Democratic and People's Republic of Algeria

Ministry of Higher Education and Scientific Research

Mohamed El Bachir El Ibrahimi University - Bordj Bou Arreridj-

Faculty of Mathematics and Informatics

Informatics Department

DISSERTATION

Presented in fulfilment of the requirements of obtaining the degree Master in Informatics

Specialty: **Information and communication technologies**

THEME

Deep Learning Algorithms for Remote Sensing

Presented by

Bouketir Hadda

Habbeche Dounia

Supervised by

Dr. Beghoura Mohamed Amine

Publicly defended on: 20/06/2024 In front of the jury composed of

Bouziane Abderaouf	President
Attia Safa	Examiner

Promotion: 2023/2024

Acknowledgment

We would like to express our sincere thanks to all the professors who have taught us and who, through their expertise, have supported us in our academic pursuits.

Secondly, we would like to extend our gratitude particularly to ***Dr. Beghoura Mohamed Amine*** *for his guidance, trust, patience, and supervision throughout our thesis, for the numerous pieces of advice he has given us, as well as for the time he has spent with us to ensure the success of this thesis.*

We also thank the academic team of our Mathematics and Computer Science faculty for answering our questions throughout these five years.

Our heartfelt thanks also go to the Members of the Jury for their interest in our research, for accepting to evaluate our work, and for enriching it with their suggestions.

Dedication

Whoever said "I am for her... shall obtain her"
and I'm for her even if she refuses, I approached her
praise be to God with love gratitude and appreciation
for the beginning and the end…

To the one who sweated on their forehead and taught me that success comes only through patience and perseverance
to the light that illuminated my path and the lamp whose light never fades in my heart,
from whom I received the precious and invaluable
from whom I derived my strength and pride in myself
My dear father Saleh

To the one under whose feet God placed paradise
who eased hardships for me with her prayers
to the great woman who always wished to see me on a day like this
My beloved mother Wahiba

To my steadfast support and the safety of my days
to those whom I leaned on and they became springs from which I quenched my thirst
to the best of my days and their purity
to the delight of my eyes
my dear brothers Nabil Nassim Yahya and my sister Widad

To all who provided assistance and support on this journey...
to my most loyal friends, companions of years…
those who shared hardships and crises
to those who poured out their feelings and offered sincere advice

I dedicate to you this achievement and the fruit of my success that I have always wished for
Here I am today…
having completed its first fruits by the grace of God Almighty

So
praise be to God for what He has bestowed upon me, and may He help and make me blessed wherever I may be…

Hadda

Dedication

First of all, I would like to dedicate this work to my dear parents
for their love and encouragement, for their support
their sacrifices and the trust they placed in me, my God protect them

To my dear sister
Ikram

To my fiance

To my partner Hanane bouketir and her family

And my last dedication is to my cousin Rania Belmecili

she was a student of this faculty God rest her soul

Dounia

Abstract

Deep learning has revolutionized the analysis of data collected from unmanned aerial vehicle (UAV) imagery, allowing for more profound insights, precise analysis, and enhanced data extraction. This advancement has significantly contributed to the refinement of semantic segmentation techniques. Particularly convolutional neural networks (CNNs) have emerged as powerful tools in this domain, outperforming traditional methodologies. Nonetheless, challenges persist, including feature extraction, class imbalance issues, overfitting, and vanishing gradients that hinder deep neural network training, consequently impacting segmentation performance. To address these challenges, we propose a novel approach by integrating the U-Net architecture with ResNet34 backbone leveraging its strong feature extraction capabilities. These features are further improved by using trained weights from the ImageNet dataset. We train and evaluate the proposed model on several UAV datasets, including Aerial Semantic Segmentation, LandCover.ai, UAVid, and AeroScapes. We achieve remarkable performance, higher accuracy, precision, recall, F1-score, and miou compared to other methods.

Keywords: Remote Sensing, UAV, Semantic Segmentation, Deep Learning, U-Net, ResNet34

Résume

L'apprentissage profond a révolutionné l'analyse des données collectées à partir d'images de véhicules aériens sans pilote (UAV), permettant des insights plus profonds. Cette avancée a contribué de manière significative au perfectionnement des techniques de segmentation sémantique. En particulier, les réseaux neuronaux convolutionnels (CNN) ont émergé comme des outils puissants dans ce domaine, surpassant les méthodologies traditionnelles. Néanmoins, des défis persistent, notamment l'extraction des caractéristiques, les problèmes de déséquilibre de classe, le surajustement et les gradients disparus qui entravent l'entraînement des réseaux neuronaux profonds, affectant ainsi la performance de segmentation. Pour relever ces défis, nous proposons une approche novatrice en intégrant l'architecture U-Net avec le rétroviseur ResNet34 exploitant ses capacités d'extraction de fonctionnalités solides. Ces fonctionnalités sont ensuite améliorées en utilisant des poids entraînés à partir de l'ensemble de données ImageNet. Nous formons et évaluons le modèle proposé sur plusieurs ensembles de données UAV, notamment LandCover.ai, UAVid et AeroScapes. Nous obtenons des performances remarquables par rapport à d'autres méthodes.

Mots-clés : Télédétection, UAV, Segmentation Sémantique, Apprentissage Profond, U-Net, ResNet34

الملخص

أحدث التعلم العميق ثورة في تحليل البيانات المجمعة من صور المركبات الجوية غير المأهولة(UAV) ، مما أتاح رؤى أعمق وتحليلات دقيقة واستخراجًا محسنًا للبيانات. وقد ساهم هذا التقدم بشكل كبير في تحسين تقنيات التقسيم الدلالي. حيث برزت الشبكات العصبية التلافيفية (CNNs) كأدوات قوية في هذا المجال، متفوقة على المنهجيات التقليدية. ومع ذلك، لا تزال تواجه العديد من التحديات، بما في ذلك استخراج الميزات، ومشاكل توازن الفئات، وفرط التكييف، وتلاشي التدرجات التي تعيق تدريب الشبكات العصبية العميقة، مما يؤثر بالتالي على أداء التقسيم. لمعالجة هذه التحديات، نقترح نهجًا جديدًا بدمج بنية U-Net مع العمود الفقري ResNet34 مستفيدين من قدراته القوية على استخراج الميزات. يتم تحسين هذه الميزات بشكل أكبر باستخدام الأوزان المدربة من مجموعة بيانات .ImageNet نقوم بتدريب وتقييم النموذج المقترح على عدة مجموعات بيانات، بما في ذلك Aerial Semantic Segmentation , LandCover.ai, UAVid وAeroScapes حيث حقق نهجنا أداءً ملحوظًا، مقارنة بالطرق الأخرى.

الكلمات المفتاحية: الاستشعار عن بعد، المركبات الجوية غير المأهولة، التقسيم الدلالي، التعلم العميق، U-Net، ResNet34

Table of Contents

Abreviations list

DL: Deep Learning
ML: Machine Learning
U-Net: U-shaped Network
ResNet: Residual Network
KNN: K-Nearest Neighbors
ReLu: Rectified Linear Unit
DBM: Deep Belief Networks
DNA: Deep Neural Networks
SVM: Support Vector Machines
RNN: Recurrent Neural Network
ANN: Artificial Neural Networks
UAS: Unmanned Aircraft System
UAV: Unmanned Aircraft Vehicles
SGD: Stochastic Gradient Descent
LSTM: Long-Short Term Memory
NLP: Natural Language Processing
STN: Spatial Transformer Networks
FCN: Fully Convolutional Networks
PCA: Principal Component Analysis
CNN: Convolutional Neural Network
HOG: Histogram of Oriented Gradients
GAN: Generative Adversarial Networks
TCN: Temporal Convolutional Networks
RPAS: Remotely Piloted Aircraft System
CSN: Combined Segmentation Networks
PSPNet: Pyramid Scene Parsing Network
DDCN: Dual-Deep Convolutional Network

List of figures

List of tables

General Introduction

General Introduction

1. Context

Remote sensing is the process of collecting and interpreting data about the Earth's surface from a distance, using satellites or unmanned aerial vehicles. It has been successfully applied in various fields, such as classification, semantic segmentation, and change detection. However, the increasing volume of "big data" from Earth observation and rapid advances in machine learning, has necessitated the development of more effective methods for processing sensor data. Thus, a generic ML framework called Deep Learning (DL) has been developed, which is a part of artificial intelligence that deals with artificial neural networks and deep neural networks, inspired by the structure and function of the human brain. These neural networks consist of multiple layers of interconnected nodes (neurons) that process and transform data, allowing the system to learn patterns and complex representations directly from raw data.

Several deep learning models have been developed to address different types of tasks, such as Recurrent Neural Networks (RNNs), which deal with temporal, textual, and audio data, Long Short-Term Memory (LSTM) networks, which effectively process sequential data, especially in natural language processing tasks, Generative Adversarial Networks (GANs), used for generating synthetic data closely resembling real data, Convolutional Neural Networks (CNNs), designed for image recognition and computer vision tasks. These have been widely used in many tasks such as segmentation, which is the focus of this dissertation, due to their ability to extract detailed information about land cover. This process involves the division of an image into semantically meaningful regions, facilitating the urban planning identification and the delineation of various land cover types such as water bodies, vegetation, buildings, and roads.

2. Problematic and Objectives

Despite the success of CNNs, they still face many challenges, such as precise feature extraction, class imbalance issues, overfitting, and long training times, especially the problem of vanishing gradients, which make training deep neural networks impossible, affecting segmentation performance.

To develop a more robust and accurate segmentation method for remote sensing applications. This dissertation proposes integrating the U-Net architecture with the ResNet34 backbone to leverage the strengths of both models and enhance segmentation accuracy.

3. Methodology and results

We propose a new approach by integrating the U-Net architecture with the ResNet34 backbone. The U-Net architecture, commonly used in segmentation due to its adaptability and distinctive design that includes encoding and decoding components, offers promising potential for improving segmentation accuracy. Meanwhile, ResNet34, a version of the neural network architecture that's a form of deep learning convolutional neural network created to tackle the problem of vanishing gradients.

To evaluate the effectiveness of our approach, we use four diverse datasets for unmanned aerial vehicles: LandCover.ai, Semantic Aerial Segmentation, UAVid, and AeroScapes. These datasets present situations and environmental types that provide an assessment of the proposed method's performance in various natural and urban settings. Our results show high accuracy in identifying land cover categories and urban features, surpassing the performance of alternative image processing methods.

4. Dissertation Structure

The dissertation is organized into five main chapters:

- **Chapter 1** provides general information on the remote sensing process.
- **Chapter 2** presents a comprehensive overview of deep learning, tracing the fundamental concepts and models such as artificial neural networks, convolutional neural networks, and recurrent neural networks.
- **Chapter 3** discusses the integration of deep learning and remote sensing through UAVs.
- **Chapter 4** introduces our methodology using U-Net with ResNet34 as the backbone.
- **Chapter 5** presents the experimental results and comparative analysis of our approach.

Chapter 1

Remote sensing

1. Introduction

The fundamental pillars of economic development shared by both affluent and impoverished nations hinge upon the rational and sustainable utilization of natural resources the mitigation of pollution and the enhancement of environmental and societal well-being. Achieving these imperatives necessitates a comprehensive comprehension of the human-environment interface and proactive measures to address challenges such as environmental degradation and population pressures, since the 1970s remote sensing and other spatial sciences have emerged as pivotal tools in environmental management natural resource governance economic and urban planning, and pollution control endeavors. Through a myriad of remote sensing applications crucial information has been garnered to inform decision-making processes across diverse environmental domains. The adoption of remote sensing technology in decision-making and monitoring initiatives underscores its significance in facilitating informed and effective policy formulation and implementation [1].

The inception of the term "remote sensing" traces back to 1960 marking the commencement of a field that has since burgeoned into a crucial source of extensive earth-related data, remote sensing continues to evolve and ascend in significance propelled by enhanced accessibility to the wealth of information it yields. The initial foray into remote sensing involved aerial photography, dating back to 1839 when the first atmospheric photograph was captured from an altitude of 80 meters above a French village. Subsequent advancements, such as the invention of the airplane by the wright brothers in 1903 further propelled the refinement of photographic techniques in remote sensing endeavors.

The use of aerial images in the Arab regions was through the first world war by the west by depicting the regions of Suez and some of Egypt but most were limited to military purposes and economic operations, at the beginning of the space age and satellite communications the United States of America launched a rocket in 1946 for space exploration at an altitude of 120 kilometres. In mid-1972 the American satellite (1-ERSAT) now known as Landsat 1-LANDSAT) was placed in orbit around earth and one of the most important areas that benefited from the experiments of the space plant was forest agriculture, weather and climate and land use. Each country has sought to possess a series of satellites to control this technology [2] [3].

In this chapter we outlined the foundational concepts of remote sensing and separate its operational aspects, including energy sources transport methods reception and processing, we will also explore its relevance and broad applications across diverse areas such as: agriculture, forestry, geology, meteorology, mapping and its main contribution to modern scientific and technological developments.

2. Definition of remote sensing

Remote sensing is a technique used to gather information about the Earth's surface by analyzing electromagnetic radiation from a specific source. This method captures and interprets geospatial data to gain insights into terrestrial phenomena and objects. Satellites are the primary tools for collecting this information and transmitting it to ground-based stations. The data obtained is usually presented as images or digital outputs, which are then stored in the computer systems of these stations [4].

3. Remote sensing process

Remote sensing is a process resulting from the interaction between three elements fundamental: an energy source, a target and a vector, remote sensing includes the following steps:

3.1 Energy source

An integral aspect of remote sensing entails the utilization of an energy source often in the form of electromagnetic radiation, to illuminate the object or area of interest. This energy serves as the foundation for remote sensing systems which are classified according to the specific source from which this energy emanates as shown in **Figure 1** [5]:

a. Passive sensing system: this system is reliant upon the natural source of electromagnetic energy namely "the Sun".

b. Active sensing system: this system operates based on an industrial source of electromagnetic energy wherein the sensor assumes the responsibility for emitting electromagnetic rays that subsequently reflect off objects before being received by the sensor once more, this process is commonly referred to as radar.

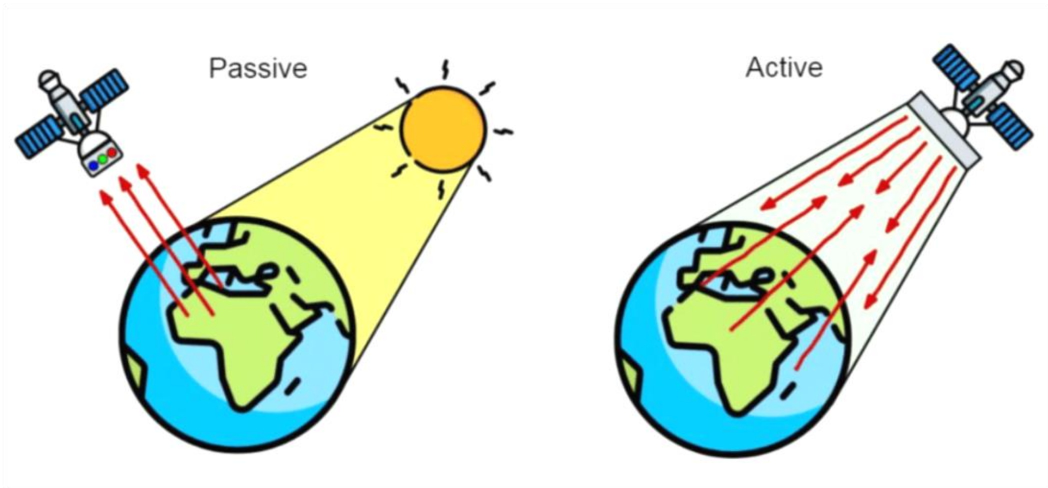


Figure 1: Types of Energy Source

3.2 Energy transmission method

The energy is transmitted as waves to the Earth's surface, where it is reflected and captured by satellite sensors. These signals from surface objects travel through the atmosphere, which affects the energy waves through factors like water vapor, soil, and gases [6].

3.3 Interaction with Target

The sensor in focus is directed towards the earth's surface thus upon energy transmission through the atmosphere, it elicits a response from the target. This reaction is contingent upon not only the chemical and physical attributes of the target and its immediate surroundings but also the radiation properties inherent to the system [7].

3.4 Recording of energy by the sensor

Following the dispersion or emission of energy from the target it is essential to remotely capture this energy using a sensor that remains independent of direct contact with the target. This procedure ultimately facilitates the recording of the captured energy.

3.5 Transmission, reception, and processing process

The energy registered by the sensor is transmitted electronically to the ground reception station, where it undergoes processing into an image format and is subsequently archived.

3.6 Interpretation and analysis

At this point the previously processed image is interpreted either visually or digitally by specialists and software to extract information about the target being studied [4].

3.7 Application

The final step in remote sensing is the application of information from thoughtful target images to understand better them and the extraction of additional information to help solve a particular problem [4].

These are the essential components of remote sensing the remote sensing procedure as depicted in the **Figure 2** [8] :

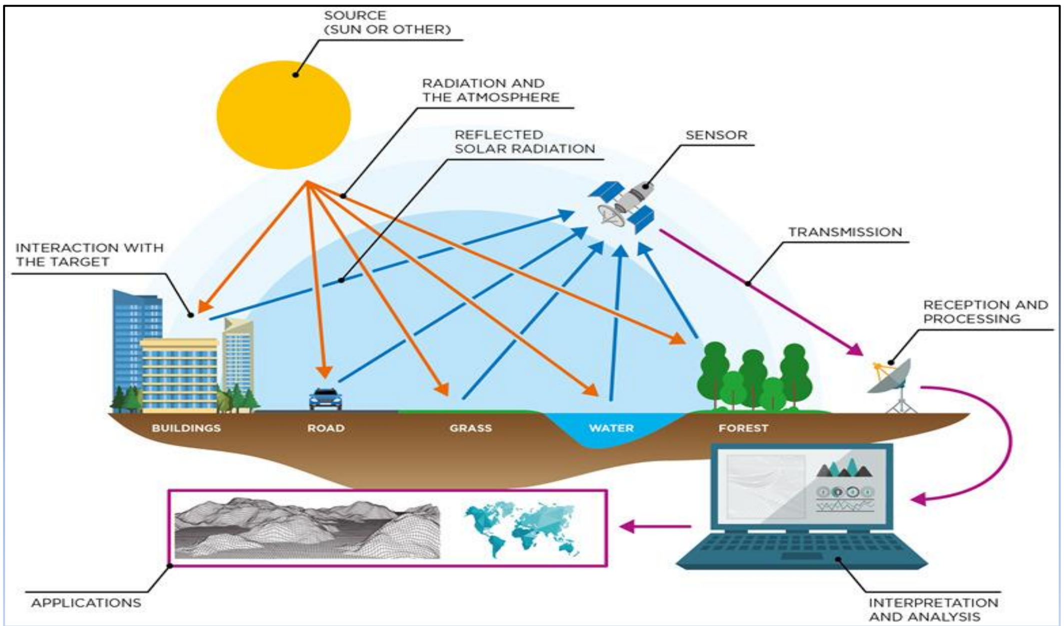


Figure 2: Remote sensing process

4. Electromagnetic radiation

Remote sensing relies on the detection and analysis of energy reflected or emitted by objects in the environment. This energy encompasses various forms of electromagnetic radiation including visible light energy (red green or blue) thermal energy and other wavelengths.

Electromagnetic radiation is characterized by two fields: electric field (E) and magnetic field (M). The electric field varies in intensity and oscillates in a vertical plane perpendicular to the direction of propagation while the magnetic field is oriented perpendicular to both the electric field and the direction of propagation. Both fields propagate at the speed of light (c), which is approximately 300 million meters per second [4].

An important feature of electromagnetic radiation in remote sensing is wavelength (λ) and frequency (F).

a. Wavelength (λ): in the realm of electromagnetic phenomena the spatial interval between successive crests or troughs in a magnetic wave commonly referred to as the wavelength, is quantified in various units of measurement, including meters (m), nanometres (nm), micrometres (µm), or centimetres (cm) [4] [9].

b. Frequency (F): a fundamental concept in wave theory, represents the rate at which waveguides traverse a fixed point in space within a specified time unit, typically expressed in seconds. In the realm of measurement frequency finds its quantification in hertz (Hz) denoting cycles per second [9].

The correlation between wavelength (λ) and frequency (ν) is mathematically articulated by the Equation (1):

$$C = \lambda . \nu \qquad \textbf{(1)}$$

Where,

c: represents the speed of light.

λ: represents the wavelength.

ν: represents the frequency.

Based on the findings outlined in this equation it can be deduced that there exists an inverse correlation between wavelength and frequency as shown in the **Figure 3** Specifically, when the wavelength is greater the frequency tends to be lower and conversely, when the wavelength decreases the frequency increases[7].

