THE PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
THE MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

AMAR TELIDJI UNIVERSITY OF LAGHOUAT
FACULTY OF TECHNOLOGY
DEPARTMENT OF ELECTRONIC

Master's dessertation

Domain : Science and Technology

Option : Telecommunications and Networks

Theme

Processing And Extracting Data From A Dataset For Artificial Intelligence IN Neural Radiance Fields

---

Traitement et extractiondes données d'un dataset via l'IA pour le Neural Radiance Fields

---

by :

DEMANA FERIAL
GUESMIA ZINEB

chairman: M^{lle} *Abdelhafidi Nadjma*
Examiner: D^{r} *Chaker Saleh*

Supervised by: D^{r} *BOUZIDI MOHAMMED REDHA*
Co Supervisor: M^{me} TOUHAMI IMANE

Academic year 2023/2024

ملخص :

تتناول هذه المذكرة دراسة استخدام تقنيات الذكاء الاصطناعي في معالجة الصور، مع التركيز على إزالة الخلفيات في الصور المعقدة مثل الأشجار. تحلل المذكرة منصة روبوفلو كأداة قوية لإدارة البيانات وتحسينها، وتوضح كيفية استخدامها في نقل التعلم لتحسين أداء النموذج. كما تم مناقشة أدوات مثل فوتوشوب وتقنيات التعلم العميق مثل الشبكات العصبية التلافيفية التوليدية والشبكات العصبية التلافيفية التنافسية، مع تسليط الضوء على كيفية استخدام جوجل كولاب بفعالية لاختبار النماذج الأولية. أظهرت النتائج أن الذكاء الاصطناعي يحسن بشكل كبير دقة وكفاءة إزالة الخلفية، مما يقلل من الجهد اليدوي المطلوب. تؤكد المذكرة على أهمية التطوير المستمر لتقنيات الذكاء الاصطناعي لتحسين معالجة الصور، مما يمهد الطريق لتطبيقات جديدة ومحسنة في المستقبل

الكلمات المفتاحية : الذكاء الاصطناعي , معالجة الصور , إزالة الخلفيات ,التعلم العميق ,النموذج, الشبكات العصبية التلافيفية , نقل التعلم.

ABSTRACT

In this works the use of artificial intelligence techniques in image processing, with a focus on background removal in complex images such as trees. It analyzes the Roboflow platform as a powerful tool for data management and enhancement, demonstrating its use in transfer learning to improve model performance. Tools like Photoshop and deep learning techniques such as U-Net and GANs are also discussed, highlighting how Google Colab can be used effectively for prototyping models. The results show that AI significantly improves the accuracy and efficiency of background removal, reducing the manual effort required. The thesis works the continuous development of AI techniques to enhance image processing, paving the way for new and improved applications in the future.

Keywords: *Artificial intelligence, image processing, background removal, deep learning, model, convolutional neural networks, transfer learning..*

RÉSUMÉ

Dans cette travail examine l'utilisation des techniques d'intelligence artificielle dans le traitement des images, en se concentrant sur la suppression de l'arrière-plan dans les images complexes telles que les arbres. Elle analyse la plateforme Roboflow en tant qu'outil puissant pour la gestion et l'amélioration des données, en démontrant son utilisation dans l'apprentissage par transfert pour améliorer les performances du modèle. Des outils tels que Photoshop et des techniques d'apprentissage profond telles que U-Net et GANs sont également discutés, mettant en évidence comment Google Colab peut être utilisé efficacement pour le prototypage de modèles. Les résultats montrent que l'IA améliore considérablement la précision et l'efficacité de la suppression de l'arrière-plan, réduisant ainsi l'effort manuel requis. Le travail souligne le développement continu des techniques d'IA pour améliorer le traitement des images, ouvrant la voie à des applications nouvelles et améliorées à l'avenir.

Mots clés : *Intelligence artificielle, traitement d'images, suppression de l'arrière-plan, apprentissage profond, modèle, réseaux de neurones convolutifs, apprentissage par transfert.*

Acknowledgements

First and foremost, we express our gratitude to Almighty God, who granted us the courage and determination to embark on this humble endeavor. We also take this opportunity to sincerely thank everyone who assisted us in any way to accomplish this thesis. Our heartfelt thanks go to Mr. Bouzidi Mohamed Redha, as the thesis supervisor, and to Professor Touhami Iman, who was our companion every step of the way, may God reward them generously, for their invaluable advice and guidance throughout our research period. We also extend our gratitude to the committee members who evaluated our work. Finally, we deeply appreciate all the officials and teachers in the Electronics Department at Amar Thelidji University of Laghouat who contributed to our education and training.

Dedication

With pride and gratitude, I dedicate this memorandum to dear ones who have profoundly impacted my journey in ways beyond words.

To my warm-hearted mother, Fatima Rahmoun, thank you for your boundless love and unwavering support through all the challenging and easy moments.
To my beloved father, El Azhari, your example in perseverance and generosity, your wise guidance, and constant belief in my abilities have enabled me to reach this moment.
To my supportive siblings, Massoud, Karima, Madani, Kauthar, and kouider, my uncle Youssef Rahmoun you complete my life and nurture my spirit with continuous encouragement and understanding.
The two pieces of my heart, Adam and Ghaith, your innocent smiles and sincere love fill our lives with happiness and joy.
To my departed grandfather, may Allah bless him, your wise upbringing and the legacy you left continue to guide us. May Allah envelop you in His mercy and grant you a place in His vast gardens.
To Zineb, my college companion and the most beautiful serendipity on my academic journey.

To my dear family, loyal friends, colleagues, and all my teachers, thank you for your encouragement and assistance, whether in words or deeds. I am grateful for everything.
This memorandum carries each of your fingerprints, and it would not be complete without you. I hope you will always be in my life, inspiring me to move forward and achieve more success and accomplishments.

DEMANA FERIAL

Dedication

With enormous pleasure, an open heart and immense joy, I dedicate my work : To my loving parents: "Ismail & Fatna" No dedication can express my respect, my eternal love and my consideration for the sacrifices you have made for my education and my well-being. May God, the Most High, grant you health, happiness and long life and ensure that I never disappoint you.

To the soul of my dearest: my grandmother" Fatna Gueffaf",I wished so much that you would see me raise my hat high and see your smilemay God have mercy on her. To my second father, who is always present in my heart, my uncle "AbdelRahman Guesmia", may God have mercy on him. to my dear grandfather:" Mohamed Gueffaf" show I wish you would be by my side in this beautiful moment of my life may God have mercy on him. To my brother and sisters: Bachir, Naoual, Asma, I ask God Almighty to grant you health, happiness and long life, I want to thank them very much for their stand and support.

To the Guesmia and Gueffaf families: for their encouragement throughout time, thank you very much for supporting you. Mighty God gives you good health,happiness and success. To My advisor, and faithful my cousin: Chouireb Amina who helped me assisted in difficult times,specially in this work.I am very grateful to you, and I do not I can never thank you enough for your kindness, your generosity, your precious help.

To Ferial, my friend and companion of my university career, I thank you for being by my side all this time and for a long time, God willing. To all my friends, colleagues and professors who accompanied me throughout my university career, thank you all.

ZINEB GUESMIA

Contents

List of Figures **IV**

GENERAL INTRODUCTION **1**

1 OBJECT DETECTION TECHNIQUES **4**

1.1 Introduction 5
1.2 Artificial intelligence 5
1.2.1 Artificial technologies 5
1.2.2 Neuronal environment 6
1.3 Machine learning 8
1.3.1 Machine learning algorithm: 8
1.3.2 List of common machine learning algorithms : . . . 9
1.3.3 Notation of dataset 9
1.3.4 Training set and test set: 10
1.4 Type of machine learning 10
1.5 Traditional machine learning 12
1.6 Deep learning 14
1.6.1 Introduction to deep learning: 14
1.6.2 A brief history of deep learning: 15
1.6.3 Reasons for the success of deep learning: 16
1.7 AI model 16
1.7.1 What is the ai model: 16
1.7.2 Types of ai models: 17
1.7.3 Model yolo : 18

1.8 Transfer learning 24
1.8.1 Types of transferlearning: 24
1.8.2 Utilizations of transfer learning: 25
1.9 Conclusion 27

2 ARTIFICIAL INTELLIGENCE IN AGRICULTURE 28
2.1 Introduction 29
2.2 Importance 29
2.2.1 Impact of ai in agriculture 35
2.3 Applications: 37
2.3.1 Overview: 37
2.4 Neural radiance fields NeRF 38
2.4.1 Context: 38
2.4.2 Neural radiance field scene representation: 40
2.4.3 Volume rendering with radiance fields: 42
2.4.4 Other relevant investigations: 45
2.5 Conclusion : 48

3 BACKGROUND EXTRACTION WITH ARTIFICIAL INTELLIGENCE 49
3.1 Introduction 50
3.2 Action steps 50
3.2.1 Background extraction 50
3.2.2 Mention all the metods: 50
3.2.3 Problems 52
3.3 Tools and datasets 52
3.3.1 Roboflow 53
3.3.2 Google colab 53
3.3.3 Python 53
3.3.4 Choosing the database: 54
3.3.5 Splitting data 54

3.3.6 Training models: 56
3.4 Result and discussion 56
3.4.1 Model yolov8 accuracy and loss : 58
3.4.2 Model sam accuracy and loss : 60
3.5 Comparison of results 61
3.6 Conclusion 61

GENERAL CONCLUSION 63

Bibliography 65

List of Figures

1.1 The biological neuron 7
1.2 Diagram artificial neural 8
1.3 The relationship between artificial intelligence, machine learning, and deep learning 14
1.4 History of Deep Learning. 15
1.5 YOLO model architecture 19
1.6 Divide image into (S*S) grids 19
1.7 The predicted vector in the case of a single box 20
1.8 Union on Intersection 21
1.9 Examples of IoU: courtesy 21
1.10 The predicted vector in the scenario of multiple boxes within the same cell 22
1.11 A tensor representing the bounding box coordinates and class probabilities. 23
1.12 The result of non-maximum suppression. 24

