New Perspectives.," January 2013. [Online]. Available: http://arxiv.org/pdf/1206.5538v2.pdf.

[39] F. Chollet, ""Deep learning with Python".," *Shelter Island: Manning Publications Co,,* 2017..

[40] Y. Y. B. a. G. H. LeCun, ""Deep learning."," *Nature, Vol. 521,No. 7553,*, pp. Pp. 436-444,, 2015..

[41] F. Chollet, ""Deep learning with Python"," *Shelter Island: Manning Publications Co,*, 2017..

[42] F. Chollet, "“Deep learning with Python”.," *Shelter Island: Manning Publications Co,*, 2017..

[43] S. T. A. M. a. S. A.-Z. Albawi, "Understanding of a convolutional neural network.," in *In 2017 International Conference on Engineering and Technology (ICET),*, Antalya, Turkey, , 2017.

[44] X. Y. C. S. W. a. L. Z. Du, ""Overview of deep learning."," in *Youth Academic Annual Conference of Chinese Association of Automation(YAC)*, wuhan, 2016.

[45] L. a. D. Y. Deng, ""Deep learning: methods and applications."," *Foundations and trends in signal processing,*, Vols. Vol. 7, No. 3–4, , pp. Pp. 197-387,, 2014.

[46] L. Jiao, F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng and R. Qu, "A Survey of Deep Learning-Based Object Detection," *IEEE*, pp. 128837-128868, 2019.

[47] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," *IEEE*, pp. 142-158, 2015.

[48] S. Ren, K. He, R. Girshick and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks.," in *In Proceedings of the 28th International Conference on Neural Information Processing Systems*, Montreal, QC, Canada, 2015.

[49] S. D. R. G. A. F. Joseph Redmon, "You Only Look Once: Unified, Real-Time Object Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2016.

[50] D. A. D. E. C. S. R. C.-Y. F. A. C. B. Wei Liu, "SSD: Single Shot MultiBox Detector," in *In proceedings of Europian Conference computer vision*, amesterdam,netherlands, 2016.

[51] J. D. T. D. J. M. Ross Girshick, "Rich feature hierarchies for accurate object detection and semantic segmentation," [Online]. Available: https://arxiv.org/pdf/1311.2524.

[52] Z. X. R. B. D. &. L. H. Xue, "YOLO-tea: A tea disease detection model improved by YOLOv5," 2023. [Online]. Available: https://doi.org/10.3390/f14020415.

[53] P. E. D. L. F. C. Y. &. M. B. Jiang, " A review of Yolo algorithm developments Proc. Comput. Sci.," 2022. [Online]. Available: https://doi.org/10.1016/j.procs.2022.01.135.

[54] C. Y. B. A. &. L. H. Y. M. Wang, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2207.02696.

[55] C. Y. B. A. &. L. H. Y. M. Wang, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.0269," 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2207.02696.

[56] V. N. D. &. D. C. Pham, " Road Damages Detection and Classifcation with YOLOv7. arXiv preprint arXiv:2211.00091," 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2211.00091.

[57] F. Z. X. &. L. B. Yang, "Video object tracking based on YOLOv7 and DeepSORT. arXiv preprint arXiv:2207.12202," 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2207.12202.

[58] S. B. G. W. &. Y. L. Jincheng Chen, "Research on YOLOv7-based defect detection method for automotive running lights,," *Systems Science & Control Engineering,* vol. 11, no. 1, 2023.

[59] K. B. a. L. P. B. Julien Maitre, "Mineral grains recognition using computer vision and machine learning," *Computers and Geosciences,* vol. 130, p. 84–93, 2019.

[60] Y. M. N. O. K. H. A. S. a. Y. K. Natsuo Okada, "Automated Identification of Mineral Types and Grain Size Using Hyperspectral Imaging and Deep Learning for Mineral Processing," *Minerals,* vol. 10, no. 809, pp. 1-12, 2020.

[61] H. W. H. J. A. S. A. L. W. XIAOBO LIU, "Method, Research on Intelligent Identification of Rock Types Based on Faster R-CNN," *Multidisciplinary,* vol. 8, pp. 21804-21812, 2020.

[62] F. C. ,. C. L. a. Y. D. Guangpeng Fan, "A Deep Learning Model for Quick and Accurate Rock Recognition with Smartphones," *Mobile Information Systems,* pp. 1-14, 2020.

[63] Z. Z. X. L. W. L. A. X. X. YANG LIU, "Deep Learning Based Mineral Image Classification Combined With Visual Attention Mechanism," *Multidisciplinary,* vol. 9, pp. 98091-98109, 2021.