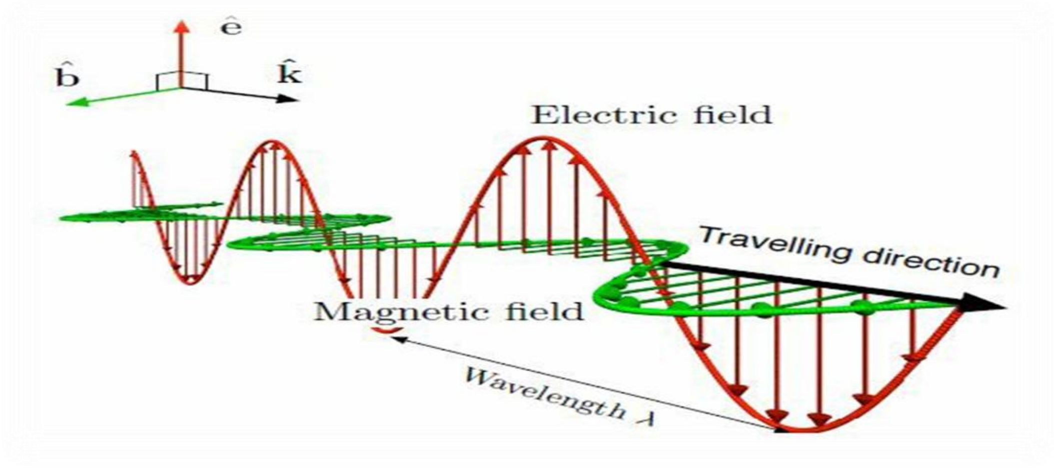


Figure 3: Relationship between wavelength and frequency

The electromagnetic spectrum is categorized based on wavelength into various regions. These regions encompass short waves including gamma rays and X-rays, and long waves comprising microwaves and radio waves. Additionally, the spectrum includes the ultraviolet and infrared regions, delineating distinct segments of electromagnetic radiation as illustrated in **Figure 4** [10]:

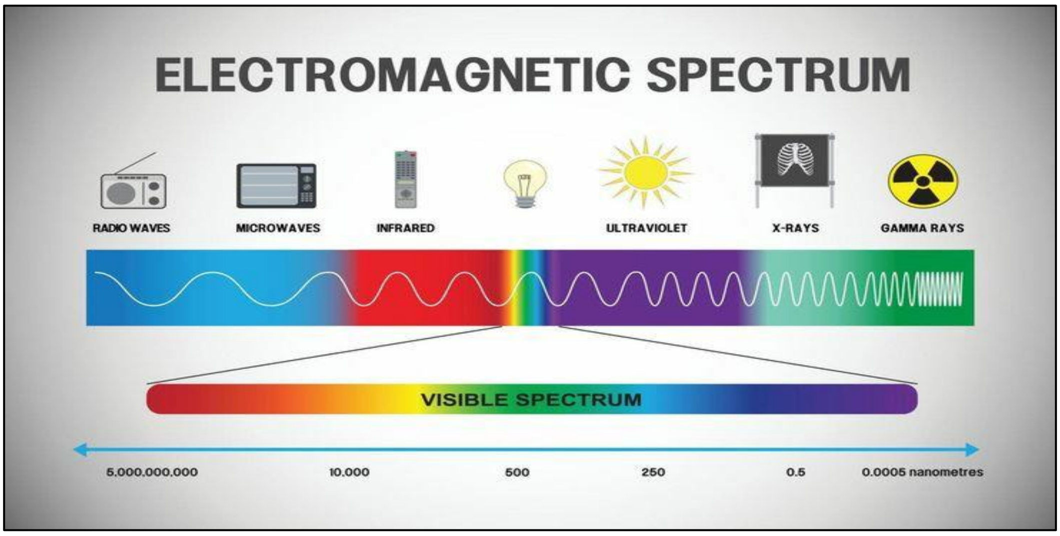


Figure 4: The Electromagnetic Spectrum

4.1 Interaction with the atmosphere

The constituents such as granules and gases present in the atmosphere exert influence over incident light and reflected radiation thereby impacting the quality of space-based imagery. This phenomenon occurs due to:

a. Scattering: dispersion phenomena arise because of the interaction between electromagnetic waves and various constituents within the atmosphere including sizable particulates and gaseous elements. This interaction leads to alterations in the propagation path of the waves. The extent of dispersion is contingent upon several factors including the wavelength of the radiation, the dimensions of the particulates and gases involved, as well as the distance traversed by the radiation through the atmospheric medium [7] [6].

b. Absorption: this phenomenon entails the absorption of energy by molecules within the atmosphere across varying wavelengths. Ozone, carbon dioxide and water vapor stand as the primary constituents of the atmosphere responsible for the absorption of radiation [7] [6].

4.2 Interaction with target

Electromagnetic radiation that traverses the earth's atmosphere without being absorbed encounters the planet's surface where reflection takes place. The most accurate representation of the variability in this reflected radiation is elucidated through an understanding of the reflection coefficient, commonly referred to as albedo. This metric quantifies the percentage of incident solar radiation that is reflected from the surface back into space [11].

- **Specular reflection:** occurs when radiation impinges upon a smooth surface such as a mirror (the surface of water bodies). The rays are reflected away from the surface in a single direction, with the angle of reflection equalling the angle of incidence. Typically, these phenomena manifest as dark colors in the image.

- **Diffuse or Lambertian reflection:** occurs when a surface exhibits a coarse texture such as the rugged rooftops of Rocky Mountains. In this phenomenon, incident rays are scattered or dispersed in multiple directions in a manner nearly identical to each other. Consequently, a suitable portion of energy is diffusely reflected towards imaging sensors allowing for the depiction of surface morphology. Typically, surfaces exhibiting diffuse reflection manifest as lighter tones in imagery.
- **Corner Reflector:** radar waves (microwaves) are specifically exposed to this type of reflection as a result of demonstrations and industrial installations on the earth's surface, here these installations such as buildings, ships, fencing and others with the earth's surface underlying existing angles to the dispersion of some radar rays where it acts as if it were as vertical mirror and looks very bright on the space image.

5. Registration of the energy released or reflected by the target

In the context of remote sensing, particularly within the fourth phase, the sensor plays a pivotal role in capturing reflected energy. It is essential to understand the nature of the sensor and its mechanism for capturing this energy.

5.1 The Sensor

The sensor is an apparatus designed to capture electromagnetic radiation emitted or reflected from a designated target, subsequently converting it into an electrical signal. This device encompasses two primary sensor types [6]:

- **Active sensor:** that emits radiation to illuminate thoughtful manifestations such as radar systems.
- **Passive sensor**: senses the energy reflected from thoughtful manifestations such as the HRV mobile sensor on the SPOT satellite.

The sensor's performance is delineated by its unique set of capabilities [6]:

a. **Spatial resolution:** the capacity to discriminate between two proximate entities within an image quantified in units of meters, denotes a fundamental aspect of visual perception and analysis.

b. **Spectral resolution**: capability is defined as the spectral range within which a sensor can effectively operate and measure, typically expressed in nanometres.

c. **Radiometric resolution:** it is the level of radiation intensity, i.e. the number of levels between white and black in the picture and often estimated at 256 values.

5.2 Platforms used in remote sensing

Sensors within remote sensing systems are strategically positioned on platforms conducive to their operational objectives. The efficacy of these platforms is directly correlated with their elevation greater heights afford broader spatial coverage and enhanced data acquisition capabilities. Selection of the platform type and its specific characteristics is contingent upon several factors, including the inherent properties of the sensor to be deployed, the intended scope of tasks and other operational requisites. This categorization typically manifests into three distinct sections each delineated by its unique operational parameters and objectives [12]:

a. Ground platforms: geospatial targets and phenomena encompass a spectrum of data sources which can be situated in both laboratory and field settings. These sources may range from handheld devices to specialized vehicles equipped with various apparatus such as levers, ground towers, among others. Illustrative examples comprise platforms outfitted with cameras and radar devices tailored for managing traffic flow and mitigating congestion.

b. Air Born Platforms: these platforms spanning balloons, conventional aircraft, helicopters and drones, serve diverse purposes such as aerial photography weather surveillance traffic management, and military reconnaissance. Distinguished by their relatively low economic burden in contrast to space-based platforms, they typically operate within altitudes of up to 37 kilometres.

c. Space-borne platforms: these platforms encompass spacecraft and satellites capable of reaching altitudes up to 36,000 kilometres. They are distinguished by their capacity to cover expansive geographical areas and typically do not necessitate special authorizations from individual states for operation. presently, satellites stand as the predominant technology employed in remote sensing applications.

6. Data transmission, reception, and processing

The energy captured by the sensor is frequently transmitted via electronic methods to a reception station. At this station, the transmitted information undergoes processing to generate images, which can be either in digital or photographic formats.

6.1 Data Transmission

In remote sensing conducted via aircraft, the collected data is analyzed using the aircraft's onboard equipment. In contrast, transmitting satellite data to the Earth's surface requires digital broadcasting [7].

6.2 Data Reception

The output of sensors in remote sensing comprises satellite images which are digitally rendered and presented by partitioning them into uniform spaces known as pixels, each pixel corresponds to a specific area on the earth's surface and is assigned a numerical value representing the intensity of radiation reflected or emitted from that particular location as shown in **Figure 5** [13] . These pixel values encode information about the characteristics and properties of the surface features captured within the corresponding spatial unit [3] [14].

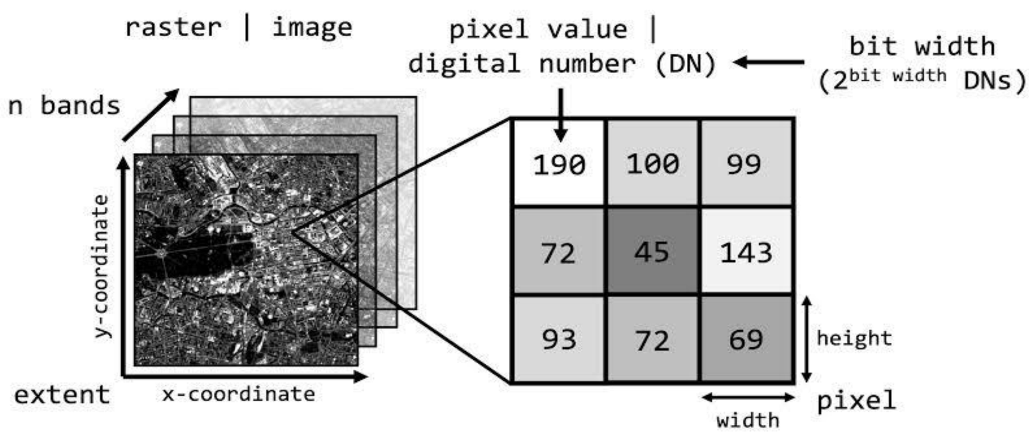


Figure 5: The Components of The Satellite Image

Each pixel within a satellite image is characterized by the following attributes:

- **Coordinates**: each pixel is identified by its position within the image, denoted by its row (i) and column (j) coordinates. These coordinates specify the location of the pixel within the image grid [15].
- **Pixel Value:** the value assigned to each pixel denoted as I (i, j) represent a quantitative measure of the electromagnetic radiation reflected or emitted from the corresponding area on the earth's surface. This value encapsulates information about the properties and characteristics of the surface feature captured by the pixel.
- **The value I (i, j):** is represented by

0: black, 1: white, gray: from 0 to 255, and colour: by RGB space (0..255,0..255,0..255).

6.3 Data processing

Image analysis plays a crucial role in remote sensing as it facilitates the study and comprehension of various characteristics within a given area, there are two primary types of image analysis and interpretation in remote sensing:

a. Visual analysis and interpretation of images: image analysis can be conducted through visual interpretation which involves examining printed or displayed images with the naked eye, this method relies on human perception and interpretation of various visual attributes such as shape, size, pattern, color intensity (relative brightness or color saturation of specific targets) and shade, by observing these visual cues, analysts can infer information about the characteristics and features present in the imagery aiding in the identification and interpretation of different land cover types objects and environmental phenomena [1, 16].

b. Digitally analyze and interpret images: analysis conducted using computer software typically occurs through three key steps:

- **Pre-processing**
  - **Radiological correction of digital data:** correction of deformation arising from sensor malfunctions or atmospheric conditions influencing the reflected or emitted radiation [16].
  - **Geometric correction of digital data:** correction of deformation arising from factors such as satellite dynamics sensor characteristics, earth's rotational motion and alterations in surface phenomena is imperative in remote sensing applications.
- **Image improvement phase:** mathematical processes are applied to the spectral ranges of each layer of the image to show and accentuate the components of the image.
- **Image classification and analysis phase:** image classification constitutes a task or a series of methodologies that adhere to a unified theoretical framework aimed at utilizing images for supplemental analyses or mapping purposes. It is crucial to convert the frequency data inherent in images into objective information regarding land cover or vegetation, typically, two primary approaches are accessible: supervised classification and unsupervised classification.

The objective of image classification is to devise a system with the capability to automatically assign labels to images, this system facilitates the execution of tasks that demand specialized expertise which may be prohibitively resource-intensive for a human observer due to physical limitations such as sustained concentration, fatigue and the substantial time investment necessitated by the analysis of extensive volumes of image data.

7. Remote sensing applications

➢ **Geology:** geology usually explores mineral and petroleum ores with remote sensors, and uses processed images in mining fields, this is based on the fact that each type of rock and mineral has its specific absorption, we must point out here that many attempts have been made to use satellite images in the field of mining. Oil and these attempts were various research attempts knowing that these satellite images deal with surface phenomena. however, while the process of the oil industry is mainly based on dealing with what are called subsurface phenomena it must be said that monitoring ground movements, earthquakes and volcanoes is one of the geological uses of remote sensing working to monitor the movements that occur in the ground layers, follow up on ground cracks. Exploration for raw materials in their natural sources [17].

➢ **Agriculture:** this includes inventorying crops discovering plant diseases studying the specific distribution of land and soil, in addition to identifying the condition of the land, disclosing the expected quantity and amount of crops, detection of diseases and pests affecting crops, drawing up special policies to protect agricultural areas from pollution, drawing maps of the agricultural field to detect agricultural areas and determine the area they cover, monitoring and reducing the phenomenon of desertification, follow the forests and detect fires as soon as they break out [18].

➢ **Water:** by using remote sensing it is possible to monitor the dryness of lands the drying up of lakes and the movement of rivers. In addition, it is possible to deal with expected torrents and floods and this is by comparing images taken at intervals. in addition, it is also possible to search and prospect for water. Groundwater found under the desert sand, via radar images, drawing accurate maps and preparing them, especially for water areas, monitoring, studying and protecting seawater, rivers, and oceans from pollution [19].

➢ **Risks and disasters:** this are represented in reducing floods, earthquakes and torrents, in addition to following up on the displaced and afflicted searching for them and searching for nuclear explosions, in addition to studying the extent of their impact on forest fires and the surrounding areas.

➢ **Outer space:** the role of remote sensing is to monitor the stars and planets and classify some military applications related to space programs into offensive and defensive technologies and applications. We must note that defensive applications are dominant to this day and the credit for this is due to a group of united nations resolutions and agreements, which stipulates the correct and proper uses of outer space.

➢ **The soil:** as for the role that remote sensing plays in the field of soil it is classifying soil into its types and studying it closely, preparing soil climate maps improving the soil through studying and preserving it. Follow up and monitor the lands, and take the necessary preventive measures to prevent drying out of the soil and small water bodies.

➢ **Integrating Artificial Intelligence (AI):** the amalgamation of Artificial Intelligence (AI) methodologies and remote sensing techniques presents immense possibilities for transforming data analysis and applications across numerous Earth scientific areas. There are variety of AI applications in remote sensing, such as data fusion, object and change identification, picture categorization, land cover mapping, and hyperspectral and radar data processing.

8. Conclusion

In recent years there has been remarkable advancement in remote sensing technology, fundamentally altering our comprehension and engagement with the surrounding world. leveraging space-based technologies, we have gained the capacity to make informed decisions and obtain precise insights into the environment, earth and its resources which were once beyond reach. Remote sensing is the cornerstone of our endeavours to achieve sustainable development and promote understanding of our ecological footprint.

In this chapter we have presented the process of remote sensing, and its effective contribution to various applications including artificial intelligence (AI). The focus of our dissertation is on enhancing the synergy between remote sensing and AI through the use of machine learning and deep learning techniques, due to their ability to extract diverse and valuable information from remote sensing data that may not be immediately apparent to humans. Consequently, this enables more efficient processing of remote sensing data, facilitating their analysis and comprehension to a greater extent.

Chapter 2

An overview on Deep Learning

1. Introduction

Deep learning a subset of machine learning techniques often referred to as representation learning, holds particular significance representation learning or feature learning, endows machines with the capacity to autonomously discern relationships from raw data a pivotal capability facilitated by learning across various layers within the network architecture, in contrast to traditional machine learning methods reliant on manually derived representations (feature selection).

The advent of deep learning has supplanted conventional methodologies by virtue of its ability to automatically extract features tailored to diverse problem domains recent years have witnessed deep learning emerge as a primary catalyst for innovative AI solutions buoyed by the proliferation of data enhanced computational resources and refinements in deep network training techniques consequently, grasping the mechanics of deep learning assumes paramount importance for contemporary software professionals given its transformative impact on the technology landscape.

This chapter delves into the intricacies of deep learning, tracing its historical evolution and exploring its fundamental concepts and models. From artificial neural networks to convolutional neural networks and recurrent neural networks, deep learning methodologies have revolutionized the landscape of data analysis and problem-solving, enabling machines to comprehend and process complex information with unprecedented accuracy and efficiency.

2. Machine Learning

Machine learning a subset of artificial intelligence enables machines to autonomously improve their performance through experience rather than explicit programming it has become integral across industries driving automated decision-making in online business, advertising, education and healthcare its adaptability to labeled and unlabeled data makes it a potent technology central to machine learning which is designing domain-specific models trained on data, allowing algorithms to refine performance iteratively the goal is to enable machines to exhibit human-like behavior when faced with new data in the same domain.

Machine learning encompasses various approaches with four primary ones being supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Each approach offers distinct methodologies for training models and extracting insights from data, contributing to the diverse applications and advancements in the field [20].

2.1 Supervised Learning

Supervised learning operates akin to a teacher guiding a child's learning process where the supervisor provides explicit feedback at each step in this method the algorithm learns from known input-output pairs akin to a child recognizing fruits, colors and numbers under the teacher direction [21]. In supervised machine learning algorithms, the desired outcome is already established with a correspondence between input and output data consequently, extensive training data is utilized to train the model enabling it to learn patterns and relationships between inputs and outputs, this training phase allows the model to attain a certain level of accuracy in predicting outcomes. Upon completion of training the model is prepared to process new input data and generate predictions supervised learning encompasses two primary methods:

a. **Classification:** predicting the type or category of a given object.
b. **Regression:** estimating a specific point along a numerical axis.

The most important supervised learning algorithms are: Decision trees, the nearest neighbor (K-Nearest Neighbour), linear regression Support vector machine, and neural networks.

2.2 Unsupervised Learning

Unsupervised learning operates independently devoid of external supervision analogous to a fish learning to swim without guidance this autonomous learning process is devoid of predefined target values associated with input data rendering it unlabeled. In this paradigm, the system must autonomously discern patterns and structures within the input data given the absence of known output values to guide model construction, various techniques are employed to extract inherent data rules patterns and clusters facilitating a deeper understanding of the data and revealing meaningful insights the input data in unsupervised learning lacks a predefined structure akin to supervised training data potentially containing outliers and noisy elements, these inputs are presented collectively to the system which then organizes them into clusters during the training phase, this clustering process aids in uncovering inherent structures and relationships within the data enabling the model to discern underlying patterns and associations autonomously [22]. There are three main types of unsupervised learning methods, these are:

a. **Clustering:** the assembly process divides objects based on unknown features.
b. **Dimension reduction or circulation:** a process of assembling specific features within features with a Level wider and higher.
c. **Learning Rules:** a way to look for Patterns in order flow.

The most important unsupervised learning algorithms are: A priori, K-means clustering, and other association rule mining algorithms.

2.3 Semi-supervised Learning

Semi-supervised learning amalgamates elements of both supervised and unsupervised learning paradigms in this approach, a significant volume of unlabeled data is utilized alongside labeled data to discern underlying patterns by leveraging the abundance of unlabeled data semi-supervised learning aims to reduce the labeling efforts required while concurrently achieving high levels of accuracy in model predictions [23].