2.1 Some applications 38
2.2 Multi-View Image Synthesis with Implicit Texture Representation and Disentangled Shape Modeling 38
2.3 Differentiable Neural Radiance Fields for Image Synthesis and Scene Optimization 40
2.4 View-Dependent Radiance in Neural Scene Representations 43
2.5 Importance of View-Dependent Emitted Radiance and High-Frequency Positional Encoding in Scene Representation 44

2.6 Enhanced Scene Modeling and Rendering from Multi-View Images using Disentangled Representations and Implicit Textures 46
3.1 An example of background removal in an PHOTOSHOP . 51
3.2 An example of background removal in an AI Tools 52
3.3 Original photo 57
3.4 segmented image 57
3.5 background with segmented areas 57
3.6 Original photo 57
3.7 segmented image 57
3.8 background with segmented areas 57
3.9 Original photo 57
3.10 segmented image 57
3.11 background with segmented areas 57
3.13 Performance test results of the retrained YOLOv8 model on subclasses 57
3.14 Performance test results of the retrained SAM model on subclasses 58
(a) Class 1 tree 58
(b) Class 2 tree 58
(c) Class 3 tree 58
3.15 Accuracy and Loss Ratios Curve of the YOLOv8 Model . 59
3.16 Accuracy and Loss Ratios Curve of the SAM Model 60

List of Symbols

AI : Artificial Intelligence

SVM : Support Vector Machines

$SIFT$: Scale Invariant Feature Transform

$SURF$: Speeded Up Robust Feature

HOG : Histogram of Oriented Gradients

ML : Machine Learning

$CNNs$: Convolutional Neural Networks

$RNNs$: Recurrent Neural Network

$GANs$: Generative Adversarial Networks

$VAEs$: Variational Autoencoders

$GNNs$: Graph Neural Networks

$YOLO$: You Only Look Once,

IoU : Intersection over Union

KNN : K-Nearest Neighbors

NLP : Natural Language Processing

NMS : Non-Maximum Suppression

OOP : object-oriented programming

IR : Information Retrieval

CAD : computer-aided design

$mRNA$: Messenger Ribonucleic Acid

IoT : Internet of Things

VGA : Video Graphics Array

IBM : International Business Machines Corporation

API : Application Programming Interface

$UAVs$: Unmanned Aerial Vehicles

$COCO$: Common Objects in Context

VOC : Visual Object Classes

MLP : Multi-Layer Perceptron

$NeRF$: Neural Radiance Fields

RGB : Red, Green, Blue

$ReLU$: Rectified Linear Unit

SDF : Signed Distance Function

BTF : Bidirectional Texture Functions

SH : Spherical Harmonic

uv : Ultraviolet

ITR : Implicit texture representation

RM : Rendering module

GPU : Graphics Processing Unit

CPU : Central Processing Unit

SAM : Segment Anything Model

$U - Net$: U-Net Architecture (a convolutional neural network)

OAK : Object Aware Kernel

Sec : Section

$R\&D$: research and development

GENERAL INTRODUCTION

In recent years, machine learning and deep learning have experienced tremendous growth, introducing new techniques and technologies in data analysis and visualization. These works techniques include Neural Radiation Fields (NeRF), which shows the ability to represent and reconstruct 3D models with high accuracy from 2D models. This promising technology opens up many possibilities in various applications, such as animation, virtual reality, and image enhancement. In this article, we discuss how NeRF can be used in scenes that contain complex information, such as trees, which present a great challenge when trying to separate them from the background using traditional methods. In addition, we review the use of the Roboflow platform to manage data management and preparation, train deep models, and add transfer learning to improve performance with limited datasets. We also explore post-filtering issues related to the use of AI-powered tools and associated challenges, as well as appropriate data preparation and classification techniques to ensure reliable results. This note seeks to provide a holistic view of how AI can best be used to visualize and separate complex objects and backgrounds, and highlights new and innovative approaches that help push the boundaries of what is possible in this work. Neural Radiance Fields (NeRF) are a powerful tool in the 3D modeling field, but when used to create trees digitally, many unique challenges and issues arise that must be addressed to ensure modeling if it is accurate and true and integrated. These complications make it difficult for NeRF to distinguish and accurately represent parts of a tree. In addition, the closing and overlapping of branches and leaves can create obstacles that make it difficult for NeRF to determine the true composition of each segment Environmental variability also poses significant challenges; Lighting conditions change dramatically throughout the day and in different locations, causing significant variability in the images used to train the NeRF model, complicating accurate predictions of color and density complexity Also, weather is like wind and rain can change the appearance of the trees between photos, further increasing inconsistencies. Digitizing trees also requires fine-grained images to capture details, where enough storage space is used to train an image, it requires high computing power requiring multi-dimensional data collection, which can be challenging for large trees or those in harsh environments, which often require specialized equipment such as drones Also, NeRF Models face problems in training and modeling, as they require significant computational resources, especially in terms of data availability in dealing with high resolution and complex scenes such as trees Accurate self-planning is essential for NeRF models to ensure that tree parts are represented accurately

When digitizing trees using Neural Radiance Fields (NeRF) technology, background extraction becomes crucial for several reasons. By isolating trees from the rest of the environment, the accuracy of the model can be significantly improved and data complexity reduced. Firstly, noise in the training data can be minimized by removing irrelevant elements such as buildings, people, or other trees, allowing NeRF to focus solely on the tree's features. Isolating the tree also enables the model to better understand and capture the fine details of branches and leaves without distraction from other elements in the scene. Secondly, with the background removed, the input data becomes simpler and more consistent, facilitating the modeling process and reducing computational complexity, thus making the process faster and more efficient. Thirdly, lighting conditions around the tree can be standardized by removing the background, helping to create a more consistent model despite natural lighting changes throughout the day, and reducing the effects of shadows and reflections that could complicate the model's learning process. Fourthly, once the tree is modeled without a background, it can be easily inserted into various virtual environments, which is particularly useful for applications in animation, video games, and architectural simulations. Models without backgrounds are also easier to manipulate and edit, allowing for precise adjustments without unwanted interference. Lastly, accurate background extraction allows for better differentiation of the edges and details of branches and leaves, which is crucial for creating realistic models, and automatic segmentation algorithms can work more effectively when there is no complex background to analyze. In summary, background extraction is a critical step in digitizing trees using NeRF, as it enhances model accuracy, simplifies data, aids in managing lighting variations, facilitates post-processing, and allows for more precise segmentation. These combined benefits lead to more realistic and usable digital representations in various applications.

the organization of the manuscript : The manuscript consists of 3 chapters organized as follows: In the first chapter, we introduced the work of this thesis. The second chapter presents artificial intelligence in agriculture . In the third chapter, we study and discuss the workflow model and its main applications in the scientific field and its results

Chapter 1

OBJECT DETECTION TECHNIQUES

1.1 Introduction

Object detection is a crucial task in computer vision, enabling the identification and localization of objects within images or videos. With the rapid increase in image and video data, machine learning techniques have become popular for their ability to learn and accurately recognize objects. In particular, deep learning has revolutionized object detection by achieving significant advancements in both accuracy and efficiency. The main goal is to explore and analyze various machine learning algorithms and methodologies used in object detection, encompassing both traditional methods and deep learning-based approaches. The importance of object detection in computer vision applications cannot be overstated. It underpins several critical tasks, including video surveillance, autonomous driving, object tracking, and augmented reality. For these applications to function effectively and make informed decisions, accurate and efficient object detection is essential.[3]

1.2 Artificial intelligence

Artificial Intelligence (AI) is currently at the top of the cyclical hype cycle What is the purpose of artificial intelligence (AI)? Artificial intelligence has been entering our lives for decades, almost without realizing that, in many ways, this is a young field, having already started in 1956 with the Dartmouth Conference. [20]Artificial intelligence (AI) is defined as "a field of science and engineering that is concerned with the computational understanding of what is commonly called intelligent behavior.

1.2.1 Artificial technologies

AI technologies include several different areas and techniques, including:

- **Machine learning:** is centered around developing methods that en-able systems to learn models and bases from data and make decisions based on this learning, instead of using static programming.

- **Natural Language Processing:** is a method that permits computers to comprehend, analyze, and produce natural language, making interactions between humans and computers more natural.
- **Computer Vision :** The development of computer vision technologies enables computers to analyze, understand, and absorb images and videos in ways that are similar to human beings. Deep Learning Deep multi-layered neural networks are used in machine learning to represent data hierarchically, allowing more complex data to be understood and analysed.
- **Reinforcement Learning:** The concept of promoting correct behavior through rewards and penalties is the basis of machine learning patterns.