[64] M. Y. ,. F. M. ,. M. H. a. H. L. Liqin Jia, "Mineral Photos Recognition Based on Feature Fusion and Online Hard Sample Mining," *minerals,* vol. 11, no. 1354, pp. 1-16, 2021.

[65] M. L. S. H. ,. Q. R. a. J. S. Ye Zhang, "Intelligent Identification for Rock-Mineral Microscopic Images Using Ensemble Machine Learning Algorithms," *Sensors,* vol. 19, no. 3914, pp. 1-14, 2019.

[66] Q. Y. W. a. S. J. Deng, ""Design science research in information systems: systematic literature review 2001-2015," in *In international conference on information resources management,,* ontario, 2017.

[67] N. a. S. D.-C. Mandran, ""Supporting experimental methods in information system research."," in *In 2018 12th International Conference on Research Challenges in Information Science (RCIS),* 2018.

[68] D. B. Bisandu, ""Design science research methodology in Computer Science and Information Systems."," *International Journal of Information Technology Vol. 5,* pp. No. 4 Pp.55-60, 2016.

[69] M. A. a. K. M. Ardakan, ""Applying design research method to IT performance management: Forming a new solution."," *Journal of Applied Sciences,* pp. Pp. 1227-1237, 2009..

[70] A. H. a. A. M. J. vom Brocke, ""Introduction to Design Science Research,"," *no.november,* pp. pp.1-13, 2020.

[71] A. R. Partha Pratim Bhaduri, "WilDect-YOLO: An efficient and robust computer vision-based accurate object localization model for automated endangered wildlife detection," *science Direct,* 2021.

[72] A. V. H. R. N. S. S. E. W. O. a. G. B. Michael Gomez Selvaraj, "AI-powered banana diseases and pest detection," *Plant Methods,* vol. 15, no. 92, pp. 1-11, 2019.

[73] K. e. a. Jiang, "An attention mechanism-improved YOLOv7 object detection algorithm for hemp duck count estimation.," 2022. [Online]. Available: https://doi.org/10.3390/agriculture12101659.

[74] C. V. V. I. S. S. J. &. W. Z. Szegedy, "Rethinking the inception architecture for computer vision.," in *In Proceed ings of the IEEE conference on computer vision and pattern recognition,* 2016.

[75] M. Muhammad Rizwan, "YOLOv7 Architecture Explanation," 2023. [Online]. Available: https://www.cameralyze.co/blog/yolov7-architecture-explanation.

[76] A. B. a. H.-Y. M. L. Chien-Yao Wang, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 27 9 2022. [Online]. Available: https://paperswithcode.com/sota/real-time-object-detection-on-coco?p=yolov7-trainable-bag-of-freebies-sets-new. [Accessed 6 7 2022].

[77] Z. Z. X. L. W. L. A. X. X. YANG LIU, "Deep Learning Based Mineral Image Classification Combined With Visual Attention Mechanism," *Multidisciplinary,* vol. 9, pp. 98091-98109, 2021.

[78] T. T. M. A. R. A. S. C. KEN PEFFERS, "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems,* vol. 24, no. 3, pp. 45-77, 2008.

[79] S. T. ,. W. W. ,. P. S. S. a. C. C. Teerapong Panboonyuen, "Object Detection of Road Assets Using Transformer-Based YOLOX with Feature Pyramid Decoder on Thai Highway Panorama," *Information,* vol. 13, no. 5, pp. 1-12, 2021.

[80] S. L. X. W. a. Q. Z. Xingkui Zhu, "TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios," Beihang University, Beijing, China, 2021.

[81] S. L. W. ,. L. a. J. S. Zheng Ge, "YOLOX: Exceeding YOLO Series in 2021," Megvii Technology, 2021.

[82] S. T. ,. W. W. ,. P. S. ,. S. a. C. C. Teerapong Panboonyuen, "Object Detection of Road Assets Using Transformer-Based YOLOX with Feature Pyramid Decoder on Thai Highway Panorama," *Information,* vol. 13 , no. 2, pp. 1-12, 2022.

[83] Nixon, M.S.; Aguado, A.S. (Eds.), in *Feature Extraction & Image Processing for Computer Vision, 3rd ed,* UK, Academic Press: Oxford, 2013.

[84] Nixon, M.S.; Aguado, A.S, "Feature Extraction & Image Processing for Computer Vision," UK, Academic Press: Oxford, 2013.

[85] R. Gandhi, "towardsdatascience," 09 jul 2018. [Online]. Available: https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms. [Accessed 6 may 2023].