The most important semi-supervised algorithms: KNN, SVM and decision tree, The Bayes includes a DL model for example DBM, CNN, and LSTM.

2.4 Reinforcement Learning

In this form of learning algorithms acquire knowledge through a feedback mechanism and accumulation of past experiences the iterative nature of this process entails the algorithm's continual progression toward achieving predefined objectives at each step, the algorithm receives feedback from preceding actions and incorporates experiential learning to anticipate optimal subsequent steps, this iterative feedback loop characterizes the approach as a trial-anderror process aimed at goal attainment reinforcement learning epitomizes a protracted iterative undertaking wherein the system's accuracy improves commensurate with the quantity of feedback received fundamentally, basic reinforcement learning adheres to the principles of the Markov decision process [24]. Reinforcement learning encompasses two distinct approaches:

a. **Model-based approach.**
b. **Model-free approach.**

The most popular reinforcement learning algorithms are: Q-learning, deep adversarial networks, and Temporal difference.

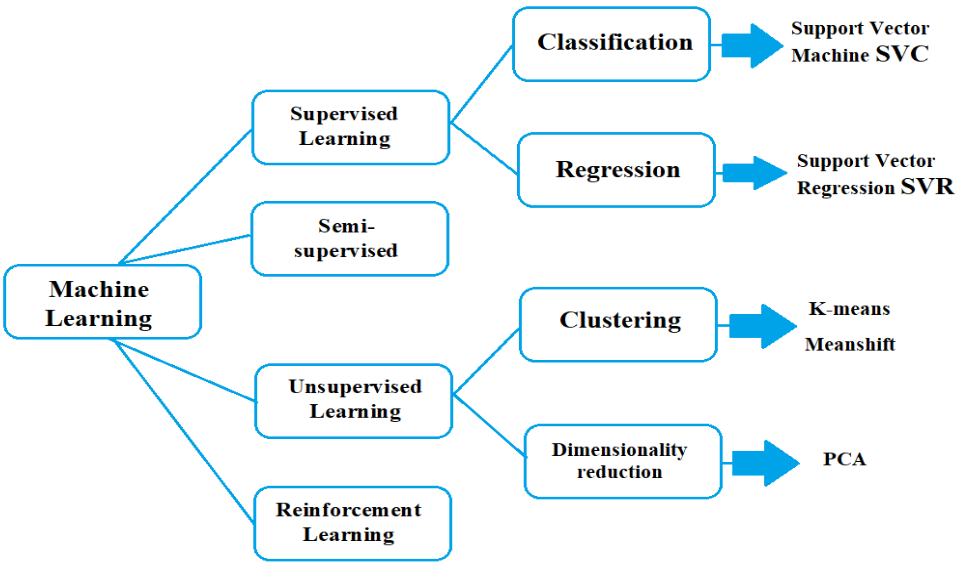


Figure 6: Machine learning encompasses various approaches

3. Deep Learning

3.1 History of Deep Learning

Since the advent of the computer era scholars have engaged in theoretical discussions regarding machine intelligence envisioning the development of intelligent computing systems capable of learning and comprehending solutions to intricate problems, the ensuing table delineates the principal stages of deep learning [25].

Year	Involvement
En 1950's & 1960's	The first neural networks appeared, including the Perceptron algorithm for image classification. However, these early cases were too simple to be widely common.
En 1980's	Researchers developed methods for retransmitting transactions to build and train multi-level neural networks.
In 2000s	Increase the layers of neural networks. these multi-layered networks led to the field of AI research. Called "deep learning" because

	algorithms process data in multiple layers of response.
En 2012	Deep neural networks began to perform better than traditional classification algorithms including machine learning algorithms, this increase in performance is largely due to the increased performance of computer processors (GPUs) and the large amount of data that is now available. Rapid digitalization has led to the production of large-scale data which is the oxygen used in teaching deep-learning models.
Since then	Every year, deep learning has continued to improve and become the best approach to solving problems in many different fields.

Table 1: Deep Learning History

3.2 Definition of Deep Learning

The deep learning emerged in the early 1980s within the domain of machine learning it pertains to the construction of predictive and adaptable algorithms through the utilization of machine learning techniques primarily via pre-training on extensive datasets central to this approach is the utilization of neural networks, which are computational models designed to emulate the functioning of biological neural networks in the human brain.

Neural networks typically comprise three primary layers: the input layer, the hidden layer (or layers) and the output layer within the hidden layer(s) sequential data undergoes weight and activation functions facilitating the transformation and propagation of information to subsequent layers for further processing [25].

4. Deep Learning Models

Deep learning models have emerged as a prominent force within the field of artificial intelligence showcasing remarkable advancements in data processing and problem-solving capabilities leveraging deep neural networks, these models excel in tasks such as classification, prediction and information extraction. Deep learning methodologies enable a nuanced understanding of complex data intricacies, thereby enhancing the performance of intelligent systems.

4.1 Artificial neural networks (ANNs)

4.1.1 Definition of Artificial Neural Networks (ANNs)

This computational model mirrors the cognitive operations and mechanisms observed in the human brain during information processing industrial neural networks function by interconnecting groups of neurons, where each neuron's output is intricately associated with the inputs of others through weighted connections within these networks processing elements apply activation functions to their inputs, akin to the information processing methods employed by the human brain [26, 27].

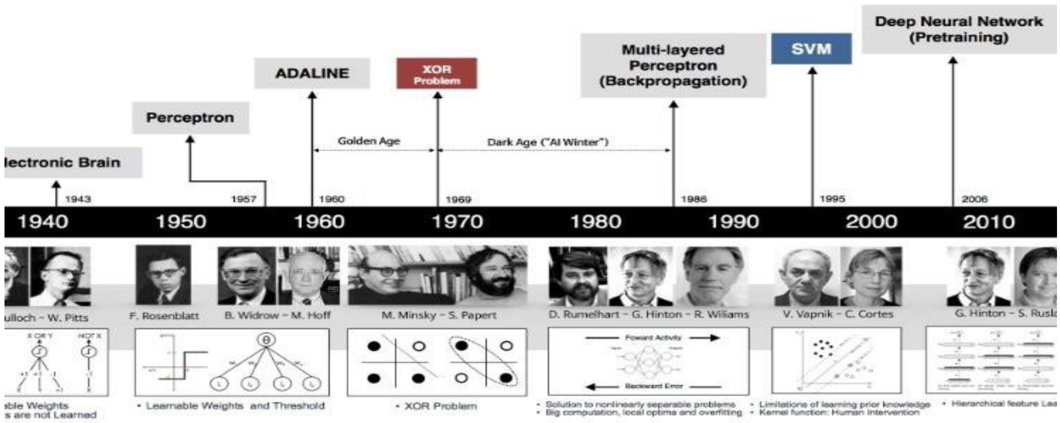


Figure 7: History of artificial neural networks

4.1.2 Neural network structure

a. Biological Neuron Components

The human brain is one of the most complex and diverse organs of the human body. It consists of billions of neurons connected to form complex networks this structural complexity is illustrated in **Figure 8** [28] :

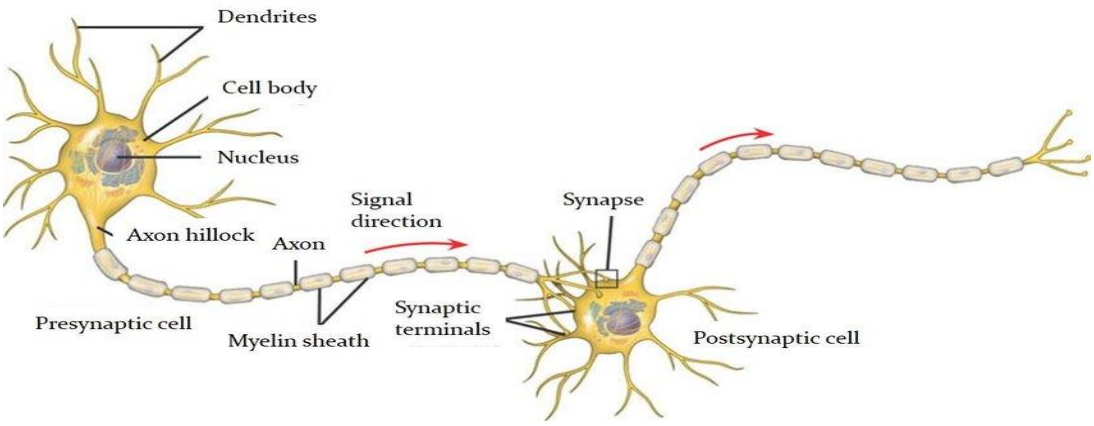


Figure 8: Simplified composition of biological neurons

Figure 8 depicts the simplified composition of the neuron, comprising:

- **Nucleus:** situated within the cell body the nucleus houses the genetic material in the form of DNA and assumes a pivotal role in orchestrating and regulating cellular activities [29].
- **Dendrites:** these intricate fibers surrounding the cell body (soma) serve as receptors for electrochemical impulses originating from neighboring neural fibers constituting the primary input of the neuron [29].
- **Soma (Nerve cell body):** serving as the epicenter for neuromorphic computing the soma acts as the integration hub for neuronal signals overseeing vital functions within the neuron [30].
- **Axon:** extending outward from the cell body axons facilitate the transmission of neural signals to other neurons or targeted organs and tissues, the neuronal axis generates electrical impulses known as action potentials, with a temporal resolution on the order of milliseconds

this internal voltage arises from the momentary amalgamation of impulses emanating from interconnected neural fibers serving as inputs [30].

➢ **Synapse:** serving as the junction between two neurons synapses enable the transmission of electrical and chemical signals from one cell to another [30].

b. Artificial neuron components

As previously noted, efforts have been made to replicate the functionality of biological neurons through the development of artificial neurons.

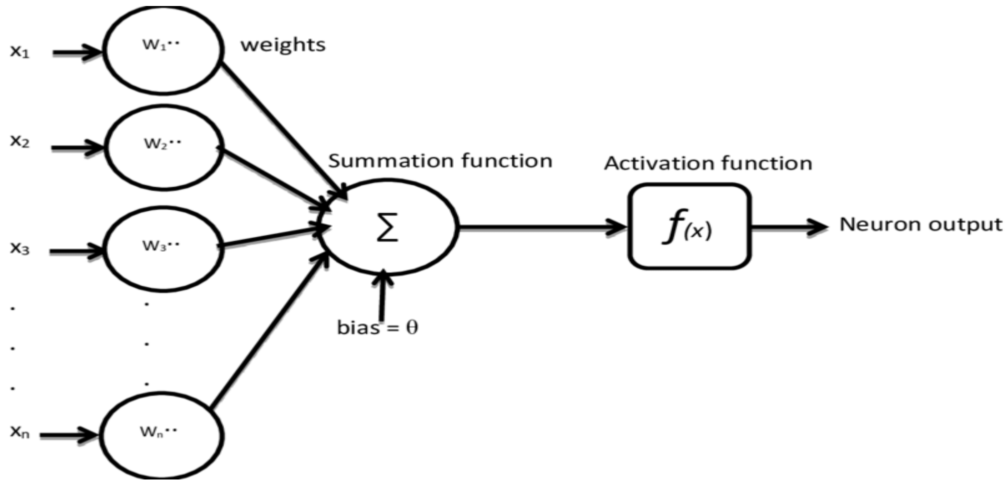


Figure 9: Components of artificial neuron

The industrial neuron comprises five distinct components, namely :

➢ **Input layer Xi:** this layer encompasses external data pertinent to the problem under consideration such as text images sound or other forms of information received by the neuron via signals from adjacent cells these inputs are denoted as **Xi** [31].

➢ **Weights Wi:** representing the strength of the connection between input signals and subsequent elements weights Wi signify the degree of correlation these weights are adjustable parameters within neural networks modulated through the learning process [29].

➢ **Summation function:** this function computes the aggregate of weighted inputs by multiplying each input Xi by its corresponding weight Wi and summing the resultant products mathematically, the total input is expressed as follows:

$$\textbf{Total Input} = \sum X_i W_i$$

➢ **Activation function:** each neuron within the neural network employs an activation function to process incoming data additionally, a threshold value is associated with each neuron this threshold represents the minimum cumulative input required to trigger neuron activation [[29, 30, 32].

- **Neuron output = 0** signifies neuronal inactivity when the cumulative input fails to meet the threshold.
- **Neuron output = 1** indicates neuronal activation when the cumulative input surpasses the threshold level.

Numerous activation functions vary in their characteristics catering to specific objectives and desired output qualities within neural networks, here are some commonly employed activation functions [33]:

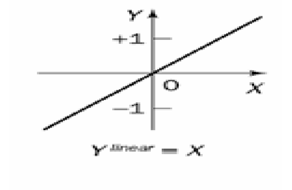
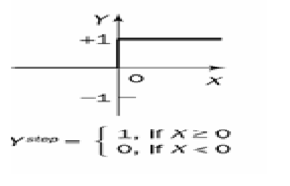
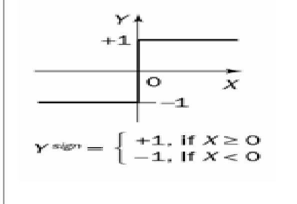
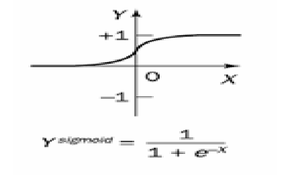
Linear function	Step function	Sign function	Sigmoid function
			
The function in which the output image is similar to the input and provides multiple and unlimited ratings **[34]**.	This function identifies the neuron output between 0 and 1	This function uses classification and pattern recognition (in which the output value is limited between -1 and 1) [35].	It is a non-linear function and its output is confined between 0 and 1 and is one of the most important functions used in artificial neuralnetworks for easy calculation [32]

Table 2: Activation function used in synthetic neural networks

➢ **Output Layer:** this layer transmits the output signal to subsequent neurons thereby transforming into an input signal for those cells.

4.1.3 Mechanical learning in a multi-layered model (MLP)

Multiple artificial neural networks comprise numerous units of cognitive elements known as perceptron elements interconnected by bonds these bonds connect the output of specific neurons to the input of adjacent neurons thereby establishing a three-layer neural network architecture as depicted in **Figure 10** [36]:

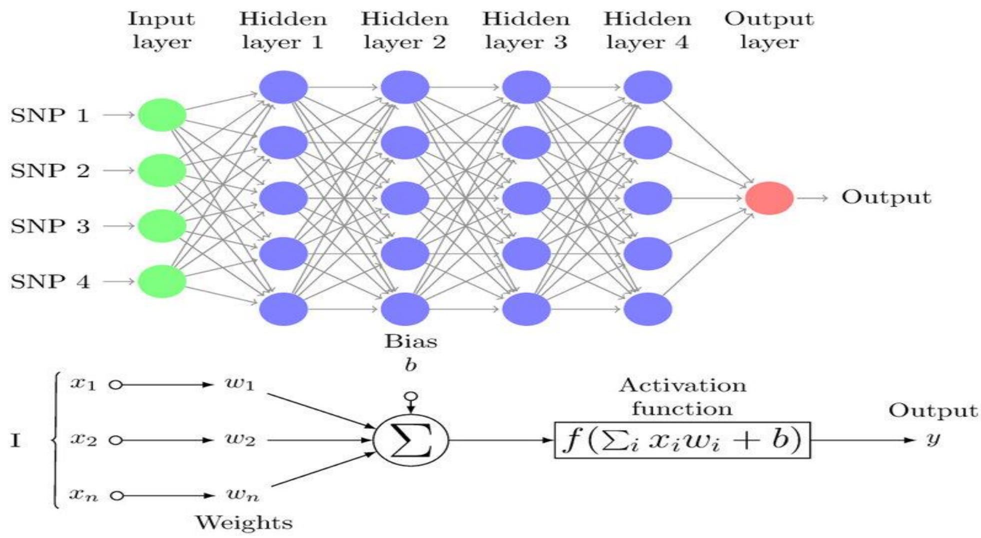


Figure 10: Multi-layered neural networks

- **Input Layer:** this layer serves as the entry point for the network's input values, which are distributed among its computational elements.
- **Hidden Layer:** this layer or group of layers is responsible for refining the input data and applying boundary values to activate computational elements.
- **Output Layer:** positioned at the network's exit point, this layer produces real output values both actual (current) outputs and desired outputs post-training.

4.2 Convolutional neural networks (CNNs)

4.2.1 CNN Architectural Evolution

Artificial intelligence (AI) has made tremendous advances in many fields, most notably computer vision, which aims to make machine vision similar to human vision by detecting, recognizing, and classifying images, analyzing videos, and processing natural language. Convolutional neural networks have been the most prominent means of building computer vision algorithms due to their ability to recognize and classify image features thanks [37].

Convolutional neural networks emerged as part of deep learning in the early 1980s, when Japanese scientist Konishi developed a Neocognitron model that was used for image recognition [38]. In the early 1990s, French researcher Yann Lucan developed the Linet model, the first model to use a CNN to recognize handwritten numbers [39]. 2012 was a turning point for the development of CNNs with the emergence of AlexNet network, which outperformed traditional image recognition models in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [40]. Since then, convolutional neural networks have enjoyed tremendous success, attracting the attention of many researchers who have sought to explore and develop many new architectures for these networks as shows in **Table 3** :

Year	CNN Architecture	Description
2014	VGGNet	Developed at the University of Oxford by the geometry group. It consists of 16-19 layers with 3x3 small convolutional filters and maximum pooling layers. It is characterized by its simplicity and depth as it relies on stacking convolutional layers with small filters to achieve high resolution and effectively capture features.
	GoogleNet (Inception)	Developed by Google researchers, they sought to minimize information overhead while maintaining representational power by using Inception modules that feature multiple parallel convolutional processes with different kernel sizes within the same layer.
2015	ResNet	Developed as a solution to the vanishing gradient issue in deep networks, it utilizes residual connections that allow residual functions to be learned instead of direct learning of the underlying maps, allowing for the training of very deep networks with hundreds of layers.
2016	DenseNet	It relies on dense communication between layers, where the outputs of all previous layers are merged into each layer, resulting in improved parameter utilization and performance.
	Xception	Is an extension of the Inception architecture, where it has been replaced by the Xception modules, which are characterized by deep and linear finite transformations, allowing for increased model accuracy.

2017	MobileNet	Developed by Google researchers, its architecture is based on narrow deep filters, achieving computational efficiency and high-speed performance.
2019	EfficentNet	Is based on the principle of convolutional mesh modeling, where models are optimized in parallel across model size, depth, and width to achieve consistently better performance.

Table 3: CNN Architecture development

Convolutional neural networks have been used in many computer vision tasks including object identification, image classification and segmentation, object search in images and videos, face recognition, and other applications that involve simulate pattern recognition [41].

4.2.2 CNN Layers

The layers of convolutional neural networks (CNNs) play an important role in the network's ability to extract, process, and interpret features from visual data. These include the Input Layer, Convolutional Layer, Pooling layer, Activation Layer, Fully Connected Layer, and SoftMax/Logistics Layer. As described in the **Figure 11** [42].

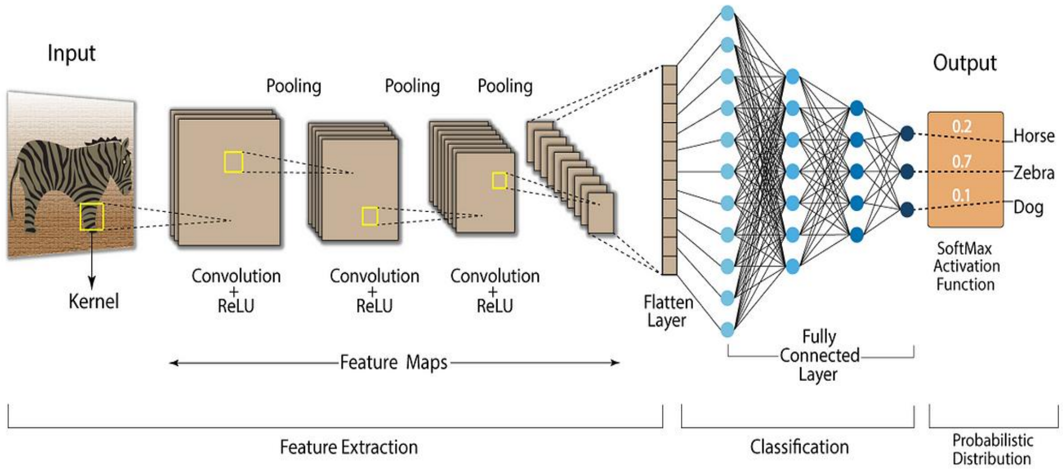


Figure 11: Example showing the CNN layers

a. The Input Layer

Is the first step in the model, representing the integrated visual data in the image as pixels arranged in 3D matrices representing the height, width, and depth of the image. Next, the data is reorganized into a single 2D matrix and converted into a vector representation. Each entry in the data is represented by a specific pixel value, which is the final input to the input layer. This process enables the model to understand the input data and start extracting features from it using the following layers in the neural network.

b. Convolutional Layer

Named for its convolution operation, this layer employs a process known as convolution to extract features or properties from images and input data using a kernel, which is a small matrix applied to the entire data by sliding, as shown in **Figure 12** [43]. The output is calculated by multiplying the data values by the kernel, resulting in the generation of a feature map. This process aids in extracting local patterns such as edges and angles, thereby augmenting the network's capacity to discern critical features within the data.