  These are just a few of the key technologies in artificial intelligence, and there are many more advanced technologies that are utilized in various applications.[2]

1.2.2 Neuronal environment

1. **Biological Neuron:** Biological Neuron Our brain contains over 100 billion neurons. A neuron receives signals from other neurons, integrates these signals, and then generates a new signal that it sends to additional neurons. Biological neurons consist of three primary components: dendrites, cell bodies, and axons (see Figure 1.1). Dendrites form a network of receptors that transmit electrical signals from other neurons to the neuron's body. This body functions as an integrator, accumulating electrical charges. When the neuron becomes

sufficiently excited (i.e., when the accumulated charge surpasses a certain threshold), it generates an electrical potential through an electrochemical process, which it then propagates through its axons to stimulate other neurons. Artificial neural networks.[14]

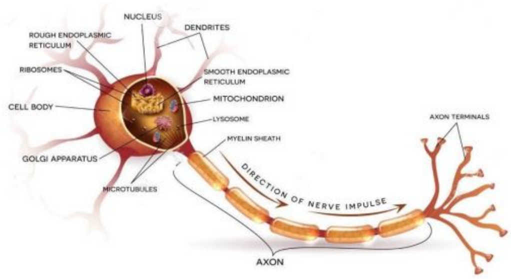


Figure 1.1: The biological neuron

2. **ARTIFICIAL NEURON:** An artificial neuron functions as a simplified representation or abstraction of biological neurons through mathematical operations. Within an artificial neural network, these neurons serve as fundamental components. Each artificial neuron accepts one or more inputs, akin to the dendrites of biological neurons, and aggregates them to generate an output, akin to the axon of a biological neuron. Typically, the inputs are weighted before aggregation, and the resultant sum undergoes a non-linear transformation, often facilitated by an activation function. While activation functions commonly exhibit a sigmoidal shape, they may also manifest as other non-linear functions. Formally, an artificial neuron can be defined as a computational element with 'n' inputs, where 'n' corresponds to either external neurons or outputs from other neurons, and it yields a singular output. This model finds mathematical expression through the following equations:[10]

$$Y = f(Xi * Wi) \tag{1.1}$$

Where w(k), xi, f(k), y are respectively, the synaptic weights , the inputs, the activation function and neuron output. We can summarize the difference between artificial and biological neuron in the(figure 1.2)

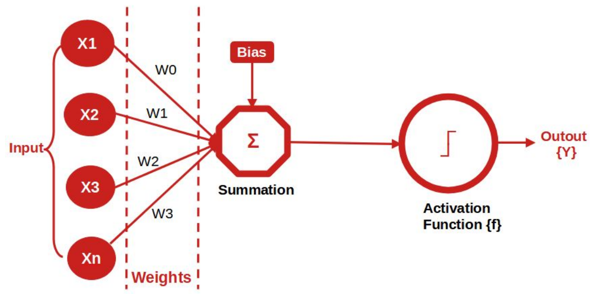


Figure 1.2: Diagram artificial neural

1.3 Machine learning

Machine learning, a branch of artificial intelligence, is the programming of computers to optimize a performance criterion using idealized data or previous experience The goal of machine learning is to develop methods that can automatically detect patterns in data, and then use the detected patterns to predict future data or other outcomes of interest.[5]

1.3.1 Machine learning algorithm:

A machine learning algorithm is a search process that selects the best function that best describes the relationships between features in a dataset, in order to achieve an intuitive understanding of the information extracted or learned from the data. The machine learning process usually consists of two main steps: Training and Inference. During the training phase, the machine learning algorithm processes the dataset and selects the function that best fits the patterns detected in the data[12]. The extracted function is then encoded in a specific form, such as a set of rules or specific equation parameters; this encoded function is known as a model. The process of analyzing the data to extract the function from the model is often referred to as model training.

1.3.2 List of common machine learning algorithms :

Here is the list of commonly used machine learning algorithms. These algorithms can be applied to almost any data problem:[1]

(a) Linear Regression.

(b) Logistic Regression.

(c) Decision Tree.

(d) SVM.

(e) Naive Bayes.

(f) KNN.

(g) K-Means.

(h) Random Forest .

(i) Dimensionality Reduction Algorithms.

(j) Gradient Boost and Adaboost.

1.3.3 Notation of dataset

Before delving deeply into machine learning, we first describe the notation of the dataset, which will be used throughout this section as well as the tutorial.[5] There are two general dataset types:

Labeled dataset:

$$X = \left\{x^{(n)} \in R^d\right\}_{n-1}^{N}, \quad Y = \left\{y^{(n)} \in R\right\}_{n-1}^{N}. \tag{1.2}$$

- X : Feature set containing N samples.
- x $\in R^n, wherendenotesthefeaturevector. Y Labelset, denotedasy \in R.$

- Each sample is a d-dimensional vector: x = [x1, x2, ..., xn] ^{T}.
- Each dimension of a vector is called an attribute, feature, variable, or element.
- In some applications, the label set is unobserved or ignored.
- Another form of the labeled dataset is described as (x, y), where each (x, y) pair is called a data pair.

Unlabeled dataset:

$$X = \left\{x^{(n)} \in R^d\right\}_{n-1}^{N} \tag{1.3}$$

- X Feature set containing N samples.
  - x $\in R^n, wherendenotesthefeaturevector.$

1.3.4 Training set and test set:

In machine learning, it is assumed that there is an unknown global dataset, which contains all possible pairs of data as well as their probability distribution in the real world. However, in reality, we usually work with only a subset of the dataset, due to memory constraints or other reasons. This acquired subset is called the training set, and is used to gain features and knowledge about the overall dataset.

Generally, it is assumed that the vectors in the training set are independently and identically (i.i.d.) drawn from the overall dataset.[5]

1.4 Type of machine learning

- **Supervised learning:**

Network is provided with a correct answer (output) for every input pattern(presence of a master who provides the desired response). Besides weights are determined to allow the network to produce answers as close as possible to the known correct answers. The back-propagation algorithm belongs into this category. This procedure is repeated until a performance criterion is satisfied. Once the learning procedure is completed, the synaptic coefficients take optimal values with regard to the stored configurations and the network can be operational

- **Unsupervised learning:**

  This type of learning does not require a correct answer associated with each input pattern in the training set and the learning procedure is based only on input values. It explores the underlying structure in the data, or correlations between patterns in the data and organizes patterns into categories from these correlations. The network is self-organized in such a way as to optimize a certain cost function, without being given the answer desired. This property is called self-organization. The Kohonen algorithm belongs into this category.

- **Semi-supervised:**

  Semi-supervised learning can be defined as a hybridization of the above-mentioned supervised and unsupervised methods, as it operates on both labeled and unlabeled data , Thus, it falls between learning "without supervision" and learning "with supervision". The ultimate goal of a semi-supervised learning model is to provide a better outcome for prediction than that produced using the labeled data alone from the model.

- **Reinforcement learning**

  Reinforcement learning is a type of machine learning algorithm that enables software agents and machines to automatically evaluate the opti mal behavior . its ultimate goal is to use insights obtained from environmental activists to take action to increase the reward or minimize the risk . It is a powerful tool for training AI models that can help increase automation or optimize the operational efficiency of sophisticated systems such as robotics, autonomous driving tasks, manufacturing and supply chain logistics, however, not preferable to use it for solving the basic or straightforward problems.

1.5 Traditional machine learning

Traditional machine learning encompasses a suite of mathematical, statistical, and computational techniques designed to develop algorithms capable of solving problems by identifying patterns within diverse input data. Unlike direct formulas, the solution is derived from the established relationships between specific features and their values, rather than from predetermined calculations. Unlike neural networks, traditional machine learning relies on feature extraction procedures. While neural networks autonomously extract features during training, traditional methods employ feature detection algorithms such as Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), Histogram of Oriented Gradients (HOG), among others. These algorithms identify key points and generate feature vectors essential for subsequent classification processes, encompassing characteristics like corners, color schemes, and image textures.

– **SIFT:**

  SIFT employs mathematical approximations to create scale-invariant image representations, effectively standardizing images. By detecting important features within these standardized images, SIFT generates feature vectors that encode the existence of these critical points.

– **SURF :**

Similar to SIFT, SURF operates on the principles of feature detection, local area description, and matching. However, it utilizes block blur to quickly approximate Gaussian differences, enhancing efficiency in points of interest detection and description.

– **HOG :**

HOG describes object appearance and shape through the distribution of intensity gradients or edge directions within an image. By dividing the image into cells and compiling histograms of directional gradients, HOG achieves robustness to lighting variations and improves accuracy through contrast normalization.

Following feature selection and extraction, a data classification model is constructed using various algorithms like Support Vector Machines (SVM), k-nearest neighbors, and Random Forests, among others. The selection of an appropriate algorithm depends on the specific task at hand and often requires domain expertise. To streamline this process, the authors suggest leveraging automatic machine learning (AutoML) techniques to determine the optimal algorithm.

Currently, traditional computer vision algorithms find suitability in simpler tasks, particularly when computational resources or datasets are limited. Applications of machine learning algorithms span across domains including robotics, augmented reality, automatic panorama stitching, virtual reality, 3D modeling, motion estima-

tion, video stabilization, motion capture, video processing, and scene understanding.[11]

1.6 Deep learning

1.6.1 Introduction to deep learning:

In the realm of artificial intelligence, deep learning stands out as a pivotal subfield, dedicated to constructing expansive neural network architectures engineered to yield precise, data-centric determinations. Deep learning thrives in environments characterized by intricate datasets and ample data reservoirs. Presently, a multitude of digital enterprises and cutting-edge consumer technologies heavily leverage deep learning methodologies. Notably, Facebook employs deep learning algorithms to scrutinize textual content within online discourse. Similarly, behemoths like Google, Baidu, and Microsoft harness deep learning prowess for image retrieval and machine translation services. [9]