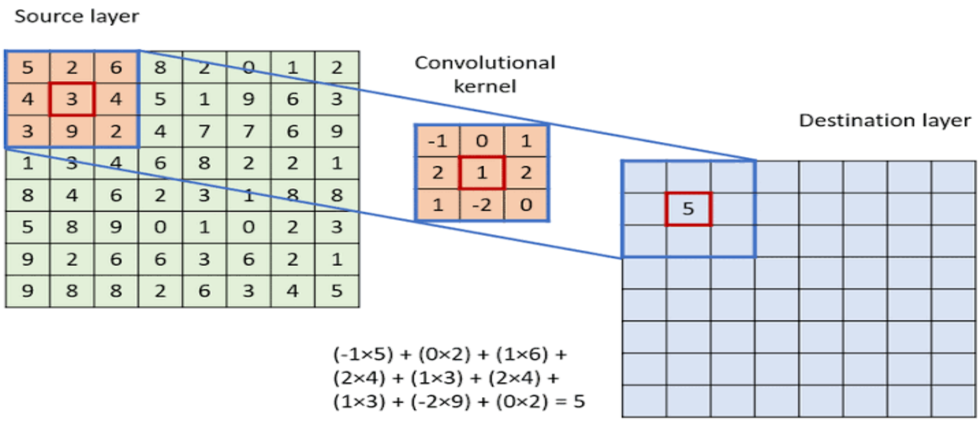


Figure 12: Convolution operation

c. Pooling layer

The pooling layer is employed to reduce the dimensions and spatial size of the input image typically following convolutional operations through techniques such as max pooling or average pooling as described in the **Figure 13** [44]. This process involves selecting either the maximum or average value within a specified range in the feature map, effectively encoding information while mitigating computational complexity which minimizes training time and prevents overfitting.

- **Maximum Pooling:** is the most commonly used method, a kernel of a certain size (n x n) traverses the matrix, selecting the maximum value from the data within its span then it positioned in the resulting matrix, this process reduces the spatial dimensionality of the representation which, allows the extraction of the important patterns and eliminating unimportant details.
- **Average Pooling:** it minimizes the size of the data and increases the efficiency of handling it. By passing a window of a certain size (n x n) through the input data, the values at each point are averaged and placed in a specific location in the resulting matrix. This process is repeated across each channel of the original matrix, ultimately resulting in an output matrix that contains the same number of channels but with a smaller size.

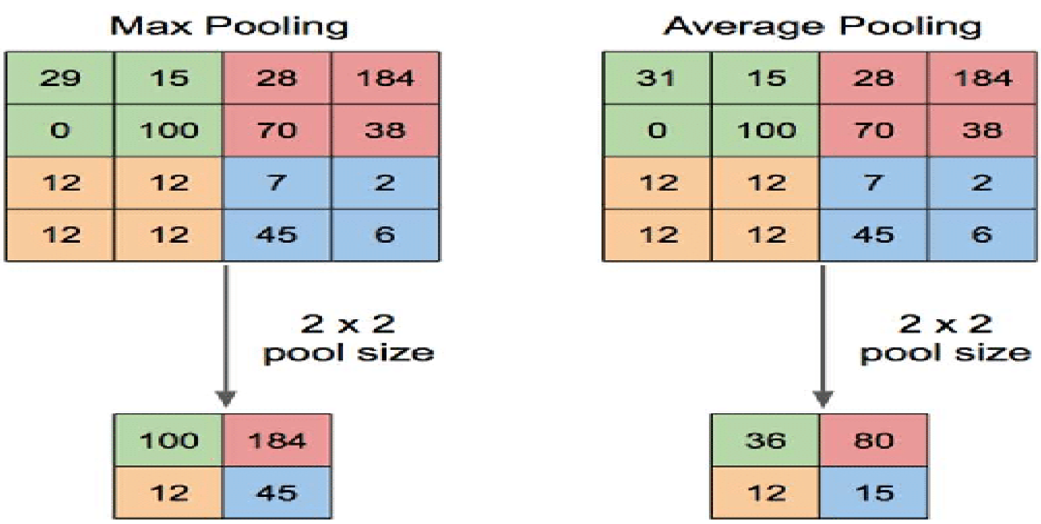


Figure 13: Pooling with a 2x2 filter

d. Activation layer

The activation function is a particular mathematical function that determines the filter's output value. In CNN, the most often used function for feature extraction is the Rectified Linear Unit (ReLU) function. These functions seek to find values inside a specific range, such as -1 to 1 or 0 to 1 depending on the function type being used. There are two primary categories of activation functions [45]:

➢ **Linear Activation Function:** the output is generated by multiplying the inputs by a constant (weight of each nerve cell) using a linear transformation between the inputs and outputs. This function can be better than nonlinear functions because it only gives yes or no answers, without complications.

➢ **Non-linear Activation Functions:** it facilitates the learning and modelling of complex data, including images, video, and audio, by allowing the model to construct intricate graphs between inputs and outputs. These functions are necessary to effectively and accurately represent non-linear or high-dimensional data.

e. Fully connected layer

This layer encompasses weights, biases and neurons facilitating connections between neurons from one layer to those of another layer it plays a crucial role in image classification tasks by combining features extracted from previous layers to classify images into different categories.

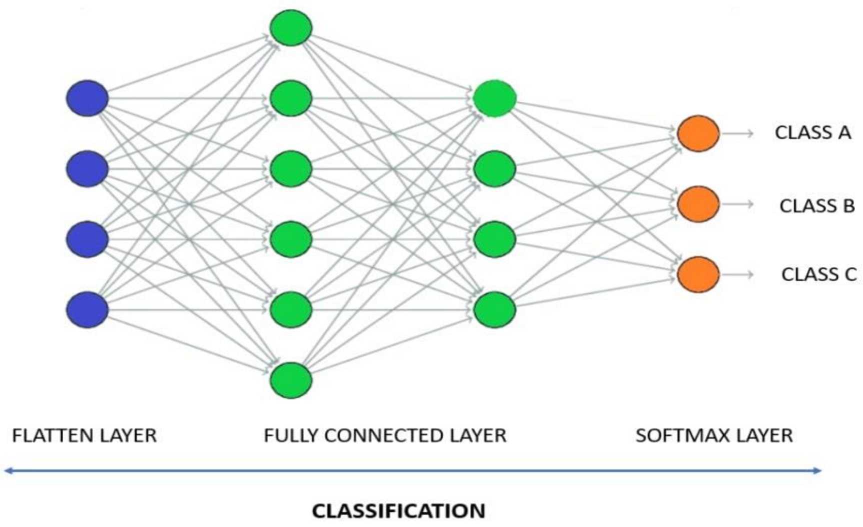


Figure 14: Fully connected layer

f. SoftMax/Logistics Layer

Positioned at the end of the fully connected layer this layer is pivotal in classification tasks the logistic function is employed for binary classification while SoftMax is utilized for multi-class classification assigning probabilities to each class output.

g. Output Layer

The last layer in CNN network where complex calculations are compiled. Based on the features extracted from the previous layers, the network's predictions are formulated, assigning probabilities to different classes, to characterize the most likely outcome.

4.2.3 Technological Advancements in CNNs

Convolutional neural networks' architecture has seen tremendous advancements in recent years, improving its effectiveness in a variety of computer vision tasks. These developments include novel techniques and approaches that boost CNN effectiveness, precision, and capacity for generalization. Among the significant developments [45]:

A. Depth and width

- CNN Based on Depth: based on deep networks with additional graphics and feature sequences approximating the objective functions, where depth plays a crucial role in improving learning ability. ResNet (Residual Network) is one of CNN's depth-based architectures.
- CNN Based on Width: focuses on width to develop effective learning strategies, making sure that networks are wide enough to maintain overall approximation properties while increasing in depth.

B. Utilization of Multiple Paths

- Address challenges such as disappearing gradients by incorporating shortcut connections or multiple paths, allowing for personalized information flow between layers and improving network training.

C. Exploitation-Based Feature-Map CNNs

- Selects relevant feature maps to enhance network generalization and performance, dynamically adjusting the weights associated with the cores. DenseNet is an example of a convolutional neural architecture that utilizes this technology.

D. Spatial Exploitation-Based CNNs

- Utilizes various filter sizes to extract fine and coarse information from the input data, boosting performance by considering the perimeter of internal pixels. Spatial Transformer Networks (STNs) are one of the modules of spatial exploitation-based CNNs.

4.3 Recurrent neural networks (RNNs)

4.3.1 Definition of Recurrent Neural Networks (RNNs)

Constitute a category of artificial neural networks adept at processing sequential or time-varying data distinguished by their capacity to preserve past information for subsequent analysis. RNNs integrate specialized memory components within each neuron this augmentation empowers the network to discern temporal contexts within the data. RNNs find extensive utility across a spectrum of applications, including natural language analysis and automated translation their proficiency in handling sequencing and temporal dynamics renders them invaluable in domains necessitating nuanced treatment of temporal data [46].

4.3.2 How RNN Works

Recurrent Neural Networks (RNNs) are designed to analyze sequential or time-varying data such as word sequences or video frames. These networks incorporate internal memory mechanisms to retain past information and analyze data in a sequential manner throughout the training process. RNNs continually adjust their weights to enhance their ability to comprehend temporal contexts and generate accurate outputs for sequences [47].

5. Deep Learning applications

- **Natural Language Processing (NLP):** deep learning methods are employed in natural language processing (NLP) to emulate human-level understanding through machine learning accommodating linguistic variations and generating appropriate responses. NLP leverages deep learning across various tasks including sentiment analysis language modeling, text classification, information retrieval, word embedding, spoken language understanding, machine translation and question-and-answer systems.

➢ **Robotics:** recent advancements in robotics owe much to the progress in artificial intelligence and deep learning artificial intelligence enables robots to perceive and interact with their environment, fostering their potential as human assistants in various domains.

➢ **Virtual Assistants:** virtual assistants utilize deep learning algorithms to enhance their understanding of user preferences ranging from dining preferences to favorite music genres these assistants leverage natural language processing capabilities to interpret user commands effectively furthermore, they offer features such as speech-to-text conversion, note-taking and appointment scheduling, enhancing user convenience and productivity.

➢ **Industrial Automation:** deep learning technologies are utilized in industrial automation to bolster worker safety in settings such as factories and warehouses these systems autonomously detect and respond to potential hazards including the presence of workers or objects in proximity to machinery.

➢ **Healthcare:** deep learning facilitates healthcare applications by leveraging data collected from smartphones and wearable devices medical tools analyze this data, including movement patterns to identify potential health issues computer vision software tracks patient movements to predict falls and changes in mental state, while deep learning algorithms aid in diagnosing conditions such as skin cancer from medical images.

➢ **Fraud Detection:** with the increasing digitization of banking transactions, the risk of digital fraud rises. deep learning techniques play a crucial role in mitigating this risk by rapidly detecting fraudulent activities, thereby safeguarding financial systems and transactions [20].

6. The difference between deep learning and machine learning

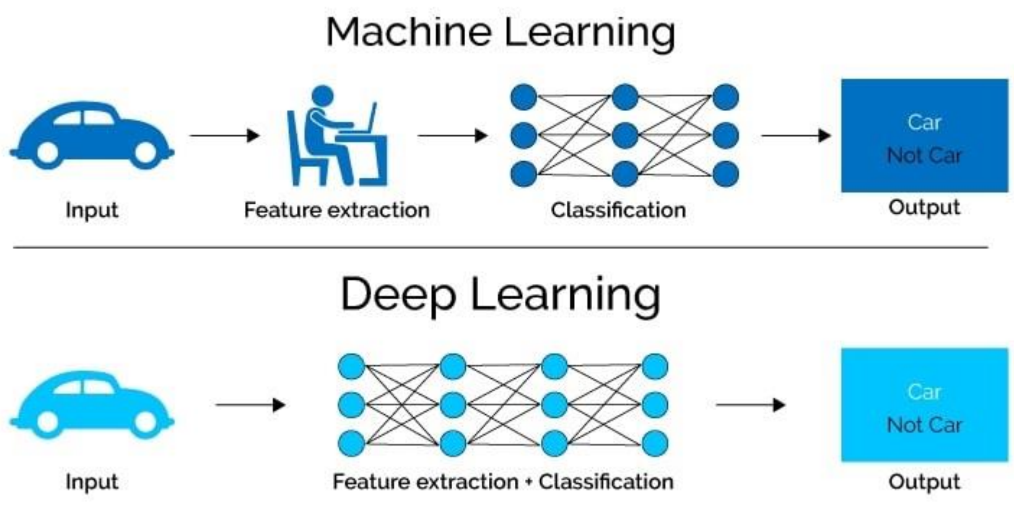


Figure 15: Machine learning vs deep learning [48]

One of the principal distinctions between deep learning and machine learning lies in their respective performance characteristics relative to the quantity of training data available, deep learning systems typically require a substantial volume of training examples to yield satisfactory results insufficient training data may hinder performance conversely, machine learning algorithms can exhibit favorable outcomes even when trained on a relatively small dataset furthermore, the implementation of deep learning necessitates advanced hardware infrastructure whereas machine learning techniques can be deployed on both low-power devices and conventional computing systems [20].

The fundamental distinction highlighting the potency of deep learning over machine learning lies in the automatic feature extraction capability inherent in these algorithms in essence, a comparison between machine learning and deep learning is summarized in **Table 4**:

Machine Learning	**Deep Learning**
A subset of AI	A subset of machine learning
Can train on smaller data sets	Requires large amounts of data
Requires more human intervention to correct and learn	Learns on its own from environment and past mistakes
Shorter training and lower accuracy	Longer training and higher accuracy
Makes simple, linear correlations	Makes non-linear, complex correlations
Can train on a CPU (central processing unit)	Needs a specialized GPU (graphics processing unit) to train

Table 4: Comparison between machine learning and deep learning

7. Conclusion

Deep learning facilitates the extraction of meaningful correlations from vast datasets and the interpretation of unstructured data. This process involves the integration of mathematical algorithms with extensive data and robust hardware infrastructure to derive actionable insights. through this methodology, digital data can be automatically extracted, categorized and analyzed, while deep learning has been in existence for several years, its widespread adoption has surged notably in the past three to four years. this trend can be attributed to various factors, including advancements in hardware resources, the refinement of sophisticated algorithms and the optimization of neural networks. It is important to note that deep learning represents an evolution of the traditional approach of artificial neural networks rather than an entirely novel concept.

The advent of deep learning technologies has revolutionized remote sensing. Which has enhanced the ability to process and analyze massive amounts of remote sensing data typically captured by satellites or aircraft, leading to more accurate and actionable insights.

Chapter 3

Deep Learning and Remote Sensing

1. Introduction

Modern technology and rapid advancements in artificial intelligence and remote sensing are pivotal elements that have redefined the way we perceive the world around us, remote sensing using deep learning emerges as an advanced innovation that combines data collection efficiency and machine intelligence to achieve significant progress in the analysis of land resources allocation and understanding and identify areas suitable for development, agriculture, conservation, or infrastructure projects [49].

Through the advanced techniques of remote sensing and, the process of data collection has been revolutionized to a very high degree to support information extraction from videos, images and cloud datasets. This advancement has significantly expanded the scope of data collection, allowing comprehensive assessments of Earth's surface on a large scale [50]. The proliferation of Unmanned Aerial Vehicles and aerial imagining, is the main engine of this revolution. UAVs, comprising hybrid models, multirotor, and fixed-wing, have emerged as essential tools in remote sensing due to their accessibility, cost-effectiveness, and robust operational capabilities in swiftly capturing high-resolution imagery [51].

Deep learning reflects the use of artificial neural models to extract information from data derived from unmanned aerial vehicles (UAVs) which enable the extraction of valuable insights from this imagery particularly through semantic segmentation techniques, by segmenting the images into meaningful regions corresponding to different land features. In this chapter we provide a comprehensive overview of remote sensing using deep learning techniques. we explore the pivotal role of deep learning in enhancing data collection, analysis, and information extraction from unmanned aerial vehicles (UAVs).

2. Remote sensing using deep learning

Remote sensing plays a vital role in data collection relying on a wide range of technologies such as satellites and unmanned aerial vehicles (UAVs). The analysis and extraction of data are facilitated through the use of deep learning a type of artificial intelligence technique that relies on the formation of deep neural networks, this connection between remote sensing and deep learning enables a deeper understanding and better extraction of information from incoming data contributing to improved analysis and comprehension.

2.1 Deep learning for feature extraction

Analyzing remote sensing data using traditional methods, which rely on manual feature engineering by domain experts, is considered one of the most challenging tasks due to its complex patterns and structures, as well as the large size of the sensed images (multispectral or hyperspectral) [52]. However, the sensors use such as single-lens RGB cameras and RGB-D sensors [53], and with the application of deep learning techniques, interpreting these images has become easier. The deep neural networks are used to effectively and automatically extract features from remote sensing data. These networks consist of several sequential layers, where raw data is used as input to the network, and the different layers learn progressively deeper representations of the data. One common method for feature extraction using deep neural networks is to use convolutional layers. These layers apply linear transformations to small parts of the input data, allowing the network to identify local patterns and features in the sensed images such as edges, corners, and patterned textures [54] Subsequently, the deep neural network follows other layers such as pooling layers, which reduce the dimensions and size of the extracted data, and fully connected layers, which transform features into more widespread and usable representations.

3. UAVs as Remote Sensing Platform

Unmanned Aerial Vehicles (UAVs), commonly known as drones represent a category of aircraft that operate without an onboard human pilot and are remotely controlled diverse in design and functionality. UAVs encompass a wide range of sizes and types , **Figure 16** [55] from hand-held drones to large long-range vehicles, common UAV configurations include multirotor aircraft such as quadcopters or octocopters as well as fixed-wing aircraft equsipped with various sensors and systems UAVs are capable of executing missions such as aerial imaging, environmental monitoring and geospatial surveying. the versatility and autonomy of these vehicles make them valuable tools in both military and civilian applications. It is noteworthy that the specific categorization and nomenclature of UAVs may vary encompassing terms like Remotely Piloted Aircraft (RPA), Remotely Piloted Aircraft System (RPAS) and Unmanned Aircraft System (UAS), each reflecting nuanced distinctions in their operational structures and components [56].

Figure 16: UAV Types

3.1 Deep learning application on UAVs remote sensing data

Recent advancements in technology and artificial intelligence demonstrate a positive interaction across various fields, including the utilization of unmanned aerial vehicles (UAVs) and deep learning techniques. This integration presents an intriguing and innovative domain where the deep learning models can be trained to perform a variety of tasks [57]:

- **Classification and identification of objects:** deep neural networks have the capability to automatically categorize different land cover types (like forests, urban areas, and water bodies) from UAVs images. Moreover, object detection models can identify specific features (such as buildings, roads, and vehicles) within large-scale images.
- **Semantic segmentation:** deep learning models can segment images into meaningful regions, which enables the understanding of the spatial distribution of various land cover classes. For example, semantic segmentation can differentiate between crops, forests, and bare soil in agricultural landscapes.

- **Change detection:** deep learning algorithms can identify changes over time by comparing UAVs images captured on different dates. Applications include the monitoring of deforestation, urbanization, and natural disasters.
- **Ultra-high resolution and noise reduction:** deep neural networks can enhance the spatial resolution of low-resolution drones' images.

4. Deep Learning and UAVs for semantic segmentation

Semantic segmentation is one of the most important tasks in computer vision, aiming to categorize each pixel in an image into certain classes (flower, person, road, or car). This task provides a more detailed understanding of the content of images, compared to classification based on object recognition [58]. Semantic segmentation has been applied in many fields such as medical imaging, satellite image analysis, and drone imagery [59].

Several deep learning architectures have been proposed for semantic segmentation, including Full Convolutional Networks (FCNs), U-Net, DeepLab, and SegNet. These architectures have made significant progress compared to traditional methods that rely on clustering pixels based on wavelength. Although deep learning methods have been successful in developing semantic segmentation, they still suffer from challenges such as layer imbalances, disparate scales, and fuzzy boundaries between layers.