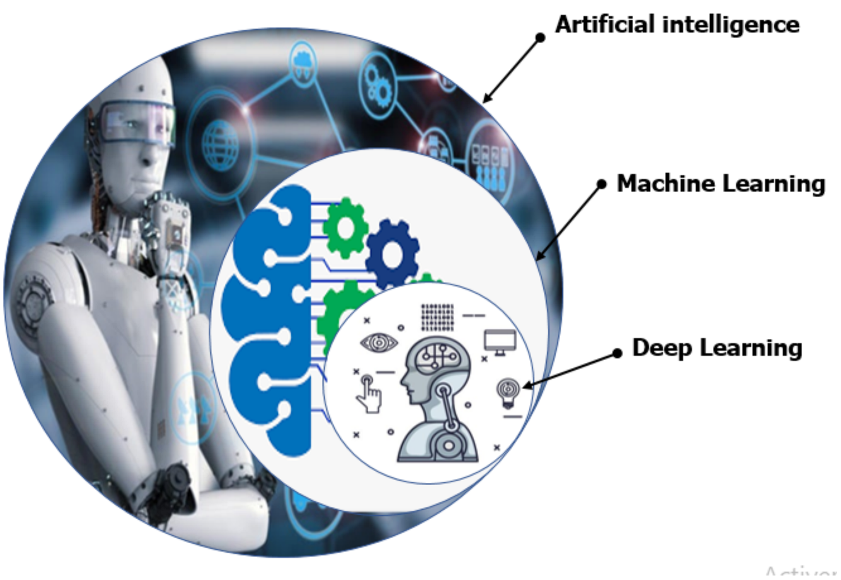


Figure 1.3: The relationship between artificial intelligence, machine learning, and deep learning

1.6.2 A brief history of deep learning:

Figure 1.4 illustrates a timeline depicting the historical milestones in the development of deep learning, emphasizing significant periods of research. These include investigations into threshold logic units (early 1940s to the mid-1960s), the era of connectionism (early 1980s to mid-1990s), and the emergence of deep learning (mid-2000s to the present). Additionally, the upper section of Figure 1.4 showcases pivotal conceptual advancements, training algorithms, and model architectures that have propelled the evolution of deep learning. Notably, the two gray rectangles in the figure symbolize the developmental phases of two crucial deep learning network architectures: Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). [9]

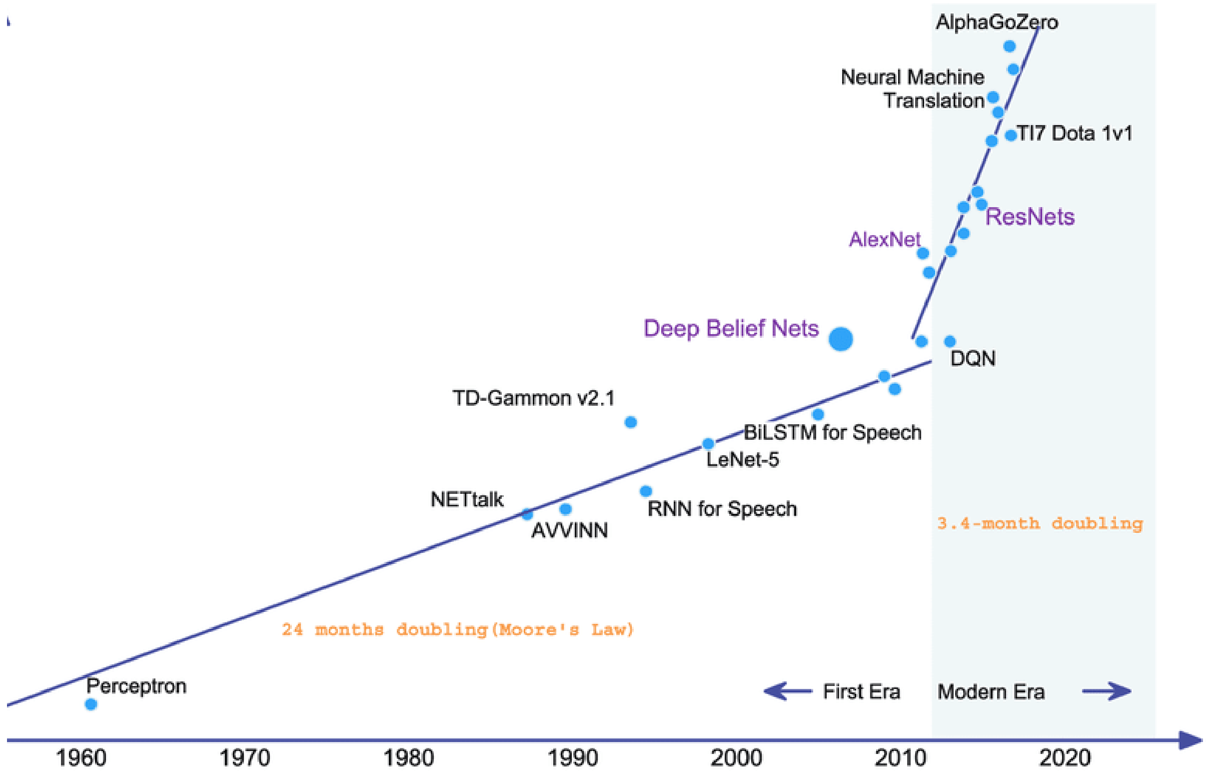


Figure 1.4: History of Deep Learning.

1.6.3 Reasons for the success of deep learning:

In any data-driven endeavor, the key to success lies in understanding what to measure and how to measure it effectively. This underscores the critical significance of feature selection and feature design within the domain of machine learning. As previously discussed, these tasks necessitate domain expertise, statistical analysis, and iterative experimentation to construct models with varying feature sets. Consequently, dataset design and preparation can consume a substantial portion of time and resources allocated to a project, sometimes accounting for up to 80Feature design stands out as a domain where deep learning demonstrates a significant advantage over traditional machine learning approaches. In conventional machine learning, feature design often demands considerable human effort. Deep learning diverges from this approach by autonomously learning the most relevant features directly from the raw data, thereby mitigating the need for extensive manual intervention. [9]

1.7 AI model

1.7.1 What is the ai model:

Artificial intelligence consists of intelligent models that are considered an essential part of artificial intelligence and machine learning applications. These models are used as computational representations to predict future events, make decisions, and perform specific tasks. These models are the driving force behind innovation in the field of artificial intelligence. They have the ability to process vast amounts of data and solve complex problems. These models rely on techniques such as computer vision, natural language processing, and machine learning to analyze complex data patterns.[13] Additionally, artificial intelligent models benefit from decision-making algorithms in learning processes. These models advance towards achieving their specific activities and goals through training processes, data collection, and analysis.

1.7.2 Types of ai models:

AI models encompass a wide array of algorithms and structures tailored to fulfill specific tasks or tackle particular challenges. Here are some prevalent types of AI models:

(a) **Machine learning models:** These models discern patterns and correlations from data to make forecasts or determinations without explicit programming. Examples comprise linear regression, decision trees, support vector machines, and random forests.

(b) **Deep learning models:** A subset of machine learning models, these employ neural networks with multiple layers to extract hierarchical representations from data. They shine in tasks like image and speech recognition, natural language processing, and reinforcement learning.

(c) **Reinforcement learning models:** Here, an agent learns to engage with an environment by taking actions and receiving feedback in the form of rewards or penalties. They find common application in robotics, game playing, and autonomous vehicle navigation.

(d) **Natural language processing (NLP) models:** These models specialize in comprehending and generating human language. They incorporate algorithms for tasks like text categorization, sentiment analysis, language translation, and question answering.

(e) **Computer vision models:** Concentrating on processing and interpreting visual data like images and videos, these models are utilized for tasks such as object detection, image classification, image segmentation, and facial recognition.

(f) **Generative models:** These models produce new data samples resembling the training data. Examples include generative adversarial networks (GANs) for creating realistic images and variational autoencoders (VAEs) for generating diverse outputs.

(g) **Graph neural networks (GNNs):** Operating on graph-structured data like social networks or molecular structures, these models are employed for tasks such as node classification, link prediction, and graph generation

(h) **Bayesian models:** These models leverage probabilistic reasoning to make forecasts or decisions, particularly valuable in scenarios with uncertainty and incomplete information. These exemplify only a fraction of the extensive spectrum of AI models utilized across industries for diverse applications. Each model type has its merits and demerits, and the selection of a model hinges on the precise requirements of the problem at hand.[17]

1.7.3 Model yolo :

YOLO, short for "You Only Look Once," is one of the popular object detection algorithms used by researchers worldwide. It was first described in 2015 in the paper by Joseph Redmon et al. The network utilizes features from the entire image to predict each bounding box. It also predicts all bounding boxes of all classes in an image simultaneously.This means the network reasons globally over the entire image and all objects it contains. The YOLO design allows for end-to-end learning and real-time speeds while maintaining high average precision. The architecture of this network is inspired by the GoogLeNet model for image classification. It consists of 24 convolutional layers followed by 2 fully connected layers. Instead of the inception modules used by GoogLeNet, it simply employs 1x1 reduction layers followed by 3x3 convolutional layers.

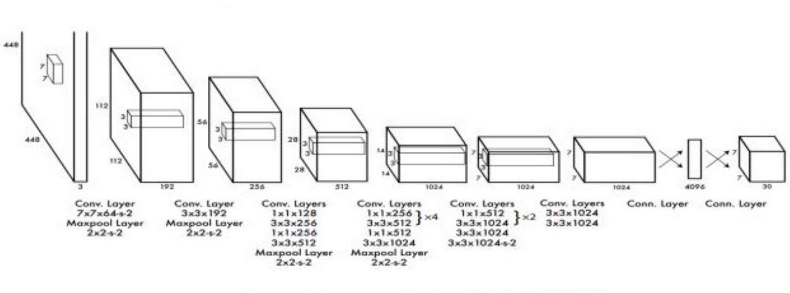


Figure 1.5: YOLO model architecture

- **Divide the image into cells with size S x S:**

  The image undergoes segmentation into a grid of S x S dimensions (for instance, 3 x 3), resulting in a total of N cells. Each cell within this grid is tasked with object detection.