4.1 Advanced Techniques in Semantic Segmentation

There are many techniques used in semantic segmentation that rely on several effective strategies and architectures that improve the accuracy of image partition and the development of segmentation models.

4.1.1 Fully Convolutional Networks (FCNs)

FCNs are a type of convolutional neural network, characterized by the ability to preserve spatial information throughout the network, allowing for dense and accurate per-pixel projections. Their design allows them to take advantage of large-scale image processing due to a series of transformation layers followed by oversampling layers. The transform layers capture spatial information at different scales, while the oversampling layers increase the spatial resolution of the feature maps to produce the final segmentation map [59].

The output size of FCNs depends on the size of the input rather than always producing a fixed-size output [60]. Thus, these types of networks are commonly used in semantic segmentation tasks for object detection, scene understanding, medical image analysis, and satellite image interpretation. Unlike CNNs that produce output maps with the same spatial dimensions as the input data, fully connected layers are replaced by convolutional layers to retain spatial information, making them important in image classification tasks.

4.1.2 Encoder-Decoder Architectures

Encoder and decoder architecture have proven to be effective in many tasks, especially in semantic hashing tasks, due to their unique architecture similar to traditional CNNs. The network is divided into two main parts: Encoder, which is used to extract sequential features from the internal image, and Decoder, which is used to convert these features into detailed results or predictions. These networks also benefit from interconnectivity, which facilitates the flow of information across the different layers of the network. Many neural networks follow this architecture, most notably U-Net [61] and SegNet [62].

4.2 Methodologies in land segmentation

This process is a crucial step in achieving high accuracy in segmentation. The corresponding details can be presented in the table below:

Training Stage	Details
Data Pre-processing	satellite images and aerial photographs are pre-processed to isolate land regions while excluding non-land features such as buildings, roads, water bodies, and vegetation. This segmentation approach enhances the dataset by focusing exclusively on land-related features, optimizing it for subsequent deep-learning tasks such as land segmentation.
Architecture selection	The dataset is divided into training and testing sets, appropriate segmentation network architectures, such as U-Net or DeepLab, are selected based on their ability to capture fine-grained spatial information. adjustments in architecture complexity, such as the number of layers and skip connections, are made to accommodate variations in land features and image resolutions.
Loss Function Definition	The selection of a suitable loss function, such as cross-entropy loss or Dice loss, is essential for measuring the discrepancy between predicted and ground truth segmentation masks during training.
Training Process	Data is fed through the segmentation network, and the model iteratively learns to generate pixel-level land segmentation masks. The training process involves computing the loss between predicted and ground truth masks and updating the model parameters using optimization techniques like stochastic gradient descent (SGD) or Adam.

Table 5 : Methodologies of Models training

5. Related Work

In recent years, semantic segmentation of UAV data for remote sensing has been increasingly popular, resulting in a plethora of applied approaches and research across multiple fields. In this section, we aim to review some of the methodologies and publications related to semantic segmentation using deep learning techniques on UAVs data.

According to existing literature traditional approaches often encounter obstacles such as parameter setups and algorithmic complexities [63]. For instance Alhichri et al. [64] paper introduced an automated system that employs VGG16 for feature extraction and SVM for vehicle classification in UAV imagery, while Moranduzzo et al. [65] proposed a vehicle detection method based on Histogram of Oriented Gradients (HOG) from UAV images. Additionally, recent research Micheal et al. [66] Focused on exploring learning techniques for object detection and tracking, in UAV data, whereas Yao et al. [67] investigated the use of fully convolutional neural networks for semantic segmentation in video frames. In the domain of aerial scene segmentation, a U-Net model was utilized in a study to generate high-resolution segmentation masks by linking encoder-decoder characteristics. Furthermore, research endeavour's by Ahram et al. [68] have explored the development of combined segmentation networks (CSNs) customized for UAV datasets and established baseline models using

DeepLabv3+ with modified structures, along with conducting experiments to assess the impact of data augmentation and model adjustments on improving results [69]. Shaar et al. [70] presented a method for segmenting satellite images in sensing to recognize aircraft that combines two models. To increase feature localization and segmentation accuracy, they used the U-Net design with skip connections. In study Kentsch et al. [71] employed transfer learning methods along, with an MLP classifier to analyze forest images. Richmond et al. [72] applied Gray Wolf Optimization (GWO) to enhance the modified U-Net model with ResNet 34 (UResNet 34) for building segmentation resulting in increased accuracy and convergence. Moreover, Yiwen et al. [73] suggested a deep learning method that combines FCN and Conv LSTM to segment UAV video frames. Lucas et al. [74] investigated the application of advanced machine learning techniques, in identifying citrus trees from images captured by drones, they contrasted models including DDCN, DeepLabV3+, U-Net, FCN, and SegNet, with DDCN exhibiting the most effective performance. Furthermore, Jenssen et al. [75] investigated pixel-wise segmentation techniques as well as correction-based classification algorithms for segmenting small objects in satellite pictures using deep convolutional neural networks. In a related study Jeon et al. [76] investigated how to track marine grass habitats in environments using UAV imagery and deep learning algorithms. They examined normalization methods and segmentation models, like U-Net, SegNet, PSPNet and DeepLab v3+. Dongzi et al. [77] introduced a simplified deep learning technique for lychee tree segmentation using UAV-collected satellite images, employing DJI P4 Multispectral equipment and DJI Terra software for image analysis and segmentation, where the methodology included image partitioning and a straightforward guidance process to facilitate learning and training. They also modified the YOLACT network for semantic segmentation with computational complexity reduction techniques, applied to high-resolution image levels to preserve crucial details during analysis. Shrikrishna et al. [78] proposed a modified U-Net network and compared it with SegNet and the original U-Net for plant semantic segmentation.

6. Conclusion

The integration of remote sensing, deep learning, and the utilization of unmanned aerial vehicles (UAVs) has significantly advanced the field semantic segmentation. This synergy has led to a notable improvement in data accuracy particularly in the precise delineation of land features and resource distribution.

In this chapter, we provided an overview of remote sensing using deep learning techniques, which have proven their efficiency in analyzing and extracting features from sensed data from UAVs. These techniques have facilitated many applications, such as object classification, change detection, and semantic segmentation, which are the focus of this thesis. We also presented a review of architectures that provide robust solutions for land feature segmentation from UAV images, enabling detailed insights into land cover patterns, through the use of Fully Convolutional Networks (FCNs) and Encoder-Decoder.

To further enhance segmentation accuracy and detail, we employ a U-Net integrated with ResNet-34 for the semantic segmentation task. This combination leverages the strengths of both architectures, providing a powerful tool for precise and detailed analysis of land features.

Chapter 4

Methodology

1. Introduction

Semantic segmentation constitutes a fundamental tool in environmental analysis and comprehension of its dynamics serving to partition images into precise and detailed meaningful regions. Modern methodologies notably deep learning leveraging neural networks assume a pivotal role in achieving this objective by automating the detection of changes with remarkable accuracy thus augmenting the quality and precision of insights concerning land usage and alterations [79]. In this chapter we present our approach to semantic segmentation of UAV remote sensing data using a fusion of the U-Net architecture with ResNet34 as backbone. we aim to furnish a comprehensive comprehension of the collaborative procedures entailed in the development and execution of our model spanning from the phase of data acquisition to model evaluation.

2. U-Net Model Architecture

We employed the U-Net architecture, a fully convolutional neural network designed for semantic segmentation tasks. U-Net incorporates skip connections between encoder and decoder blocks to preserve spatial information crucial for accurate segmentation. This symmetrical architecture facilitates efficient information flow and enables the model to capture both local and global contexts, it comprises several fundamental components [61]:

- **Encoder**

The process begins by feeding inputs into the encoder, which comprises a set of convolutional layers that extract important features from the image. These layers utilize activation functions like ReLu to introduce non-linearity, in addition to pooling layers. The number of these layers is determined based on the size of the image and the complexity of the model.

- **Bottle Neck**

The bottle neck consists of a convolutional layer that uses a large kernel size to enhance the extraction of deep features, followed by an activation function. Short connections enable information from the upper layers of the up-path to directly access information from the lower layers of the down-path, facilitating effective information transfer without significant loss.

- **Decoder**

The decoder includes upsampling layers aimed at enhancing dimensionality and reinstating high resolution in the image. It depends on concatenation processes to merge the information extracted from the encoder. Additionally, it comprises convolutional layers used to improve image quality and extract more precise information, followed by an activation function to activate the extracted information.

- **Final Layer**

At the final layer level, the final result of the model is produced. It consists of a convolutional layer followed by a SoftMax activation function, which generates a mask that identifies the class for each pixel in the image.

Building upon the comprehensive information preparation outlined earlier, we chose the U-Net architecture for its proven effectiveness in semantic segmentation, particularly its ability to handle complex spatial relationships and maintain fine details. Additionally, its adaptability and balance between computational efficiency and segmentation accuracy make it an ideal choice for our objectives.

3. Integration of ResNet34 with U-Net

ResNet, also known as residual network, is a type of convolutional neural network (CNN) that was introduced by Kaiming He et al. [80] in 2015. ResNet was designed to address the issue of vanishing gradients, which had been a major challenge in the development of deep networks [81].

3.1 ResNet Architecture

ResNet models are composed of convolutional layers and residual blocks. A residual block consists of several stacked convolutional layers that extracts features from images by applying a set of filters to the input feature maps, the output is then normalized to stabilize and accelerate the training process. Each convolutional layer is followed by an activation function (ReLu) to introduce non-linearity into the network, to enhances its ability to learn complex patterns. It consists of skip connection from the first layer to the last layer of the block. These skip connections allow gradients to flow directly during training, as they skip one or more layers [80] . This enables the network to learn residual mappings instead of trying to learn the desired output directly, which has facilitated the training of deep networks. Tracking the residual blocks with pooling layers to reduces the dimensionality and spatial size of the input image, followed by a fully connected layer and output layer which transforms the final results into the desired predictions.

ResNet architectures typically come in several variants, such as ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152. The numbers in these names correspond to the total number of layers in the network. ResNet has proven its effectiveness in many computer vision tasks, such as image classification, object detection, and semantic segmentation.

In this work, we chose ResNet34 and integrated it with U-Net to enhance semantic segmentation in aerial data. ResNet34 stands out for its ability to efficiently learn features and capture complex patterns, while its skip connections ensure the preservation of spatial information, crucial in the segmentation process. There's a balance achieved between model performance and computational cost, as ResNet34 trains faster than deeper architectures like ResNet50 and requires less memory. The integration with U-Net enhances the utilization of powerful feature extraction capabilities. These features are enhanced by utilizing pre-trained weights sourced from ImageNet.

3.2 ResNet34 Architecture

ResNet-34 consists of 34 layers including convolutional layers and residual blocks, as shown in **Figure 17** [82].

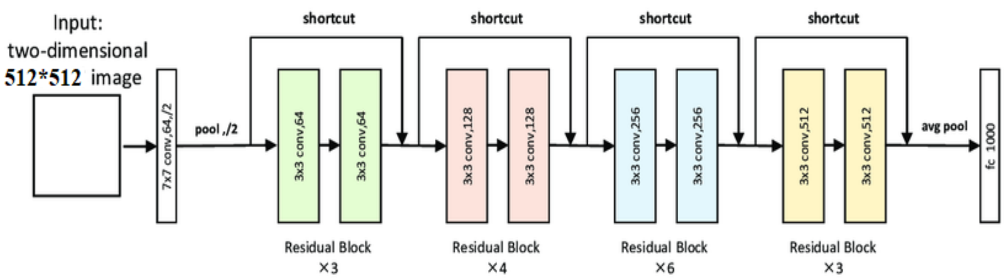


Figure 17: ResNet34 Architecture

The initial convolutional layer applies 64 filters of size 7 x 7 and stride 2 during convolution, meaning that the filters move two pixels at a time horizontally and vertically across the input image, followed by batch normalization and ReLu activation. This layer produces feature maps with low spatial dimensionality and non-linear network [83].

ResNet34 consists of 1 convolutional layer succeeded by 3 residual blocks, each block consists of two convolutional layers with 64 filters of size 3x3, using a stride of 1. By using skip connections between blocks, the vanishing gradient problem is avoided, guaranteeing effective deep neural network training. After each convolutional layer, max pooling with a kernel size of 3x3 and stride 2 is applied to downsample the feature maps which are subsequently fed into fully connected layers followed by ReLu. The final output layer translates these computations into actionable predictions [84].

4. Model Implementation

The architecture employed in this work for performing the semantic segmentation of UAV remote sensing data utilizes the fusion of the U-Net framework with ResNet34, as illustrated in **Figure 18**.

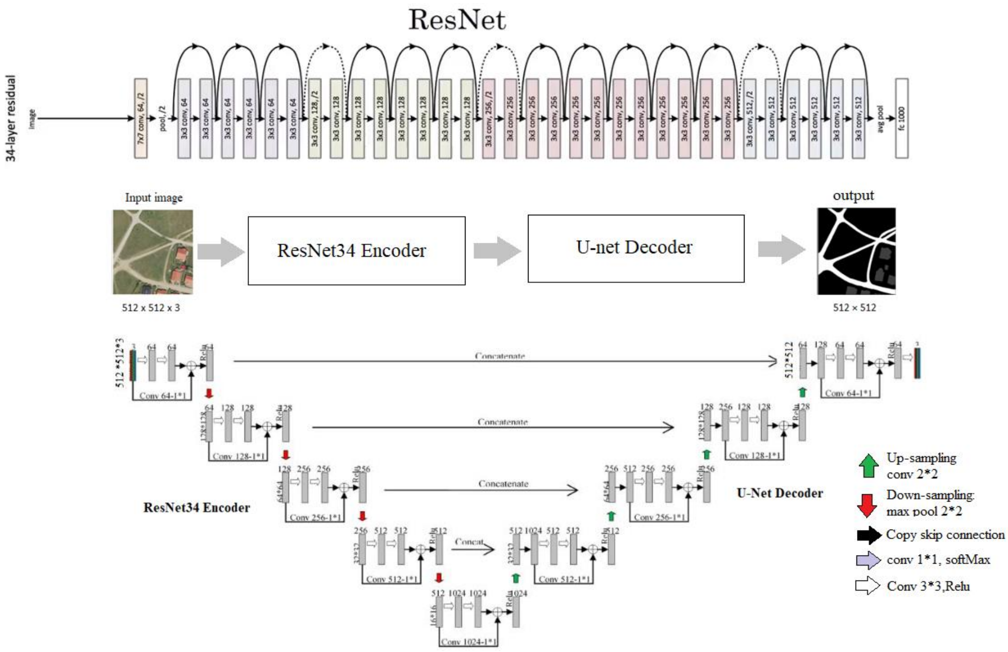


Figure 18: The used model U-Net architecture

The segmentation process begins with the input layer, where input images with a resolution of 512x512 pixels are fed into the model, serving as the initial entry point for image data.

These images are then processed through a pre-trained ResNet34 model, which functions as the encoder in the U-Net architecture. It serves as the 'eyes' of the model, where meticulously examining land cover images detect essential elements like buildings, woodlands, roads, water bodies, and background details. Similarly, aerial drone images capture a wide range of elements, such as paved areas, dirt, grass, water, vegetation, and more. the ResNet34 encoder, adhering to ResNet principles, is structured with an initial convolutional layer, multiple blocks, global average pooling, and a fully connected layer, each block features two 3 × 3 convolutional layers with ReLu activations, incorporating skip connections to link layers and enhance gradient flow for deep network training. Max-pooling aids in down-sampling for richer feature representations. Utilizing pre-trained weights, the encoder extracts discriminative features like edges and textures from UAV imagery, enabling a thorough understanding of the input data. This hierarchical feature abstraction underpins accurate semantic segmentation in subsequent model stages.

The bottleneck layer captures the most abstract features extracted by the ResNet34 encoder. These abstract features are then passed to the U-Net decoder, where max-pooling with a 2 × 2 filter is employed for down-sampling, reducing dimensions by half while doubling channel capacity. Subsequently, each decoder block incorporates a 2 × 2 up-sampling filter to recover feature map sizes, which is crucial for reconstructing detailed information and enabling precise semantic segmentation by effectively reconstructing information from the reduced feature maps.

Finally, the output layer, represented by a 1x1 convolutional layer with a SoftMax activation function, generates pixel-wise classification probabilities for each class. This layer produces a segmentation mask with the same dimensions as the input image, assigning a class label to each pixel. U-Net serves as the 'brain' of our model. It takes these detailed features and organizes them into meaningful segments, like identifying buildings, roads, woodlands, water, and other classes.

The integration of the ResNet34 encoder and the U-Net decoder forms a robust and comprehensive process for identifying features. where ResNet34's role as the 'eyes' of the model and U-Net as the 'brain,' ensures precise and detailed segmentation across a diverse range of classes and elements present in land cover and aerial drone images.

5. Evaluation Metrics

To measure the semantic segmentation performance and evaluate our models, we use:

- **Intersection over union (IoU) score:** this metric evaluates the correspondence between predicted and ground truth segmentations, elevated Iou scores signify enhanced alignment between the predicted and ground truth delineations [85].

$$\mathbf{Iou} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \qquad \textbf{(1)}$$

- **Mean Intersection over union (mIoU) score:** is the average of IoUs by calculating the intersection over union (IoU) for each class and then averaging these values across all classes [86].

$$\mathbf{Mean_{Iou}} = \frac{1}{N} \sum_{i=1}^{n} Iou_i \qquad \textbf{(2)}$$

- **Precision:** is a measure of the accuracy of positive predictions made by a model, calculated as the number of correct positive predictions (TP) divided by the total number of positive predictions (TP + FP) [87].

$$\textbf{Precesion} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

- **Recall:** also called sensitivity, is used to measure the fraction of positive patterns that are correctly classified [88].

$$\textbf{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

- **F1-score:** This metric represents the harmonic mean between recall and precision values.

$$\textbf{F1 score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

- **Accuracy:** the accuracy metric measures the ratio of correct predictions over the total number of instances evaluated [88].

$$\textbf{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{6}$$

Where: TP True Positives, TN True Negatives, FP False Positives, FN False Negatives, N class number.

6. Dataset

Four distinct UAV datasets were utilized in this study as shown in **Table 6**: Land cover.ai [89], Aerial semantic segmentation [90], AeroScapes dataset [91], and UAVid [92]. Each dataset was pre-processed, ensuring data cleanliness and consistency. Data augmentation techniques such as rotation, scaling, and flipping were applied to enhance the diversity of training samples.

Datasets	Classes	Source	Size (px)	Resolution (m)	Bands
Land Cover.ai	5	Aerial	512 × 512	0.25–0.5	RGB
Aerial Semantic Segmentation	24	UAV	6000 × 4000	5–30	RGB
UAVid	8	UAV	3840 × 2160	0–50	RGB
AeroScapes	12	UAV	1280 × 720	5–50	RGB

Table 6: Information about the datasets

6.1 Land Cover Dataset

We have used the LandCover.ai dataset for semantic segmentation purposes. Given its tailored focus on land cover analysis through aerial imagery, this dataset aptly aligns with our objective of automating the identification process for buildings, woodlands, water bodies, and roads. It comprises a curated collection of aerial images, spanning rural territories covering 216.27 sq. km within Poland, a nation situated in central Europe. These images collectively encompass an area of 39.51 sq. km and have been sourced from various aerial surveys conducted across different years spanning from 2015 to 2018, capturing a diverse range of spatial and temporal contexts. Aerial photography operations in Poland traditionally occur between April and September, thereby capturing a comprehensive spectrum of optical conditions, including variations in saturation solar, illumination angles, and shadow lengths.

Figure 19: Sample images from the dataset showcasing diverse environmental conditions, including different regions, seasons, times of day, weather patterns, and lighting conditions

To ensure maximal dataset diversity, we meticulously curated 41 orthophoto tiles from diverse counties spanning all regions. Each tile encompasses an approximate area of 5 square kilo meters. The dataset comprises 33 images with a 25 cm resolution (approximately 9000 × 9500 pixels) and 8 images with a 50 cm resolution (approximately 4200 × 4700 pixels). This rigorous selection process results in comprehensive coverage: 176.76 square kilo meters for the 25 cm resolution images and 39.51 square kilo meters for the 50 cm resolution images, aggregating to 216.27 square kilo meters overall [89].