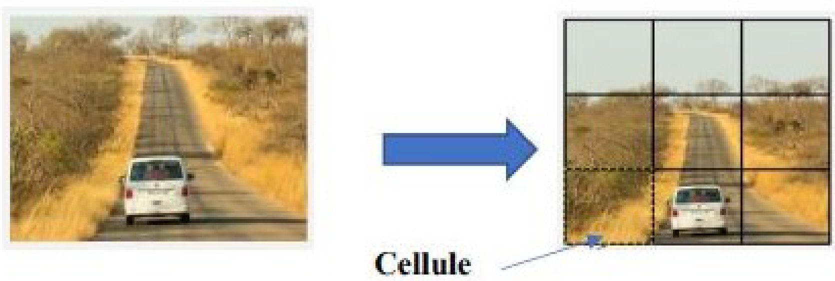


Figure 1.6: Divide image into (S*S) grids

- **Each cell predicts B bounding boxes:**

After partitioning the image into N cells, every cell within the grid forecasts B bounding boxes along with confidence scores for these boxes. Each bounding box comprises five predictions: x, y, w, h, and confidence. The (x, y) coordinates denote the box's center in relation to the grid cell boundaries, while the width and height are forecasted relative to the entire image. Ultimately, the confidence prediction indicates the Intersection over Union (IoU) between the forecasted box and any ground truth box.[14]

Example:

In a scenario with a grid of 3x3 cells (S=3), each cell makes a prediction for 1 bounding box (B=1), and identifies objects as either dogs (C=1) or humans (C=2). In this setup, the Convolutional Neural Network (CNN) generates a vector Y for each cell.

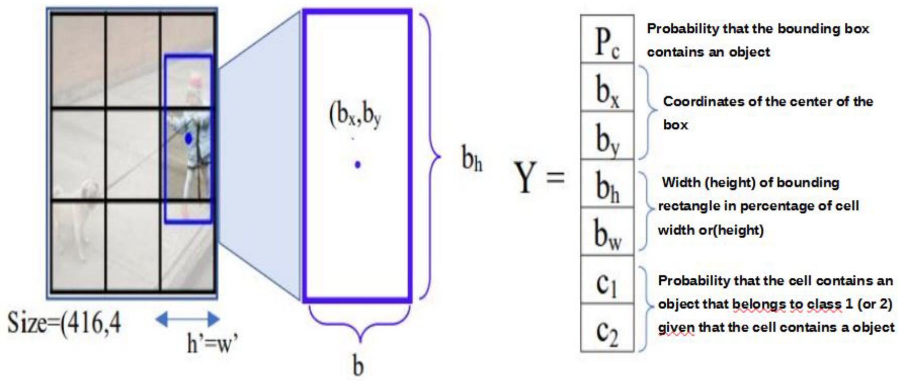


Figure 1.7: The predicted vector in the case of a single box

The elements of vector Y follow the YOLO format:

Pc: the confidence prediction, signifies the Intersection over Union (IoU) between the predicted box and the ground truth box.

$$b_z = (x - h')/h', b_y = (y - w')/w' \tag{1.4}$$

$\mathrm{b}_h = h/416, b_w = w/416$

A) Intersection over Union (IoU): Intersection over Union (IoU) serves as the standard evaluation metric in object detection tasks. It distinguishes between true positives and false positives within a set of predictions. To utilize IoU effectively as an evaluation criterion, a precision threshold needs to be defined. This metric assesses the overlap between the predicted bounding box and the ground truth bounding box, calculated by comparing their respective areas. During training, the IoU confidence

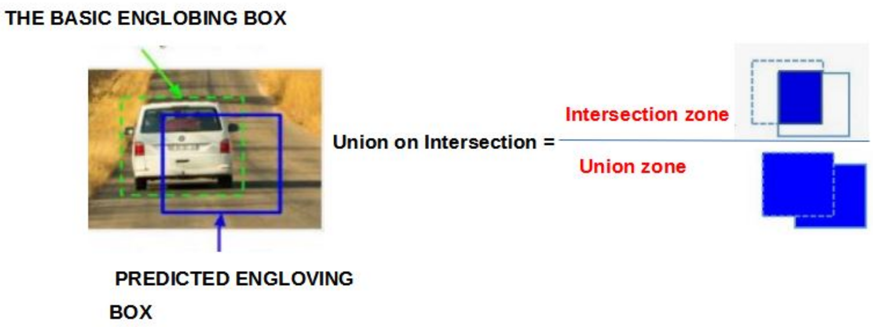


Figure 1.8: Union on Intersection

score is computed between the predicted box and the ground truth box. In the figure above, there are examples of good and bad Intersection over Union scores. As you can see, predicted bounding boxes that heavily overlap with the ground truth bounding boxes have higher scores than those with less overlap.

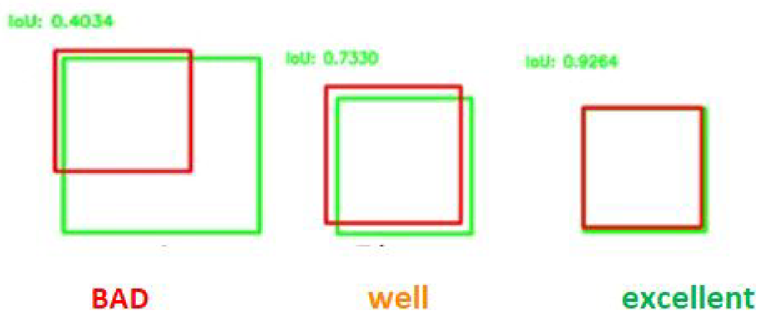


Figure 1.9: Examples of IoU: courtesy

B) Anchor box: In the earlier example, we considered the prediction of a single bounding box. However, when multiple bounding boxes exist within the same cell, the YOLO algorithm employs Anchor Boxes to address this scenario. As previously discussed, each cell corresponds to a vector representation. In situations where multiple boxes are present within a single cell, we expand the vector in the following manner:

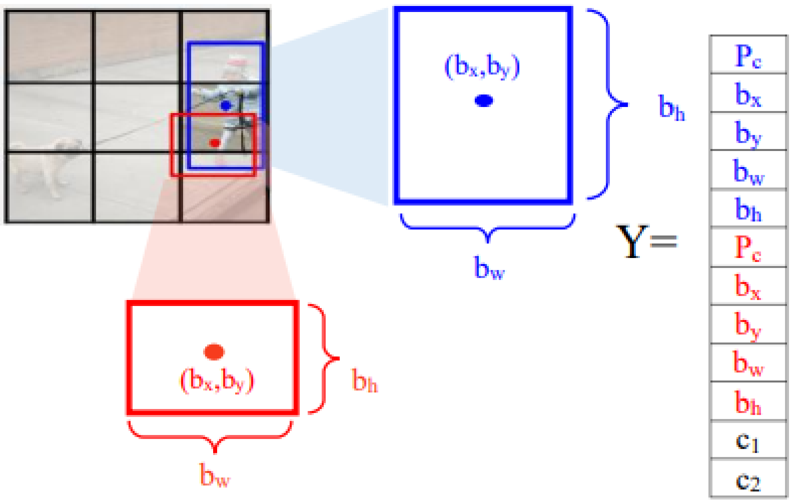


Figure 1.10: The predicted vector in the scenario of multiple boxes within the same cell

In general, if we divide the image into an S × S grid and predict B bounding boxes, along with their respective confidences and class probabilities C for each grid cell, these predictions are encoded as a tensor.

$$S \times S \times (B * 5 + C). \tag{1.5}$$

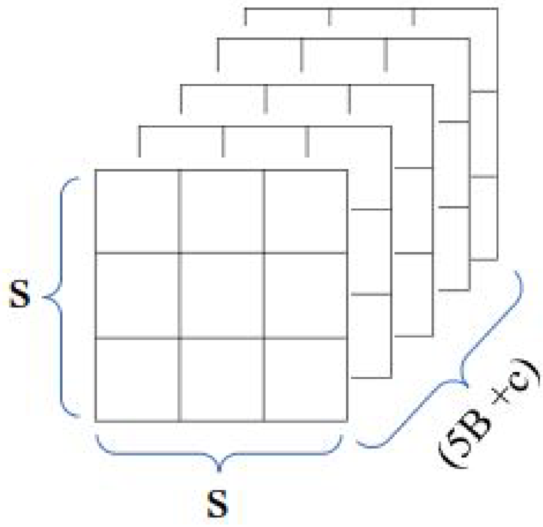


Figure 1.11: A tensor representing the bounding box coordinates and class probabilities.

C) Non-maximum suppression (NMS):

Non-Maximum Suppression serves as the concluding stage in the detection algorithm, especially when multiple bounding boxes detect the same object within an image. This method is utilized to eliminate redundant bounding boxes, ensuring that only the most probable one is retained. The technique involves a five-step process:

- **Step 1:**Select the bounding box with the highest objectiveness score.
- **step 2:** Next, compare the overlap (intersection over union) of this box with other boxes.
- **step 3:**Remove bounding boxes whose overlap (intersection over union) is :50 %
- **step 4:**Proceed to the next highest objectiveness score.
- **Step 5:**Finally, repeat steps 2 to 4 until all objects in the image are processed.

There have been 8 versions of the model so far, with each new version improving upon the previous one in terms of speed and accuracy. [14]

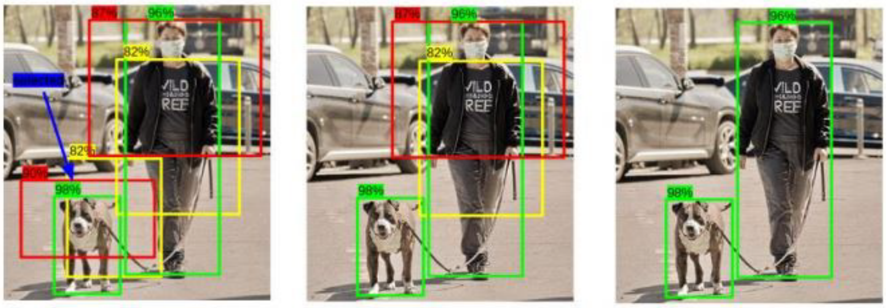


Figure 1.12: The result of non-maximum suppression.

1.8 Transfer learning

Transfer learning is driven by the observation that individuals can effectively leverage previously acquired knowledge to tackle new challenges more efficiently or effectively. For instance, proficiency in C++ programming can facilitate rapid learning of Java, given their shared object-oriented programming (OOP) principles. The concept of transfer learning, also known as learning of transfer, delves into the interdependence of human behavior, learning, or performance on past experiences. Over a century ago, researchers began exploring how individuals could transfer skills from one context to another with similar characteristics. In the realm of machine learning, the core rationale for transfer learning lies in the necessity for continual learning approaches that can preserve and apply previously gained knowledge, enabling intelligent systems to adapt to new environments or tasks with minimal human intervention. Informally, transfer learning in machine learning refers to a system's capacity to recognize and utilize knowledge and competencies acquired from previous domains or tasks to address new or unfamiliar domains that share certain commonalities.

1.8.1 Types of transferlearning:

1.Homogeneous transfer learning: Homogeneous transfer learning involves transferring knowledge from a source domain to a target domain where both domains share similar feature spaces and distributions. This means that the input data and tasks in both domains are either identical or very similar. In such cases, knowledge acquired from the source domain can be directly applied to the target domain without substantial modifications. Homogeneous transfer learning is typically easier to implement as there is a high degree of similarity between the domains. **2. Heterogeneous transfer learning:** On the other hand, heterogeneous transfer learning deals with situations where the source and target domains possess different feature spaces or distributions. This implies that the input data or tasks in the two domains are not directly compatible. Heterogeneous transfer learning aims to transfer knowledge across dissimilar domains by aligning or transforming the data or features from the source domain to suit the target domain. This often involves employing more sophisticated techniques such as domain adaptation, which accounts for the differences between domains during the transfer process. To summarize, homogeneous transfer learning assumes domain similarity and straightforwardly transfers knowledge, while heterogeneous transfer learning addresses cases where domains are dissimilar and necessitates techniques to align or adapt knowledge from the source domain to the target domain.[19]

1.8.2 Utilizations of transfer learning:

Transfer learning has recently found successful applications in various classification challenges across different domains, including Natural Language Processing (NLP), Information Retrieval (IR), recommendation systems, computer vision, image analysis, multimedia data mining, bioinformatics, activity recognition, and wireless sensor networks.

1. NLP implementations: In the realm of NLP, transfer learning, often referred to as domain adaptation, has been extensively researched to address diverse tasks such as named entity recognition, part-of-speech tagging, sentiment classification, sentiment lexicon construction, word sense disambiguation, coreference resolution, and relation extraction.

2.Web-based utilizations: Information Retrieval (IR) stands as another significant area where transfer learning methodologies have been broadly explored and applied. Common web-based applications include text classification, advertising optimization, learning to rank, and recommender systems

3. Sensor-driven implementations: Transfer learning has also been investigated for solving problems related to WiFi-based localization and sensor-based activity recognition. For instance, techniques have been developed to transfer WiFi-based localization models across different time periods, spatial locations, and mobile devices. Researchers have proposed transfer learning methods to address indoor sensor-based activity recognition challenges.

4 .Applications in computer vision: In the past decade, the adoption of transfer learning techniques has surged in computer vision, image analysis, and multimedia understanding. These applications encompass various tasks such as image classification, image retrieval, facial verification, age estimation from facial images, image semantic segmentation, video retrieval, video concept detection, event recognition from videos, and object recognition.

5. Bioinformatics applications: In Bioinformatics, transfer learning approaches have been explored to tackle computational biological problems, including identifying molecular associations of phenotypic responses, recognizing splice sites in eukaryotic genomes, predicting mRNA splicing patterns, determining protein subcellular locations, and analyzing genetic associations.

6. Other utilizations: Beyond the aforementioned applications, research efforts have examined the transfer of domain knowledge for learning relational action models across different domains in automated planning. Additionally, transfer learning methodologies have been applied to solve challenges in robotics, brain-computer interfaces, computer-aided design (CAD), and software engineering, such as solving the inverse dynamics problem for robotic manipulators, cross-project defect prediction, and learning conceptually related classifiers for CAD.[19]

1.9 Conclusion

In conclusion, object detection combines various aspects of artificial intelligence and machine learning, from traditional methods to advanced deep learning and transfer learning techniques. The ability of these systems to accurately identify and locate objects is crucial for a multitude of applications, from safety to autonomous driving. Continued progress in this area promises to make AI systems even more powerful and efficient in the future.

Chapter 2

ARTIFICIAL INTELLIGENCE IN AGRICULTURE

2.1 Introduction

Agriculture stands as one of humanity's oldest professions, evolving significantly from early civilizations to the present day. Throughout history, advancements in agricultural practices and techniques have played a crucial role in sustaining the growing population's food needs. However, with each passing generation, we have witnessed a concerning trend: a steady increase in population accompanied by a decline in agricultural production, reaching a 20 %deficit. As the gap between food demand and production widens, there arises an urgent need for innovative solutions to address this challenge. Artificial Intelligence (AI) emerges as a powerful tool capable of revolutionizing agricultural practices by automating food crop production processes, thereby augmenting food production to meet the demands of a growing population. AI solutions offer a plethora of capabilities essential for optimizing agricultural operations. They can analyze fields to optimize crop harvesting, continuously monitor crop and soil conditions, and provide accurate weather forecasts for informed decision-making. By leveraging AI in agriculture, farmers can access real-time insights and personalized data-driven recommendations, empowering them to enhance farming techniques and maximize yields. [16]

2.2 Importance

Artificial Intelligence (AI) holds immense potential to revolutionize various fields, and agriculture is no exception. By integrating AI-powered solutions, we can redefine traditional farming practices and bring about significant advancements. These solutions not only empower farmers to achieve more with less but also enhance the quality of crops and accelerate their time-to-market. In today's rapidly evolving technological landscape, AI, Big Data, and the Internet of Things (IoT) are driving digital transformation across all sectors. Therefore, it is imperative to leverage digital solutions augmented with AI to uplift the farming community while creating new opportunities for businesses and entrepreneurs. By offering smart farming as a service, we can address the challenges faced by farmers and pave the way for sustainable and efficient agricultural practices. [16]

Growth driven by internet of things (IOT):

Digital transformation is reshaping the agricultural landscape, with IoT technologies playing a pivotal role in revolutionizing food production. By leveraging IoT solutions, farmers can harness both structured and unstructured data to gain valuable insights into various aspects of agriculture. A vast amount of data is generated daily, encompassing historical weather patterns, soil reports, research findings, rainfall records, pest infestations, and imagery from drones and cameras. Cognitive IoT solutions have the capability to analyze this data comprehensively, providing actionable insights to optimize crop yield. Proximity sensing and remote sensing are two key technologies used for intelligent data fusion. One practical application of this high-resolution data is soil testing, where remote sensing involves sensors integrated into airborne or satellite systems, while proximity sensing requires sensors in direct contact with the soil. Hardware solutions, such as robotics paired with data-collecting software, are already enhancing agricultural practices. For instance, robots equipped with sensors can analyze soil composition and recommend the optimal fertilizer blend for specific crops, maximizing output. IoT devices, equipped with transducers to measure various environmental and crop parameters, are crucial components in this digital transformation. These devices can be mounted on protected mini boards with WiFi capability, microcontrollers, low-cost VGA image sensors, and mini battery packs powered by micro solar panels. Data collection can occur at predetermined intervals using WiFi hot spot towers strategically placed throughout the field or via drones equipped with active WiFi hot spots, enabling data capture and aerial imaging of the entire agricultural area. By leveraging data-driven farming techniques, which involve analyzing and correlating data on weather conditions, seed varieties, soil quality, disease probabilities, historical data, market trends, and pricing information, farmers can make informed decisions to optimize their agricultural practices. [16]

Image-based insight generation:

Precision farming has emerged as a prominent topic in modern agriculture, with technologies like drones and computer vision playing pivotal roles in optimizing crop management practices. By leveraging these advancements, farmers can enhance field analysis, crop monitoring, and decision-making processes. Drone-based imagery enables in-depth field analysis by providing high-resolution views of agricultural landscapes. Coupled with computer vision technology and IoT sensors, these images facilitate rapid data analysis and actionable insights for farmers. Real-time alerts generated from drone data feed empower farmers to take timely actions, enhancing the efficiency of precision farming practices. Companies like Aerial Tronics have integrated advanced technologies, such as IBM Watson IoT Platform and Visual Recognition APIs, into commercial drones for real-time image analysis. This enables functionalities like disease detection, pest identification, and nutrient deficiency recognition, enhancing crop health management. Computer vision technology also aids in crop readiness identification, where images captured under different lighting conditions assess the ripeness of fruits. By categorizing crops based on readiness levels, farmers can optimize harvesting schedules and improve market readiness. Furthermore, high-definition images obtained from airborne systems allow for real-time field management. By creating field maps and identifying areas requiring specific interventions like water, fertilizer, or pesticides, farmers can optimize resource allocation and maximize crop yields. In summary, the integration of drone-based imagery, computer vision technology, and IoT sensors holds immense potential for revolutionizing precision farming practices, ultimately leading to increased productivity and sustainability in agriculture. [16]

Identification of optimal mix for agronomic products:

Cognitive solutions are revolutionizing crop selection by analyzing various parameters such as soil condition, weather forecast, seed types, and pest infestation. By processing this data, these solutions provide personalized recommendations to farmers, ensuring the optimal choice of crops and hybrid seeds. Farmers can further tailor these recommendations to suit their specific needs, taking into account local conditions and historical farming data. This level of personalization enhances decision-making and maximizes crop yield. Moreover, cognitive solutions also consider external factors like marketplace trends, prices, and consumer demand. By incorporating these insights, farmers can make well-informed decisions that align with market dynamics and maximize profitability. In summary, cognitive solutions empower farmers to make strategic decisions in crop selection, leading to improved productivity and sustainability in agriculture. [16]