To optimize the processing and training efficiency of our semantic segmentation model we initiated the data preparation phase by addressing the challenge posed by the large image sizes in our dataset **"**LandCover.ai**"**, employing a methodical approach each large image was segmented into smaller patches of size 512×512 pixels, this precise partitioning process resulted in the generation of 10674 image patches along with their corresponding mask patches, meticulously organized within the output directories, subsequently to ensure comprehensive model training and evaluation we split and shuffled the 10674 image patches into distinct train test and validation sets, the training dataset consisted of 7470 patches while the test and validation datasets comprised 1602 patches each. This deliberate partitioning facilitated a balanced distribution of data, with approximately 70% designated for training and 15% each for testing and validation. This precise data distribution facilitated thorough model assessment and validation, enhancing the generalization capabilities of the trained model.

Data augmentation enhances the model's adaptability to unseen data by diversifying the training dataset. We applied horizontal and vertical flips to vary perspectives and random rotations to simulate different orientations typical of aerial imagery, and the 'reflect' fill mode preserved spatial coherence and synchronization, ensuring consistency between input images and masks. This approach equips the model with robust representations critical for accurate predictions in semantic segmentation.

6.2 Aerial Semantic Segmentation Drone Dataset

A well-known collection released by Graz University of Technology serves a crucial purpose in facilitating the training and assessment of deep learning models tailored for semantic segmentation tasks. The primary focus of this dataset is to enhance our comprehension of urban environments.

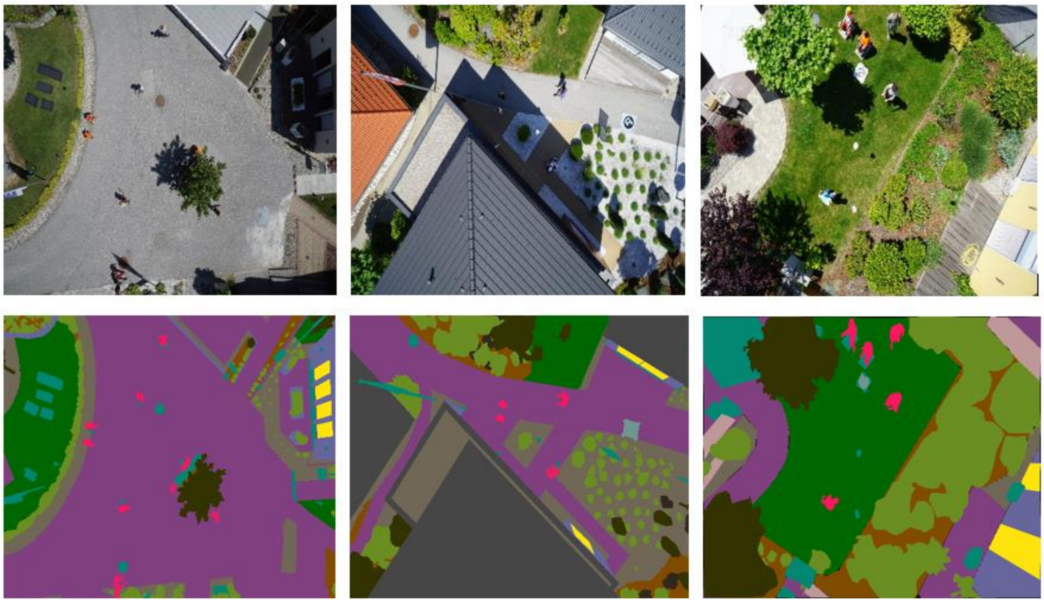


Figure 20: Sample images from the dataset with their masks

The dataset includes 400 images captured at an altitude of 5 to 30 meters above the earth's surface, taken from 20 houses in the nadir (bird's eye). Each image boasts dimensions of 6000×4000px pixels (24Mpx), meticulously paired with a corresponding mask image. These mask images, serving as label images, encode pixel-level information using numeric values representing distinct classes from 0 to 23. To ease the segmentation visualization, the dataset includes a mapping dictionary that associates each class with a specific color. **Table 7** illustrates the RGB tuple for each class as well as the class corresponding to each number in the mask image. Moreover, the dataset includes 200 private test images [90].

Class	Name	(R, G, B)	Class Counts
0	unlabeled	(0, 0, 0)	12714
1	paved-area	(128, 64, 128)	546694
2	dirt	(130, 76, 0)	66101
3	grass	(0, 102, 0)	329861
4	gravel	(112, 103, 87)	103629
5	water	(28, 42, 168)	34511
6	rocks	(48, 41, 30)	0.719
7	pool	(0, 50, 89)	38056
8	vegetation	(107, 142, 35)	94435
9	roof	(70, 70, 70)	128798
10	wall	(102, 102, 156)	39916
11	window	(254, 228, 12)	18210
12	door	(254, 148, 12)	9813

13	fence	(190, 153, 153)	13458
14	fence-pole	(153, 153, 153)	7821
15	person	(255, 22, 96)	11901
16	dog	(102, 51, 0)	7253
17	car	(9, 143, 150)	14804
18	bicycle	(119, 11, 32)	8148
19	tree	(51, 51, 0)	28498
20	bald-tree	(190, 250, 190)	20871
21	ar-marker	(112, 150, 146)	6878
22	obstacle	(2, 135, 115)	23218
23	conflicting	(255, 0, 0)	7160

Table 7: RGB tuple of each class corresponding to each number in the mask image

To efficiently handle data preprocessing and avoid memory overflow, we've implemented a data generator. This approach allows us to load the dataset in batches rather than all at once, effectively managing RAM usage. The original images in the dataset are sizable at 6000x4000 pixels, so we resize them to a more manageable size of 512x512 pixels during preprocessing.

Our dataset consists of 400 images, of which we used 320 images (80%) for the training set and the remaining 80 images (20%) for the validation set. We employ data augmentation techniques to artificially increase the dataset to prevent overfitting, given the relatively modest size of the dataset. We increase the training data fivefold, resulting in a total of 1600 images in the augmented training set and 80 images in the validation set post-augmentation. This approach optimizes memory usage ensures efficient handling of the dataset, and leverages data augmentation to enhance the robustness of our neural network model during training.

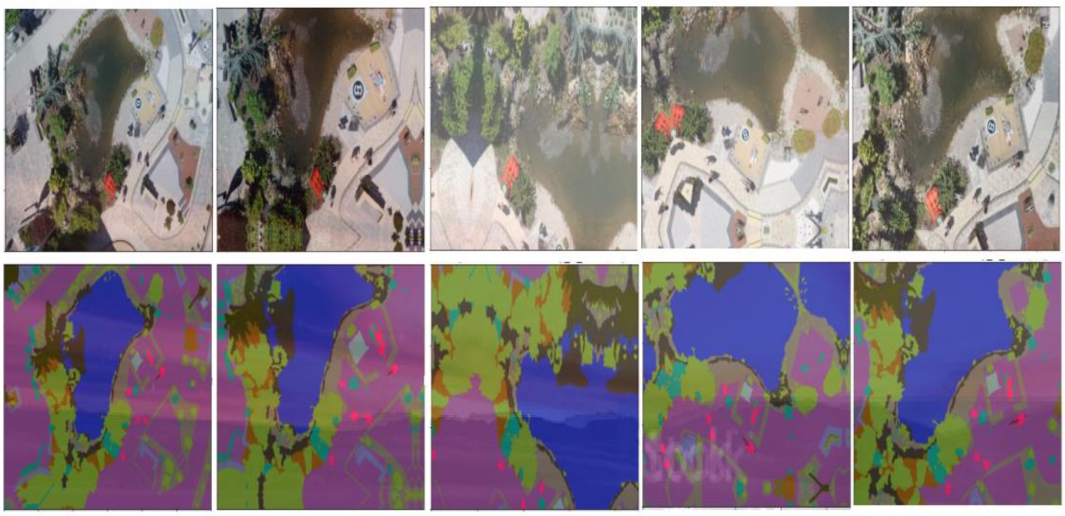


Figure 21: Samples example of augmented data

6.3 UAVid Dataset

The UAVid is a UAV benchmark dataset for semantic segmentation task, particularly in urban scenes. It was captured using a 4K resolution RGB video recorder at an altitude of approximately 50 meters.

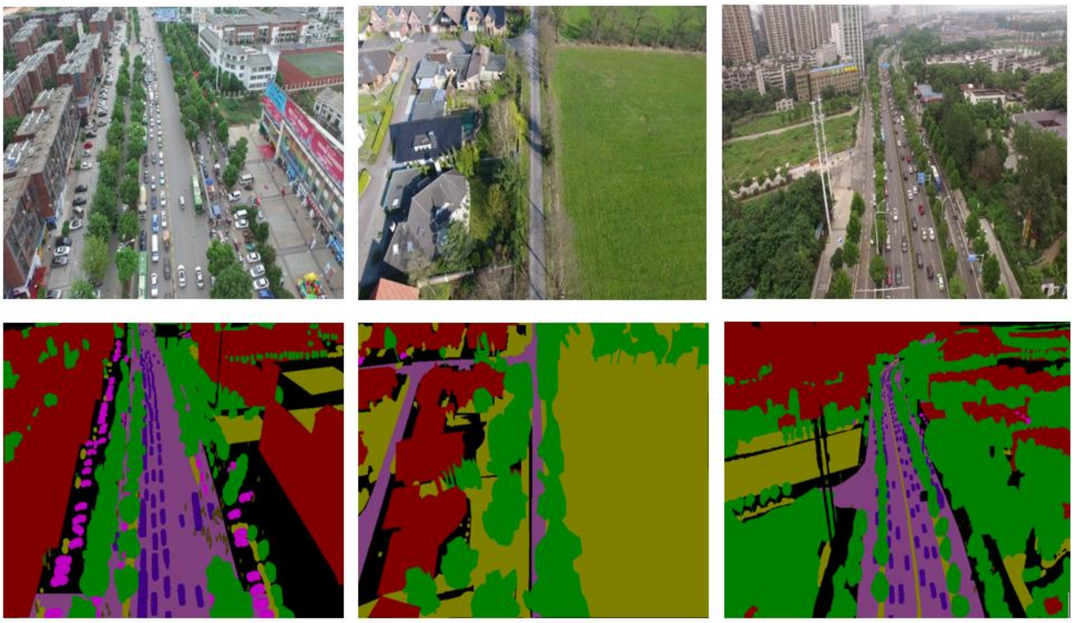


Figure 22: Sample images from the UAVid dataset with their masks

Each of the 420 images in the dataset has a resolution of 3840 ×2160 pixels and encompasses 8 classes: building, road, tree, low-vegetation, moving car, static car, background, and human. To enhance the dataset, we expanded it to 600 images, including 200 original UAV images and 400 augmented ones with shifts, noise, flips, and varied contrast. We use 600 images for training ,70 images for validation and 150 images for testing. Due to the original images hight resolution, we resized them to 512× 512 pixels [92].

6.4 AeroScapes Semantic Segmentation Dataset

The AeroScapes dataset serves as a benchmark for aerial semantic segmentation, showcasing images captured by a commercial drone at various altitudes from 5m to 50 m.

Figure 23: Close-up images from the AeroScapes dataset with their masks

Each image is sized at 1280 ×720 pixels, exhibiting significant pixel count differences across object categories as shown in **Figure 24** [91]. The dataset contains 3296 images paired with corresponding ground truth images delineating 12 classes of ground objects found in urban and suburban scenes, including background, person, bike, car, drone, boat, animal, obstacle, construction, vegetation, road, and sky. Throughout the training process, we maintain the original size, we use 2288 images for training, 654 for validation, and 327 for testing [91].

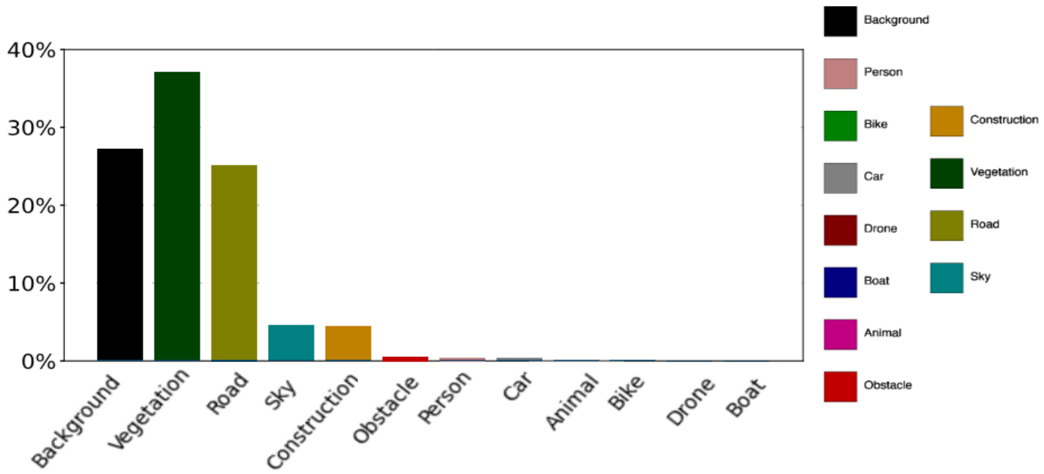


Figure 24: Pixel distribution of the AeroScapes dataset

7. Conclusion

Our goal in this chapter was to perform semantic segmentation of UAV aerial imagery based on a deep learning approach. We propose a U-Net architecture fused with ResNet34 as backbone, leveraging its robust feature extraction capabilities these features are enhanced by utilizing pre-trained weights sourced from ImageNet, to solve the vanishing gradients problem. To evaluate the effectiveness of our approach, we leverage four diverse Unmanned Aerial Vehicles datasets: LandCover.ai, Aerial Semantic Segmentation, UAVid, and AeroScapes. These datasets represent distinct scenarios and environments, providing a comprehensive assessment of the proposed approach's performance across diverse landscapes and urban setting.

To evaluate the model's performance, we present a benchmark comparison and analysis using multiple metrics such as mean Intersection over Union (mIoU), Intersection over Union (IoU), accuracy, recall, and precision. These evaluations offer insights into the efficacy and reliability of the semantic segmentation approach.

Chapter 5

Experiments and Results

1. Introduction

In the methodology chapter we outlined our approach to segmenting images using deep learning methods we detailed the specific steps we took to ensure the best possible results moving forward to this chapter, we present our findings and evaluate how well our model performed additionally, we aim to compare our results with similar research to confirm the effectiveness of our approach this comparison serves to bolster the credibility and significance of our study within the academic community.

2. Training

Training for all the datasets involved leveraging GPU resources, specifically the GPU P100, for efficient computational processing. The optimization strategies included utilizing the Adam optimizer with a learning rate of 0.0001, chosen for its adaptability to the complex loss landscape in semantic segmentation tasks. The hyperparameters, including a batch size of 16 and a random seed of 24 for reproducibility, were meticulously fine-tuned to enhance model efficacy.

For the Land Cover, UAVid and AeroScapes datasets, we employed the Dice loss function as shown in Equation **(7)** specifically tailored for semantic segmentation to effectively penalize false positives and false negatives while monitoring intersection over union (IoU) as a metric. The training spans more than 100 epochs. The model is trained to utilize the fit method on the training dataset generated by the data generator, augmented by early stopping mechanisms with patience of 10 epochs and Model Checkpoint callback functionality to safeguard optimal model weights.

$$\text{Dice loss} = 1 - \frac{2 \times \text{Intersection}(Y_{\text{true}}, Y_{\text{pred}})}{\text{Totale Pixels } Y_{(\text{true})} + \text{Totale Pixels } Y_{(\text{pred})}} \tag{7}$$

Where,

- Y_{true} : ground truth mask.
- Y_{pred} : predicted mask.
- $\text{Intersection}(Y_{\text{true}}, Y_{\text{pred}})$: calculates the number of overlapping pixels between the true and predicted masks.

For the Aerial semantic segmentation drone dataset, we employed the categorical cross-entropy loss function as shown in Equation **(8)** while monitoring accuracy as a metric. Our training spans 100 epochs, utilizing the fit method on the training dataset. This approach ensures iterative refinement of the model and optimal performance over time.

$$\text{Categorical Cross} - \text{Entropy loss} = -\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{c} y_{ij} \log(\widehat{y_{ij}}) \tag{8}$$

Where,

- N: is the number of samples in the dataset.
- C: is the number of classes.
- y_{ij} : is the ground truth probability that sample i belongs to class j.
- $\widehat{y_{ij}}$: is the predicted probability that sample i belongs to class j.

The figures below illustrate the training history, encompassing both accuracy and loss history. These visualizations offer valuable insights into the model's learning progress and overall performance during the training process.

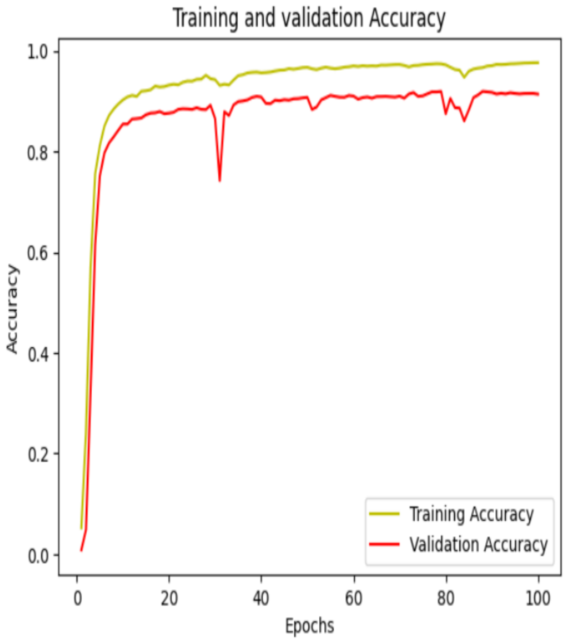


Figure 25: The training accuracy history

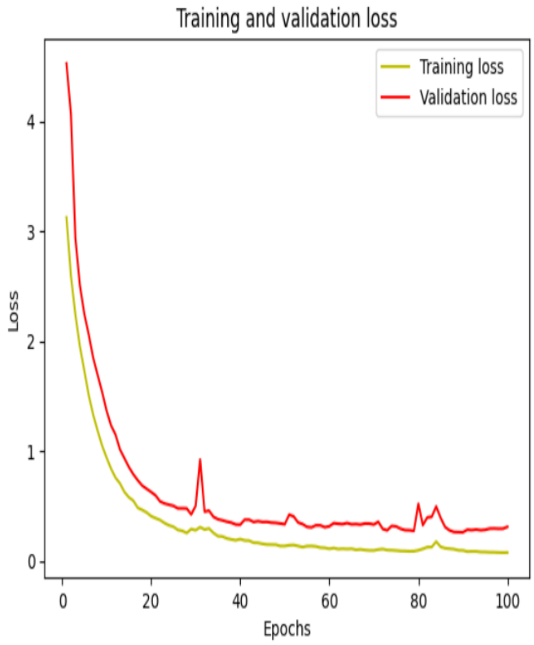


Figure 26: The training loss history

3. Benchmark comparison of U-Net+ ResNet34 with different methods

In this section, we present a benchmark comparison of our proposed method, with various other methods on 4 distinct datasets as shown in **Table 8**.

Dataset	**Method**	**Metrics**	
		mIoU	**Accuracy**
Land Cover.ai	DeepTriNet [93]	mIoU:80	
	U-Net, ResNet50 encoder [94]	mIoU:84.81	
	DeepLabv3+ OS4+augmentation [89]	mIoU:85.56	
	U-Net, ResNet34 encoder (Our method)	**mIoU:92.22**	
UAVid	Fast-SCNN [95]	mIoU:45.9	
	BiSeNet [96]	mIoU:61.5	
	FCN-8 [97]	mIoU:64	
	DeepLabv3+ [98]	mIoU:65	
	U-Net+ ConvLSTM [99]	**mIoU:76**	
	U-Net, ResNet34 encoder (Our method)	mIoU:70	
Aerial Semantic Segmentation	U-Net	mIoU:83.15	Acc: 90.72
	U-Net, ResNet18 encoder	mIoU:80.44	Acc: 89.18
	U-Net, ResNet101 encoder	mIoU:85.95	Acc: 92.40
	U-Net, ResNet34 encoder (Our method)	**mIoU:86.56**	**Acc : 92.77**
AeroScapes Dataset	FCN-Ensemble-MultiSource [91]	mIoU:57.08	
	U-Net [100]	mIoU:59.07	Acc: 68.33

	RCCT-ASPPNet [101] SS-Inhi-VGG16-FFT [92] MBMS DeepLabv3+ [100] **U-Net, ResNet34 encoder (Our method)**	mIoU:61.30 mIoU:64.43 mIoU:84.98 **mIoU: 91**	 Acc: 97.57 **Acc: 95.25**

Table 8: Benchmark comparison of U-Net+ ResNet34 with various methods, showcasing notable achievements across different datasets. In land cover.ai and AeroScapes test sets, our method achieves the highest mIoU of 92.22, 91 respectively. While The U-Net+ ConvLSTM achieves the highest mIoU of 76 compares to our competitive score of 70, our method remains competitive among several other methods. Moreover, on the Aerial semantic segmentation we got mIoU of 86.56, demonstrates robust performance across aerial scenes, outperforming other methods with deeper encoders such as ResNet 101

4. Experimental Results

4.1 Land Cover

a. Performance Evaluation

Following the successful training of the proposed model, it was applied to the test set, wherein its performance was assessed using metrics such as precision, recall, and, F1-score for each class. We obtained a mean intersection over union (mIoU) value of 92.22% on the entire test set. **Table 9** presents the results :

Class	Precision	Recall	F1-Score	Accuracy	Iou	Support
Buildings	0.97	0.95	0.96	0.97	0.93	1682065.0
Woodland	0.96	0.96	0.96	0.96	0.93	24258.0
Water	0.99	0.98	0.98	0.98	0.96	242865.0
Road	0.91	0.91	0.91	0.96	0.84	68479.0
Background	0.97	0.98	0.97	0.97	0.95	2176637.0

Table 9: Our methodology results on land cover.ai test set

We analyzed the segmentation results for various classes. **Figure 27** displays sample images from the test set alongside their corresponding segmentation masks generated by the model: buildings are dark Gray, woodlands are middle Gray, water is light Gray, and roads are white.

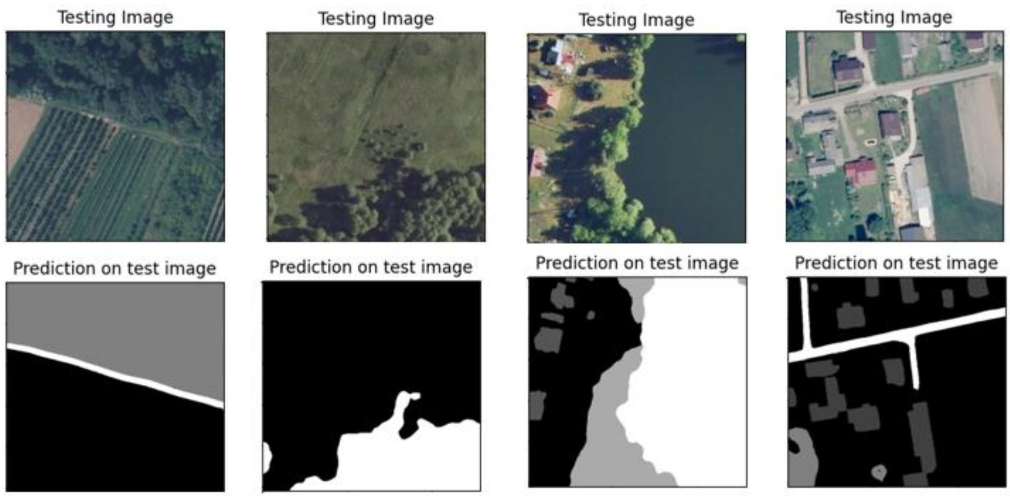


Figure 27: Close-ups of the images and their corresponding reference masks

b. Comparison to related research

To validate the efficacy of our methodology, we organized a comparative analysis of our results with the research conducted by Adrian Boguszewski et al. [89], entitled “LandCover.ai: Dataset for Automatic Mapping of Buildings, Woodlands, Water, and Roads from Aerial Imagery”.

They [89] evaluated the DeepLabv3+ model's efficacy in semantically segmenting aerial imagery. They demonstrated substantial enhancements in accuracy, attaining a mIoU of 85.56% through augmentation techniques. Challenges were identified in accurately delineating roads and buildings attributed to their diminutive dimensions and susceptibility to occlusion. Ultimately, their investigation underscores the efficacy of deep learning methodologies for automating mapping processes from aerial imagery.

The **Table 10** presents a comparative analysis of the results obtained by the DeepLabv3+ model augmented with OS 4 and the results obtained through our methodology:

Method	Buildings	Woodlands	Water	Roads	Background	Overall mIoU
DeepLabv3+ OS4+ augmentation	79.74%	91.46%	94.39%	68.74%	93.45%	**85.56%**
Our method	92.56%	93.08%	96.45%	83.89%	95.02%	**92.22%**

Table 10: Comparative results

Moreover, the visual comparative results are illustrated in **Figure 28**. The figure consists of four columns showcasing different aspects of land cover segmentation:

- **Close-ups of dataset images:** the first column displays close-up views of sample images from the dataset, capturing aerial imagery with various land cover features such as buildings, woodlands, water bodies, and roads.
- **Corresponding Annotations (Masks):** the second column presents the corresponding annotations or masks for each image. These annotations outline the ground truth labels for different land cover classes.
- **Prediction by Boguszewski and Batorski:** the third column showcases the segmentation predictions generated by DeepLabv3+OS4+augmentation.
- **Our Method Prediction:** the fourth column illustrates the segmentation predictions produced by our methodology.

The visual comparison between the annotations, DeepLabv3+OS4+augmentation predictions, and our method's predictions facilitates the evaluation of segmentation accuracy and the effectiveness of different methodologies in accurately mapping land cover from aerial imagery.

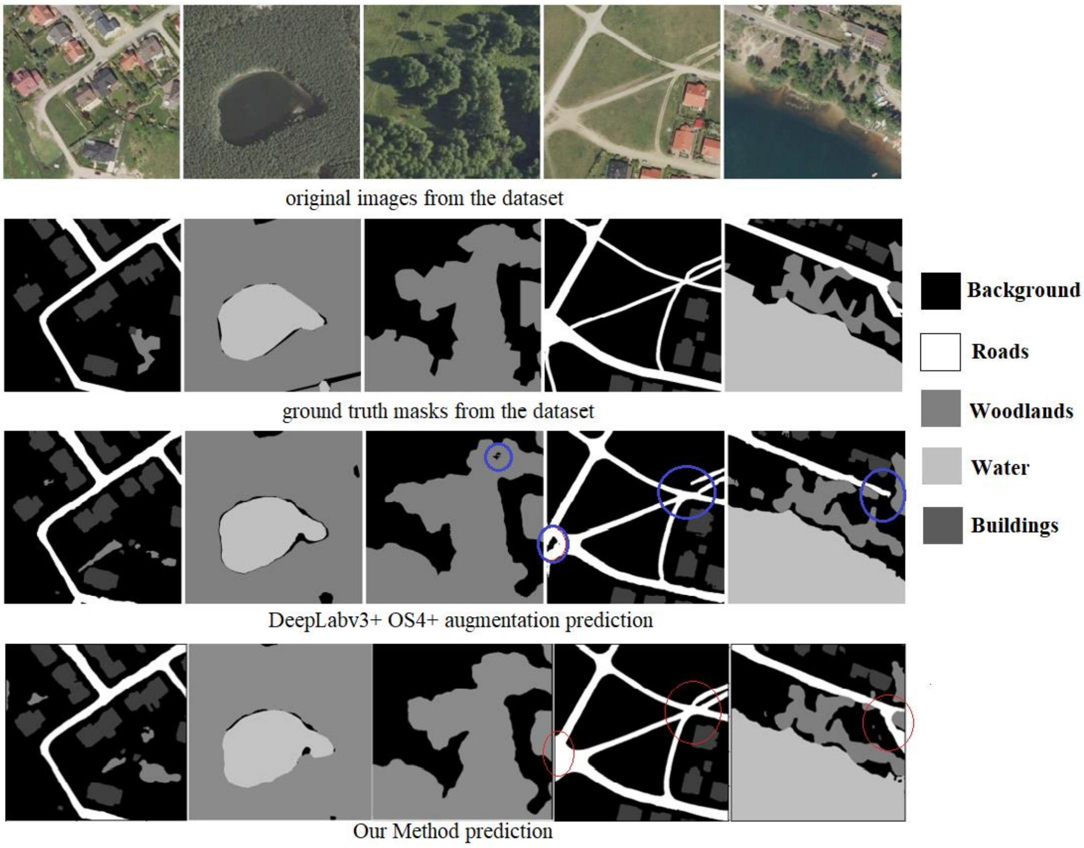


Figure 28: Visual comparison between DeepLabv3+ OS4+ augmentation, and our method for five test sets. The red circles represent areas where our method prediction closely matches the reality of the input image, The bleu circles represent areas where the DeepLabv3+ method prediction not closely matches the reality of the input image.

4.2 Aerial Semantic Segmentation

a. Performance Evaluation

In **Table 11**, we report the results obtained on the test set. The results highlight the improvement in accuracy, precision, recall, F1-score, and mIoU achieved by incorporating ResNet34 with U-Net compared to other models.

Model	Accuracy	Precision	Recall	F1-Score	MIoU
U-Net	90.72	90.61	90.72	90.64	83.15
U-Net + ResNet18	89.18	89.38	89.18	88.96	80.44
U-Net + ResNet34	**92.77**	**92.72**	**92.77**	**92.70**	**86.56**
U-Net + ResNet101	92.40	92.34	92.40	92.33	85.95
U-Net + efficientnetb0	90.46	90.52	90.46	90.29	82.58

Table 11: Comparative evaluation scores among models, where our method achieves the highest score

Figure 29 displays sample images from the test set providing visual representations of the segmentation results across various 24 classes.

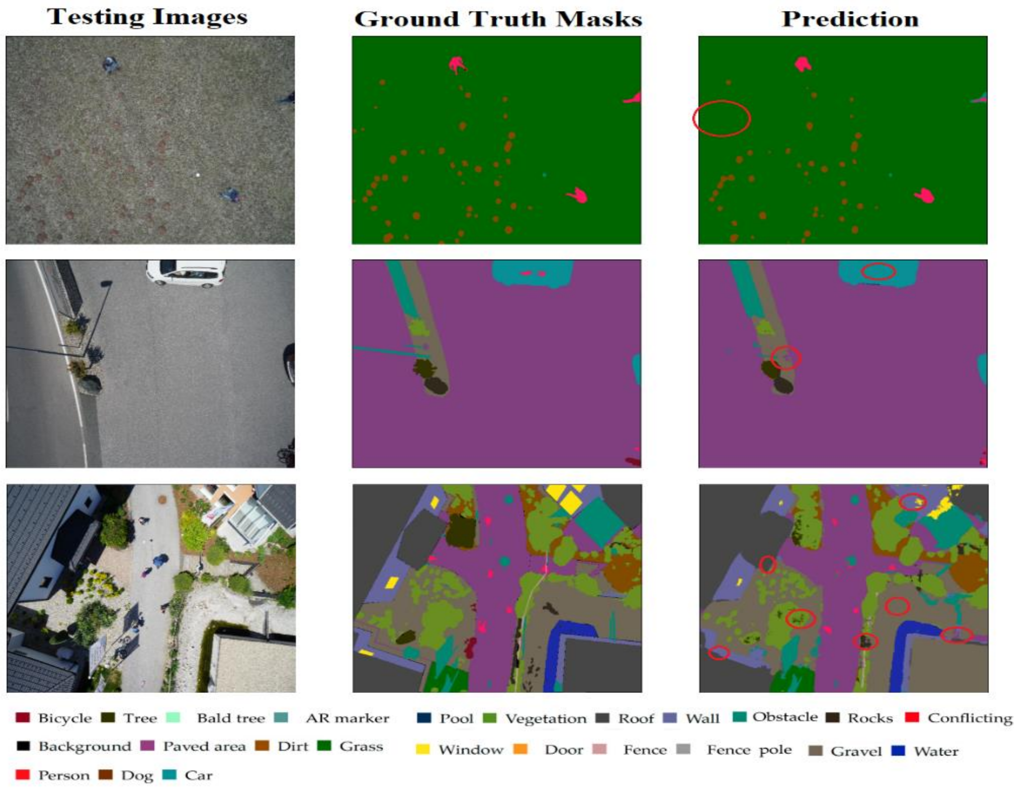


Figure 29: Sample images from the Aerial Semantic Segmentation test set providing the segmentation results

4.3 UAVid dataset

b. Performance Evaluation

In **Table 12**, we report the results obtained on the test set. We obtained a mean intersection over union (mIoU) value of 70.01% and overall accuracy value of 82.07%, The **Figure 30** displays sample images providing visual representations of the segmentation results across various 8 classes.

Classes	Precision	Recall	F1-Score	Accuracy	Iou	Support
Background clutter	0.75	0.69	0.72	0.90	0.56	1400411
Building	0.90	0.92	0.91	0.95	0.83	1983628
Road	0.81	0.88	0.85	0.95	0.73	1133775
Tree	0.83	0.87	0.85	0.91	0.73	2157555
Low vegetation	0.76	0.72	0.74	0.94	0.59	1008789
Moving car	0.67	0.49	0.57	0.99	0.40	91803
Static car	0.55	0.36	0.44	0.99	0.28	78386
Human	0.50	0.29	0.35	0.60	0.21	9973

Table 12: Our methodology results on the UAVid test set

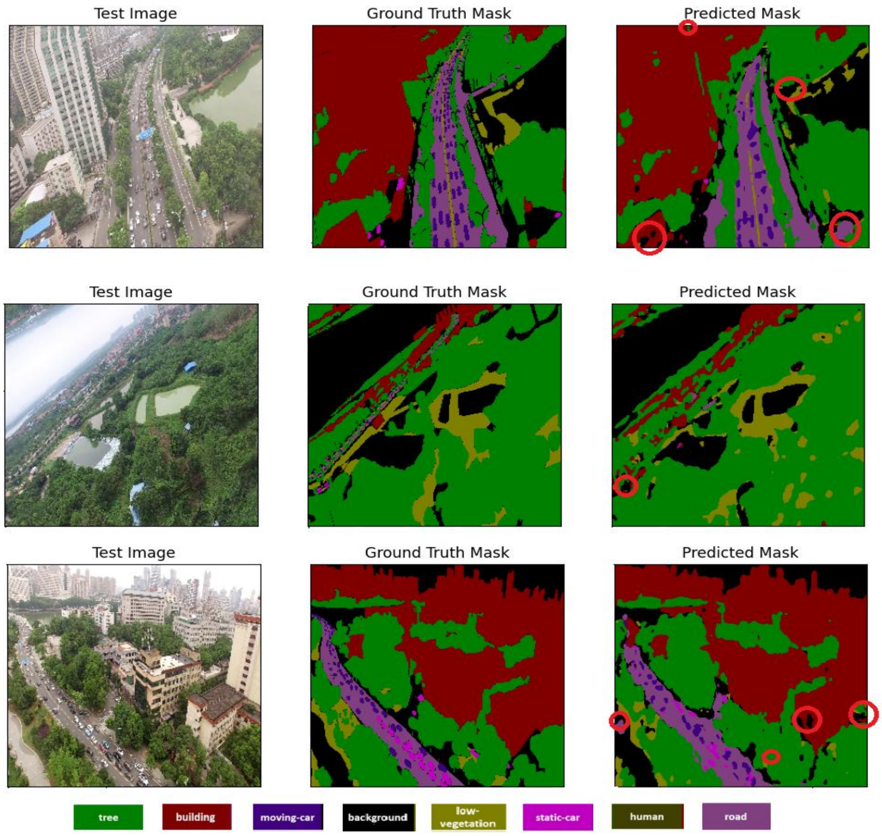


Figure 30: Close-ups images from the UAVid test set, the red circles represent areas where our method prediction doesn't closely match the reality of the input image

5. Discussion

The effectiveness of integrating the U-Net architecture with ResNet34 as its backbone in capturing intricate details and preserving crucial spatial information for accurate segmentation was evident in its strong performance compared to other methods.

In the Land Cover dataset, our method achieved promising results in the task of image segmentation, as demonstrated by the mean intersection over union (mIoU) value of 92.22% on the entire test set. This indicates the model's ability to effectively capture the spatial relationships between different land cover classes, resulting in accurate segmentation masks. For example, woodlands and water bodies achieved F1-scores of 0.96 and 0.98, indicating high accuracy in detecting water bodies. To validate the effectiveness of our approach, we compared it with the DeepLabv3+OS4+augmentation method, where our method showcased improvements, particularly in building segmenting, with a precision of 92.56% compared to their 79.74%.

Similarly, in the Aerial semantic segmentation drone dataset, we achieved promising results with an overall accuracy of 92.77%, demonstrating the effectiveness of our deep learning model in segmenting diverse urban features. While the model performed well overall, challenges were observed in accurately delineating smaller objects such as windows, doors, dirt, and obstacles due to their diminutive dimensions.

The U-Net+ ConvLSTM method achieves the highest mIoU of 76 compares to our competitive score of 70, our method remains competitive among several other methods such as Fast-SCNN, BiSeNet, FCN-8, and DeepLabv3+. We encounter challenges in accurately predicting the human class. Moreover, we achieved an impressive mIoU of 91 on AeroScapes test set. Our method excels in segmenting various objects in complex aerial scenes, surpassing most other methods including ensemble approaches and specialized architectures.

6. Challenges and Limitations

Despite the generally successful results, we encounter challenges in accurately identifying roads in the land cover dataset and struggle with smaller object categories like windows, doors and human in the Aerial semantic segmentation drone, and UAVid dataset. This highlights the need for more robust architectures tailored to such scenarios.

7. Conclusion

The integration of ResNet34 with U-Net achieves remarkable performance, scoring a mean intersection over union (mIoU) scores of 92.22% for land cover segmentation, 70% for urban scenes, and 91% for objects found in urban, and an overall accuracy of 92.77% for urban feature segmentation. These outcomes signify the robustness of our model in capturing complex spatial relationships and preserving critical spatial information. Comparative analysis with related studies, notably with the DeepLabv3 methodology, showcased significant improvements in segmentation accuracy, particularly in detecting buildings and woodlands. Despite these achievements, challenges persist in accurately identifying smaller objects such as roads, windows, human and doors. This highlights the need for continued research and the development of more robust architectures tailored to handle such scenarios effectively.

General Conclusion

General Conclusion

Unmanned Aerial Vehicles (UAVs) have revolutionized remote sensing, offering continuous access to high-resolution data from previously inaccessible areas. This data is particularly valuable for semantic segmentation, a deep learning technique crucial for tasks like land cover mapping and object detection in environmental monitoring and urban planning. However, effectively analyzing this data can be hampered by the vanishing gradient problem, where information gets progressively lost during training in deep neural networks.

This research addresses this challenge by leveraging deep learning and specifically tackling the vanishing gradient problem. We integrate a U-Net architecture with a ResNet34 backbone. The ResNet architecture is a deep learning model specifically designed to mitigate the vanishing gradient problem. This allows the model to learn complex features within UAV imagery for semantic segmentation tasks with greater depth and effectiveness. We trained and evaluated the model on diverse UAV datasets (LandCover.ai, Aerial Semantic Segmentation, UAVid, AeroScapes) leveraging GPUs, the Adam optimizer, and the Dice loss function. We meticulously optimized hyperparameters and implemented specific training strategies to maximize performance.

Our approach achieves remarkable results, with Intersection over Union (IoU) scores exceeding 90% on LandCover.ai and AeroScapes datasets (92.0% and 91.0%, respectively). Compared to existing methods (DeepLabv3, FCN-8, RCCT-ASPPNet, SS-Inhi-VGG16-FFT), our model demonstrates superior accuracy across all metrics (precision, recall, F1-score, mIoU). These findings solidify the power of deep learning for accurate and efficient large-scale geospatial analysis in remote sensing.

Looking ahead, this research paves the way for further exploration of deep learning and UAV data integration for semantic segmentation tasks. Future work can refine the proposed approach, investigate new applications, and validate its robustness in real-world scenarios.

References

References

[1] Mohamed Wissam al-Din, «Introduction to remote sensing and its applications,» Dammam University, p. 473, 2014.

[2] Sergeieva, Larysa M and Stoichyk, Tetiana I and Tarasova, Olena V and Sulyma, Tetiana S and Tarasiuk, Iryna V, «Development of the modern educational environment of the institution of professional (vocational-technical) education in the digital space,» Linguistics and Culture Review, vol. 5, n° 13, pp. 459--475, 2021.

[3] Y.Amghar, Classification des données de télédétection par une approche neuronale RBF-SGNG, Oran, Algérie: Université des sciences et de la technologie d'Oran Mohamed Boudiaf, 2006.

[4] Al-Najar Abd Al Hamid Ahmed , Introduction in Remote Sensing, Cairo, 2022.

[5] Abdul-Hameed, Abeer, «Remote Sensing Optical Images Applications in Vegetation Monitoring and Mapping,» University of Baghdad / College of Science Department of Physics, Baghdad, 2022.

[6] Al Dalimi and Khalaf Hussein, Geomatics Digital Photo Processing Book.

[7] Dr Mohammed .Juma, Foundations and Applications of Remote Sensing, Cairo: Arab Republic of Egypt, 2015, p. 208.