Health monitoring of crops:

Remote sensing techniques, coupled with hyperspectral imaging and 3D laser scanning, are indispensable for creating crop metrics over vast expanses of farmland. This technology holds the potential to bring about a revolutionary shift in how farmers monitor their land, saving time and effort significantly. Remote sensing technology enables farmers to monitor their crops throughout their entire life cycle, from planting to harvest. Furthermore, remote sensing technology facilitates the generation of automated reports, ensuring that farmers have access to real-time information about their crops. In summary, remote sensing technology represents a transformative tool for agricultural land monitoring, offering farmers invaluable insights and efficiency gains. [16]

Automation techniques in irrigation and enabling farmers : In terms of human intensive processes in farming, irrigation is one such process. Machines trained on historical weather pattern, soil quality and kind of crops to be grown, can automate irrigation and increase overall yield. With close to : 70 % of the world's fresh water being used in irrigation, automation can help farmers better manage their water problems. [16]

Drone based technology : Drones are transforming agriculture, offering innovative solutions to major challenges throughout the crop cycle. Here are six key applications of drone technology in agriculture: Soil and Field Analysis: Drones produce precise 3D maps for early soil analysis, aiding in seed planting planning and providing data for irrigation and nitrogen level management. Planting: Startups have developed drone-planting systems that reduce planting costs by : 85 %. These systems deploy pods containing seeds and nutrients into the soil, ensuring optimal conditions for crop growth. Crop Spraying: Drones scan the ground, enabling real-time spraying for uniform coverage. Aerial spraying with drones is five times faster than traditional machinery, increasing efficiency. Crop Monitoring: Drones provide efficient crop monitoring, offering time-series animations to track crop development and identify production inefficiencies, enabling better management practices. Irrigation: Sensor-equipped drones identify areas of the field requiring irrigation, optimizing water usage and improving crop health. Health Assessment: Drone-carried devices scan crops using visible and near-infrared light to track changes in plant health and detect diseases. This data helps farmers make informed decisions to maintain crop health. In the future, autonomous swarms of drones may collect data and perform tasks independently, revolutionizing agricultural practices. However, the development of high-quality sensors and advanced data processing software remains a significant challenge to realizing this vision. [16]

Models for farmers services :

Farmers can benefit from a range of service models tailored to their needs:

(a) **Chatbot:** AI-powered chatbots offer virtual assistance to farmers, providing answers and recommendations for specific queries. Using supervised and reinforced machine learning techniques, chatbots interactively address generic questions before escalating unique queries to human operators. [16]

(b) **Agri-e-calculator:** This smart application assists farmers in selecting the most suitable crops based on various factors. Farmers input their preferred crop and farm coverage area, and the calculator automatically estimates fertilizer, water, seed, equipment, labor, crop yield, market prices, and profitability using machine learning techniques and external data sources. [16]

(c) **Crop care service:** From seed sowing to harvesting, this service provides guidance using data from IoT sensors and external sources. Artificial intelligence analyzes the data and generates corrective actions using PID controller mechanisms, alerting farmers via smartphone based on severity and urgency. [16]

(d) **Price Prediction and Market Guidance:** Farmers receive predictive price and demand information throughout the crop lifecycle, helping them plan market releases and mitigate price fluctuations. [16]

(e) **Crop loan and insurance service:**This service assists farmers in securing crop loans and insurance by providing eligibility criteria, loan limits, and processing support based on smart estimations for proposed crops, ensuring mitigation against crop failures due to uncertainties or calamities. These service models empower farmers with technology-driven solutions, enhancing productivity, profitability, and resilience in agriculture.

[16]

2.2.1 Impact of ai in agriculture

Technologies powered by artificial intelligence (AI) play a crucial role in enhancing efficiency across various industries, including agriculture. They address challenges such as crop yield, irrigation, soil content sensing, crop monitoring, weeding, and crop establishment. Agricultural robots, specifically designed to leverage AI, offer valuable applications in this sector. With the global population on the rise, the agricultural industry is under pressure to meet increasing demands. However, AI presents opportunities for much-needed solutions. AI-driven technological solutions empower farmers to achieve higher output with fewer resources while improving the quality of their produce. This leads to faster time-to-market for crops. By 2020, it is projected that farmers will utilize 75 million connected devices. Furthermore, by 2050, the average farm is expected to generate an average of 4.1 million data points every day. AI has made significant contributions to the agricultural sector in various ways, including: Optimizing crop yield and resource utilization. Enhancing crop quality and marketability. Streamlining farming processes through automation. Providing real-time insights for informed decision-making. Improving sustainability practices and environmental impact. These advancements highlight the transformative potential of AI in agriculture, paving the way for a more efficient, sustainable, and resilient industry. [18]

Image recognition and perception: highlighted a recent surge in interest surrounding autonomous Unmanned Aerial Vehicles (UAVs) and their multifaceted applications. These applications encompass recognition and surveillance, human body detection and geolocalization, search and rescue operations, and forest fire detection. The growing popularity of drones or UAVs can be attributed to their versatility and advanced imaging capabilities, which range from delivery services to photography. With the ability to be piloted remotely and their agility in the air, drones are increasingly favored for reaching great heights and distances, facilitating a multitude of tasks. [18]

Skills and workforce: emphasized the role of artificial intelligence (AI) in empowering farmers to aggregate vast amounts of data from government and public sources, analyze it comprehensively, and offer solutions to various complex issues. AI also enables smarter irrigation practices, leading to increased yields for farmers. As AI continues to evolve, farming is expected to blend technological and biological skills, resulting in improved quality outcomes and reduced losses and workloads for farmers. The United Nations predicts that by 2050, twothirds of the world's population will reside in urban areas, underscoring the importance of alleviating pressure on farmers. AI applications in agriculture can automate processes, mitigate risks, and streamline farming practices, offering farmers a more efficient and manageable approach to agriculture. [18]

Maximize the output: asserted in their work that variety selection and seed quality establish the maximum performance level for all plants. Emerging technologies have facilitated optimal crop selection and enhanced the choice of hybrid seeds tailored to farmers' requirements. This has been achieved by comprehending how seeds respond to diverse weather conditions and soil types, thereby reducing the risk of plant diseases. As a result, farmers can now align with market trends, anticipate yearly outcomes, and meet consumer demands more efficiently, thus maximizing crop yields. [18]

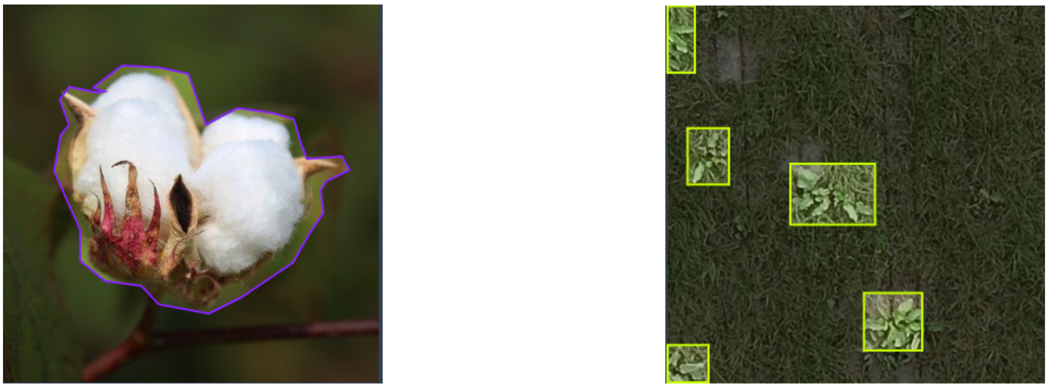
Chatbots for farmers: Chatbots are essentially conversational virtual assistants that automate interactions with end users. With the advent of artificial intelligence and machine learning techniques, chatbots have become capable of understanding natural language and engaging with users in a more personalized manner. Originally designed for industries such as retail, travel, and media, chatbots have found utility in agriculture as well. They assist farmers by providing answers to their queries, offering advice, and delivering various recommendations. [18]

2.3 Applications:

2.3.1 Overview:

In the realm of object detection research, Microsoft COCO and Pascal VOC have emerged as the predominant benchmarks for training and assessing model performance. However, these models are typically fine-tuned on smaller datasets tailored to specific imagery and target categories before being released to the public. While these general benchmarks offer insights into model performance under similar conditions, specialized domain-specific training remains invaluable. Verifiably, protest location datasets are made by gathering a large corpus of pictures and sourcing annotators to name objects in a settled set of classes.The Pascal VOC extend could be a collection of datasets made to upgrade question discovery errands and empower analysts to make models that recognize objects in reasonable scenes within the frame of a challenge. The Pascal VOC challenges begun in 2005 and laid the basis for a unused era of state-of-the-art benchmarks.ImageNet is an picture dataset comprised of over 14 million pictures each depicted by word expressions called "synset". Among the examples extracted from roboflow 100 are the grass weeds Computer Vision Project example and the cotton Computer Vision Project example.

[6]



(a) cotton

(b) grass weeds

Figure 2.1: Some applications

2.4 Neural radiance fields NeRF

2.4.1 Context:

In this study, we propose a novel approach to the enduring issue of view synthesis by directly optimizing the parameters of a continuous 5D scene representation to minimize rendering errors in a set of captured images. Our method utilizes a deep fully-connected neural network, without convolutional layers, known as a multilayer perceptron (MLP), to represent this function, mapping from a single 5D coordinate (x, y, z, ,) to a volume density and view-dependent RGB color. We demonstrate our technique by visualizing

input views of a synthetic Drums scene, randomly captured on a surrounding hemisphere, and present two novel views rendered from our optimized NeRF representation.