[8] Geolearn, «The Basic Concept of Remote Sensing,» Geolearn, 2020.

[9] D. Rao, «Project Survey of Natural and Agricultural Resources using Remote Sensing Techniques and Geographic Information System,» General Authority for Remote Sensing, vol. 33, n° 17, pp. 4-10, 2010.

[10] Knowledge Hub GCSE, «The electromagnetic spectrum,» Online Learning College, 2022.

[11] Lemenkova, Polina and Debeir, Olivier, «Recognizing the Wadi Fluvial Structure and Stream Network in the Qena Bend of the Nile River, Egypt, on Landsat 8-9 OLI Images,» Information, vol. 14, n° 14, p. 249, 2023.

[12] Dierssen, Heidi M and Ackleson, Steven G and Joyce, Karen E and Hestir, Erin L and Castagna, Alexandre and Lavender, Samantha and McManus, Margaret A, «Living up to the hype of hyperspectral aquatic remote sensing: science, resources and outlook,» Frontiers in Environmental Science, vol. 9, p. 649528, 2021.

[13] D. o. Geography, «Introduction to Remote Sensing,» Humboldt-Universität, Berlin, 2023.

[14] C. Diltsm, «Numérisez vos collections,» RCIP, 2001.

[15] Boukhelkhal Salaheddine and Lassar Yahia Haitham, «Study and implementation of a detection system for plant diseases by image processing at the level of agricultural greenhouses,» Universitie Yahia Fares, Medea, 2020.

[16] Van der Meer, Freek and Van Dijk, Paul and Van Der Werff, Harald and Yang, Hong, «Remote sensing and petroleum seepage: a review and case study,» Terra Nova, vol. 14, n° 11, pp. 1--17, 2002.

[17] Navalgund, Ranganath R and Jayaraman, V and Roy, PS, «Remote sensing applications: An overview,» current science, pp. 1747--1766, 2007.

[18] A. rabia, «Together we make the difference in education,» chez Remote Sensing Applications Lesson, University of Rochester, 2022, p. 1.

[19] Arifjanov, Aybek and Akmalov, Shamshodbek and Akhmedov, Islombek and Atakulov, Dinislom, «Evaluation of deformation procedure in waterbed of rivers,» IOP Publishing, vol. 403, n° 11, p. 12155, 2019.

[20] Farouk Paty, Ghassan Bin, «Artificial Intelligence And Machine Learning,» Journal of Information Studies and Technology, vol. 2023, n° 112, p. 12, 2023.

[21] N. Vladimir, «An overview of the supervised machine learning methods,» Horizons, vol. 4, n° 151-62, p. 56, 2017.

[22] T. M. Mustafa, «Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions,» Computers, vol. 12, n° 15, p. 91, 2023.

[23] Gowda, Harsha S and Suhil, Mahamad and Guru, DS and Raju, Lavanya Narayana, «Semi-Supervised Text Categorization using Recursive,» Springer, pp. 217--227, 2017.

[24] DatabaseTown, «Supervised Learning: Algorithms, Examples, and How It Works,» DatabaseTown, 2020.

[25] B. Nada, «Artificial Intelligence Techniques for the Development of Statistical Machine Learning,» Iraqi Journal of Information Technology, vol. 9, pp. 452--459, 2019.

[26] C. Gershenson, « Artificial Neural Networks for Beginners,» researchgate, p. 2, 2003.

[27] P. S. Alireza Hajian, Application of Soft Computing and Intelligent Methods in Geophysics, Switzerland: springer international publishing, 2018, p. 10.

[28] Aboukarima, Abdulwahed and Elsoury, Hussien and Menyawi, «Artificial Neural Network Model for the Prediction of the Cotton Crop Leaf Area,» International Journal of Plant & Soil Science, vol. 8, pp. 1-13, 2015.

[29] B. Fatima, «The contribution of artificial networks to forecasting the volume of sales to support Administrative decision-making in economic institutions: a study of certain

administrative institutions(Doctoral thesis),» Faculty of Economics and Management Sciences, 2019.

[30] ali Sharqawi mohammed, Artificial Intelligence and Neural Networks, Cairo: Office Printers Egyptian Modern, 2016.

[31] Ali Bashar Al Sharif, and Abdul Ghani Juma, «Applications on Synthetic Neural Networks Distinguish Arabic alphabets by Matlab,» Faculty of IT Engineering and Communication, p. 9, 2012.

[32] Hamid Ashour, Synthetic Neural Networks, 5- 47: research, 2020.

[33] Ashour Marwan and Abdul Hamid, «Synthetic Neural Networks,» researchgate, 2020.

[34] Omar Saber, Qasim Isra, and Rustam Mohammed, «Analytical Mathematical Study of Algorithms Synthetic Neural Networks in Suitability Model for Medical Diagnosis,» Rafidain Journal of Science Computer & Mathematics, n° 11, p. 187 190, 2013.

[35] L. Mohammed, «Entrance to artificial intelligence and machine learning,» computer and Computer Academy, pp. 44-103, 2020.

[36] Miguel Pérez-Enciso and Laura M. Zingaretti , «A Guide on Deep Learning for Complex Trait Genomic Prediction,» Genes , vol. 10, n° %17, p. 553, 2019.

[37] Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E, «ImageNet classification with deep convolutional neural networks,» Communications of the ACM, vol. 60, n° 16, pp. 84--90, 2017.

[38] Fukushima, Kunihiko, «Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,» Biological cybernetics, vol. 36, n° 14, pp. 193-202, 1980.

[39] LeCun, Yann and Boser, Bernhard and Denker, John S and Henderson, Donnie and Howard, Richard E and Hubbard, Wayne and Jackel, Lawrence D, «Backpropagation applied to handwritten zip code recognition,» Neural computation, vol. 1, n° 14, pp. 541-551, 1989.

[40] Russakovsky, Olga and Deng, Jia and Su, Hao and Krause, Jonathan and Satheesh, Sanjeev and Ma, Sean and Huang, Zhiheng and Karpathy, Andrej and Khosla, Aditya and Bernstein, Michael and others, «Imagenet large scale visual recognition challenge,» International journal of computer vision, vol. 115, pp. 211--252, 2015.

[41] Saaid Abd elkarim, «Understanding neural networks from neurons to repetitive neural networks RNN, bypass neural networks CNN and deep learning Deep Learning,» bhrres, Bahrin, 2022.

[42] L. Farihi, «Classifying images of Cats and Dogs using Convolutional Neural Networks (CNNs),» Medium, 2021.

[43] Podareanu, Damian and Codreanu, Valeriu and Aigner, Sandra and Leeuwen, Caspar and Weinberg, Volker, «Best Practice Guide - Deep Learning,» ResearchGate, 2019.

[44] Yani Muhamad and Irawan, and Setianingsih Casi, «Application of Transfer Learning Using Convolutional Neural Network Method for Early Detection of Terry's Nail,» Journal of Physics: Conference Series, vol. 1201, p. 12052, 2019.

[45] Bhatt, Dulari and Patel, Chirag and Talsania, Hardik and Patel, Jigar and Vaghela, Rasmika and Pandya, Sharnil and Modi, Kirit and Ghayvat, Hemant, «CNN variants for computer vision: History, architecture, application, challenges and future scope,» Electronics, vol. 10, n° 120, p. 2470, 2021.

[46] T. Beysolow II, Introduction to deep learning using R: A step-by-step guide to learning and implementing deep learning models using R, Apress, 2017, p. 1.5.7.

[47] Xin, Yang and Kong, Lingshuang and Liu, Zhi and Chen, Yuling and Li, Yanmiao and Zhu, Hongliang and Gao, Mingcheng and Hou, Haixia and Wang, Chunhua, «Machine Learning and Deep Learning Methods for Cybersecurity Corresponding,» Ieee access, vol. 6, pp. 35365-35381, 2018.

[48] Novikova Anna, «AI's Powerful duo: Difference between machine learning and deep learning,» intersog, 2023.

[49] Stefan Ionescu, Camelia Delcea, Nora Chirita and Ionut Nica, «Exploring the Use of Artificial Intelligence in Agent-Based Modeling Applications: A Bibliometric Study,» Algorithms, vol. 17, n° 11, p. 21, 2024.

[50] A.Hasani, H.Majidizadeh, «Semantic segmentation of UAV images based on U-NET in urban area,» ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 10, n° 14, pp. 451--457, 2023 .

[51] Lucas Prado Osco, José Marcato Junior , Ana Paula Marques Ramos, Lúcio André de Castro Jorge, «A Review on Deep Learning in UAV Remote Sensing,» arXiv:2101.10861 [cs.CV], vol. 4, p. 1, 2023.

[52] Yao, Huang and Qin, Rongjun and Chen, Xiaoyu, «Unmanned aerial vehicle for remote sensing applications—A review,» Remote Sensing, vol. 11, n° 112, p. 1443, 2019.

[53] Carlos Sampedro, Alejandro Rodriguez-Ramos, and Pascual Campoy, «A Review of Deep Learning Methods and Applications for Unmanned Aerial Vehicles,» Sensors, vol. 2017, p. 26, 2017.

[54] Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey , «ImageNet classification with deep convolutional neural networks,» Communications of the ACM, vol. 60, n° 16, pp. 84--90, 2017.

[55] Aabid, Abdul and Parveez, Bisma and Parveen, Nagma and Khan, Sher and Raheman, Md Abdul and Zayan, Mohammed and Ahmed, Omar, «reviews on Design and

Development of Unmanned Aerial Vehicle (Drone) for Different Applications},» Journal of Mechanical Engineering Research and Developments, vol. 45, pp. 53-69, 2022.

[56] Zhang, Zhengxin and Zhu, Lixue, «A Review on Unmanned Aerial Vehicle Remote Sensing: Platforms, Sensors, Data Processing Methods, and Applications,» Drones, vol. 7, n° 16, p. 398, 2023.

[57] Cheng, Xinglu and Sun, Yonghua and Zhang, Wangkuan and Wang, Yihan and Cao, Xuyue and Wang, Yanzhao, «Application of Deep Learning in Multitemporal Remote Sensing Image Classification,» Remote Sensing, vol. 15, n° 115, p. 3859, 2023.

[58] Weinland, Daniel and Ronfard, Remi and Boyer, Edmond, «A survey of vision-based methods for action representation, segmentation and recognition,» Computer vision and image understanding, vol. 115, n° 12, pp. 224--241, 2011.

[59] Yanming Guo, Yu Liu, Theodoros Georgiou and Michael lew, «A review of semantic segmentation using deep neural networks,» International Journal of Multimedia Information Retrieval, vol. 7, p. 87–93, 2018.

[60] Long, Jonathan and Shelhamer, Evan and Darrell, Trevor, «Fully convolutional networks for semantic segmentation,» chez Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431--3440.

[61] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, «U-Net: Convolutional Networks for Biomedical Image Segmentation. In Medical Image Computing and Computer-Assisted,» Medical Image Computing and Computer-Assisted Intervention, vol. 1, p. 234–241, 2015.

[62] Badrinarayanan, Vijay and Kendall, Alex and Cipolla, Roberto, «Segnet: A deep convolutional encoder-decoder architecture for image segmentation,» IEEE transactions on pattern analysis and machine intelligence, vol. 39, n° 112, pp. 2481--2495, 2017.

[63] M.Ullah, A.Mohammed and F.Alaya Cheikh, «PedNet: A spatio-temporal deep convolutional neural network for pedestrian segmentation,» Journal of Imaging, vol. 4, p. 107, 2018.

[64] Nassim Ammour, Haikel Alhichri, Yakoub Bazi, Bilel Benjdira, Naif Alajlan and Mansour Zuair, «Deep learning approach for car detection in UAV imagery,» Remote Sensing, vol. 9, n° 14, p. 312, 2017.

[65] Thomas Moranduzzo and Farid Melgani, «Detecting cars in UAV images with a catalog-based approach,» IEEE Transactions on Geoscience and Remote Sensing, vol. 52, n° 110, pp. 6356-6367, 2014.

[66] A. Micheal, A.Vani, K.Sanjeevi and C.HLin, «Object detection and tracking with UAV data using deep learning,» Journal of the Indian Society of Remote Sensing, vol. 49, n° 13, pp. 463-469, 2021.

[67] Z.Qiu, T.Yao and T.Mei, «Learning deep spatio-temporal dependence for semantic video segmentation,» IEEE Transactions on Multimedia, vol. 20, n° %14, pp. 939-949, 2017.

[68] S. Ahram, «Deep Learning-Based Semantic Segmentation of Urban Areas Using Heterogeneous Unmanned Aerial Vehicle Datasets,» MDPI, vol. 10, pp. 3-10, 2023.

[69] Natalia Ziemba-Jankowska, Tomasz Dziedzic, «LandCover.ai: Dataset for Automatic Mapping of Buildings, Woodlands, Water,» arXiv:2005.02264, vol. 3, p. 4, 2021.

[70] Fadi Shaar , Arif Yılmaz , Ahmet Ercan, Topcu Yehia, Ibrahim Alzoubi, «Remote Sensing Image Segmentation for Aircraft Recognition Using U-Net as Deep Learning Architecture,» Applied Sciences, vol. 14, n° 16, 21 March 2024.

[71] S.Kentsch , ML.Lopez Caceres , D.Serrano , F.Roure and Y.Diez , «Computer vision and deep learning techniques for the analysis of drone-acquired forest images, a transfer learning study,» Remote Sensing, vol. 12, n° 18, p. 1287, 2020.

[72] Richmond Akwasi Nsiah, Saviour Mantey and Yao Yevenyo Ziggah, «Building segmentation from UAV orthomosaics using unet-resnet-34 optimised with grey wolf optimisation algorithm,» Smart Construction and Sustainable Cities, pp. 6-18, 2023.

[73] Yiwen Wang, Ye Lyu, Yanpeng Cao and Michael Ying Yang, «Deep learning for semantic segmentation of UAV videos,» IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 2459-2462, 2019.

[74] Lucas Prado Osco, Keiller Nogueira, Ana Paula Marques Ramos, Mayara Maezano Faita Pinheiro, «Semantic segmentation of citrus-orchard using deep neural networks and multispectral UAV-based imagery,» Precision Agriculture, vol. 22, p. 1171–1188, 2021.

[75] M. Kampffmeyer, B. Salberg and R. Jenssen, «Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks. In Computer Vision and Pattern Recognition Workshops (CVPRW),» IEEE Conference, p. pages 680–688, 2016.

[76] Eui-ik Jeon , Sunghak Kim , Soyoung Park , Juwon Kwak and Imho Choi, «Semantic segmentation of seagrass habitat from drone imagery based on deep learning: A comparative study Ecological Informatics,» Ecological Informatics, vol. 66, p. 101430, December 2021.

[77] Jiawei Mo ,Yubin Lan ,Dongzi Yang ,Fei Wen ,Hongbin Qiu ,Xin Chen and Xiaoling Deng, «Deep Learning-Based Instance Segmentation Method of Litchi Canopy from UAV-Acquired Images,» Remote Sensing, vol. 13, n° 119, p. 3919, 2021.

[78] Shrikrishna Kolhar , Jayant Jagtap, «Convolutional neural network based encoder-decoder architectures for semantic segmentation of plants,» Ecological Informatics, vol. 64, p. 101373, 2021.

[79] Emek Soylu, Busra and Guzel, Mehmet Serdar and Bostanci, Gazi Erkan and Ekinci, Fatih and Asuroglu, Tunc and Acici, Koray, «Deep-Learning-Based Approaches for

Semantic Segmentation of Natural Scene Images: A Review,» Electronics, vol. 12, n° 112, p. 2730, 2023.

[80] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun, «Deep Residual Learning for Image Recognition In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),» 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), p. 770–778, 2016.

[81] B. Babina, «Deep Residual Learning for Image Recognition (ResNet Explained),» analyticsvidhya, p. 2, 2023.

[82] Zhang, Zijian and Wang, Shuqi and Li, Zhen and Gao, Feng and Wang, Huaqiang, «A Multi-Dimensional Covert Transaction Recognition Scheme for Blockchain,» Mathematics, vol. 11, p. 1015, 2023.

[83] Almoosawi, Noor Mohammed and Khudeyer, Raidah Salim, «ResNet-34/DR: a residual convolutional neural network for the diagnosis of diabetic retinopathy,» Informatica, vol. 45, n° 17, 2021.

[84] Korfiatis, Panagiotis and Kline, Timothy L and Lachance, Daniel H and Parney, Ian F and Buckner, Jan C and Erickson, Bradley J, «Residual deep convolutional neural network predicts MGMT methylation status,» Journal of digital imaging, vol. 30, pp. 622-628, 2017.

[85] Amirreza Shaban, Shray Bansal, Zhen Liu, Irfan Essa, and Byron Boots, «One-shot learning for semantic segmentation,» arXiv preprint arXiv:1709.03410, 2017.

[86] Xiaolin Zhang, Yunchao Wei, Yi Yang, and Thomas Huang, «Sg-one: Similarity guidance network for one-shot semantic segmentation,» IEEE Transactions on Cybernetics, vol. 50, n° 19, pp. 3855 - 3865, 2020.

[87] A. Tharwat, «Classification assessment methods,» Applied Computing and Informatics, vol. 17, n° 11, p. 30, 2020.

[88] M.Hossin, MN.Sulaiman and Hossin Mohammad, «A review on evaluation metrics for data classification evaluations,» International journal of data mining & Knowledge Management Process, vol. 5, n° 11, p. 1, 2015.

[89] Adrian Boguszewski, Dominik Batorski, Natalia Ziemba-Jankowska, Tomasz Dziedzic and Anna Zambrzycka, «LandCover.ai: Dataset for Automatic Mapping of Buildings, Woodlands, Water,» Agency for Restructuring and Modernisation of Agriculture, vol. 3, pp. 1102--1110, 2021.

[90] G. U. o. Technology, «The Semantic Drone Dataset,» Graz University of Technology, 2019.

[91] Ishan Nigam, Chen Huang and Deva Ramanan, «Ensemble Knowledge Transfer for Semantic Segmentation,» IEEE Winter Conference on Applications of Computer Vision, Vols. 1 sur 2 2018-Janua, p. 1499–1508, 2018.

[92] Ihsan Ullah, Sean Reilly and Michael G. Madden, «Enhancing Semantic Segmentation of Aerial Images with Inhibitory Neurons,» 2020 25th International Conference on Pattern Recognition (ICPR), pp. 5451--5458, 2021.

[93] Tareque Bashar Ovi, Shakil, Nomaiya Bashree, Md Shofiqul Islam and Muhammad Nazrul Islam, «DeepTriNet: A Tri-Level Attention Based DeepLabv3+ Architecture for Semantic Segmentation,» Computer Science and Engineering, vol. 1, p. 9, 2023.

[94] Adam J. Stewart, Caleb Robinson and Isaac A. Corley, «TorchGeo: Deep Learning With Geospatial Data,» SIGSPATIAL 22: Proceedings of the 30th International Conference on Advances in Geographic Information Systems, vol. 4, p. 8, 2022.

[95] Rudra PK Poudel and Stephan Liwicki, «Fast-SCNN: Fast Semantic Segmentation Network,» arXiv preprint arXiv:1902.04502, vol. 1, 2019.

[96] Libo Wang, Rui Li , Dongzhi Wang , Chenxi Duan , Teng Wang and Xiaoliang Meng, «Transformer Meets Convolution: A Bilateral Awareness Network for Semantic Segmentation of Very Fine Resolution Urban Scene Images,» Remote Sensing, vol. 13, n° 116, p. 14, 2021.

[97] Jonathan Long, Evan Shelhamer and Trevor Darrell, «Fully convolutional networks for semantic segmentation,» Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431--3440, 2015.

[98] LC. Chen, YG. Zhu, F P. Schroff, and H. Adam, «Encoder-decoder with atrous separable convolution for semantic image segmentation,» Proc Eur. Conf. Comput, vol. 14, n° 16, p. 801–818, 2018.

[99] Girisha S, Ujjwal Verma , Manohara Pai and M.Senior Member, «UVid-Net: Enhanced Semantic Segmentation of UAV Aerial Videos by Embedding Temporal Information,» IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, vol. 14, p. 4115, 2021.

[100] Jiangxin Hui and Hong Zhang, «A semantic segmentation network based on multibranch structures and multi-scale modules,» Research Square, vol. 1, p. 17, 2022.

[101] Yazhou Li , Zhiyou Cheng, Chuanjian Wang, Jinling Zhao and Linsheng Huang, «RCCT-ASPPNet: Dual-Encoder Remote Image Segmentation Based on Transformer and ASPP,» Remote Sensing, vol. 15, n° 12, p. 379, 2023.