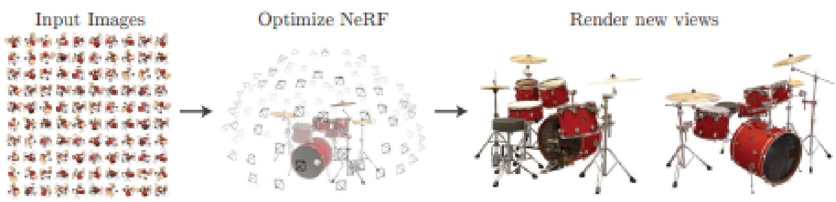


Figure 2.2: Multi-View Image Synthesis with Implicit Texture Representation and Disentangled Shape Modeling

We present a method that optimizes a continuous 5D neural radiance field representation (volume density and view-dependent color at any continuous location) of a scene from a set of input images. We use techniques from volume rendering to accumulate samples of this scene representation along rays to render the scene from any viewpoint. Here, we visualize the set of 100 input views of the synthetic Drums scene randomly captured on a surrounding hemisphere, and we show two novel views rendered from our optimized NeRF representation.

From a specific viewpoint, our process involves: 1) tracing camera rays through the scene to generate a sampled set of 3D points, 2) using these points and their corresponding 2D viewing directions as input to the neural network to produce an output set of colors and densities, and 3) employing classical volume rendering techniques to aggregate these colors and densities into a 2D image. As this process is inherently differentiable, we can employ gradient descent to optimize the model by minimizing the error between each observed image and the corresponding views rendered from our representation. This minimization across multiple views encourages the network to generate a coherent scene model by assigning high volume densities and accurate colors to locations containing the true underlying scene content. Our approach addresses the limitations of basic implementations of optimizing neural radiance field representations for complex scenes, which often fail to converge to a sufficiently high-resolution representation and are inefficient in the required number of samples per camera ray. We overcome these issues by transforming input 5D coordinates using positional encoding, allowing the MLP to represent higher frequency functions, and proposing a hierarchical sampling procedure to reduce the number of queries necessary to adequately sample this high-frequency scene representation. Our method inherits the advantages of volumetric representations, as it can effectively depict complex real-world geometry and appearance and is well-suited for gradient-based optimization using projected images. We present: an approach for representing continuous scenes with complex geometry and materials as 5D neural radiance fields, parameterized using basic MLP networks; and a positional encoding scheme to map each input 5D coordinate into a higher-dimensional space, enabling successful optimization of neural radiance fields to represent high-frequency scene content. We demonstrate that our resulting neural radiance field method outperforms state-of-the-art view synthesis methods, both quantitatively and qualitatively, including those fitting neural 3D representations to scenes and training deep convolutional networks to predict sampled volumetric representations. To our knowledge, this paper presents the first continuous neural scene representation capable of rendering high-resolution photorealistic novel views of real objects and scenes from RGB images captured in natural settings. [15]

2.4.2 Neural radiance field scene representation:

We depict a continuous scene through a 5D vector-valued function, where the input consists of a 3D location x=(x,y,z) and a 2D viewing direction (,). The output comprises an emitted color c=(r,g,b) and volume density An overview of our neural radiance field scene representation and

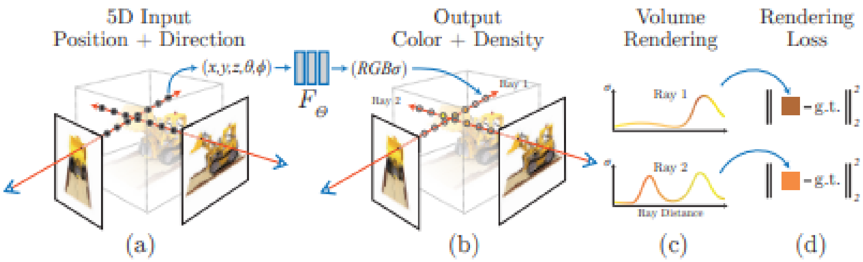


Figure 2.3: Differentiable Neural Radiance Fields for Image Synthesis and Scene Optimization

differentiable rendering procedure. We synthesize images by sampling 5D coordinates(location and viewing direction) along camera rays (a), feeding those locations into an MLP to produce a color and volume density (b), and using volume rendering techniques to composite these values into an image (c). This rendering function is differentiable, so we can optimize our scene representation by minimizing the residual between synthesized and ground truth observed images (d).

Image synthesis involves sampling 5D coordinates (location and viewing direction) along camera rays (a), inputting these locations into a Multi-layer Perceptron (MLP) to generate color and volume density (b), and employing volume rendering techniques to combine these values into an image (c). We approximate this continuous 5D scene representation with an MLP network F:(x,d)→(c,) and optimize its weights to map each input 5D coordinate to its corresponding volume density and directional emitted color. We ensure multiview consistency by constraining the network to predict volume density solely as a function of location x, while allowing RGB color c to be predicted as a function of both location and viewing direction. To achieve this, the MLP F initially processes the input 3D coordinate x with 8 fully-connected layers (employing ReLU activations and 256 channels per layer), yielding and a 256- dimensional feature vector. This feature vector is then concatenated with the camera ray's viewing direction and forwarded to an additional fully-connected layer (employing a ReLU activation and 128 channels) which produces the view-dependent RGB color. Refer to Figure 3 for an illustration of how our approach leverages the input viewing direction to represent non-Lambertian effects. [15]

2.4.3 Volume rendering with radiance fields:

Our 5D neural radiance field characterizes a scene by its volume density and directional emitted radiance at any given point in space. To render the color of a ray traversing the scene, we utilize principles from classical volume rendering. The volume density (x) can be interpreted as the differential probability of a ray ending at an infinitesimal particle located at x. The expected color C(r) of a camera ray r(t)=o+td with near and far bounds tn and tf is calculated as follows:

$$C(\tau) = \int_{t_n}^{t'} T(t)\sigma(\tau(t))c(\tau(t), d)dt, \quad \text{where} \quad T(t) = \exp\left(-\int_{t_n}^{t} \sigma(\tau(s))ds\right) \tag{2.1}$$

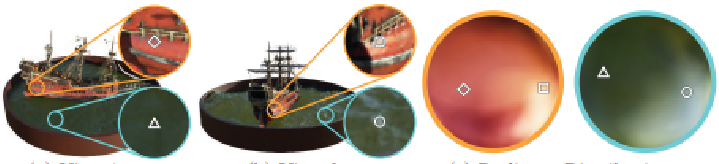


Figure 2.4: View-Dependent Radiance in Neural Scene Representations

A visualization of view-dependent emitted radiance. Our neural radiance field representation outputs RGB color as a 5D function of both spatial position x and viewing direction d. Here, we visualize example directional color distributions for two spatial locations in our neural representation of the Ship scene.In (a) and (b), we show the appearance of two fixed 3D points from two different camera positions: one on the side of the ship (orange insets) and one on the surface of the water (blue insets). Our method predicts the changing specular appearance of these two 3D points, and in (c) we show how this behaviorgeneralizes continuously across the whole hemisphere of viewing directions.

Our neural radiance field representation outputs RGB color as a 5D function of both spatial position x and viewing direction d. We illustrate example directional color distributions for two spatial locations in our neural representation of the Ship scene.Rendering a view from our continuous neural radiance field involves estimating the integral C(r) for each camera ray traced through every pixel of the desired virtual camera. We achieve this numerically using quadrature. However, deterministic quadrature, typically used for rendering discretized voxel grids, would limit our representation's resolution since the MLP would only be evaluated at a fixed discrete set of locations.Instead, we employ a stratified sampling approach, where we divide the interval [tn ,tf] into N evenly-spaced bins and randomly draw one sample from within each bin:

$$t_1 \sim \mathcal{U}\left[t_n + \frac{i-1}{N}\left(t_f - t_n\right), t_n + \frac{i}{N}\left(t_f - t_n\right)\right]. \tag{2.2}$$

Although we use a discrete set of samples for the integral estimation, stratified sampling enables us to represent a continuous scene because it ensures the MLP is evaluated at continuous positions during optimization.We use these samples to estimate C(r) with the quadrature rule discussed in the volume rendering review by Max [26]:

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i \left(1 - \exp\left(-\sigma_i \delta_i\right)\right) \mathbf{c}_i, \text{ where } T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right) \tag{2.3}$$

where

$$i = ti + 1ti \tag{2.4}$$

is the distance between adjacent samples. This function for calculating $\frown$ C(r) from the set of (ci, i) values is trivially differentiable and reduces to traditional alpha compositing with alpha values i = 1 exp (ii).

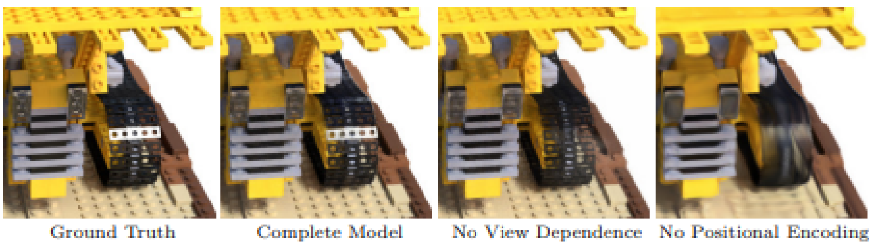


Figure 2.5: Importance of View-Dependent Emitted Radiance and High-Frequency Positional Encoding in Scene Representation

Here we visualize how our full model benefits from representing viewdependent emitted radiance and from passing our input coordinates through a high-frequency positional encoding. Removing the positional encoding drastically decreases the model's ability to represent high frequency geometry and texture, resulting in an oversmoothed appearance.

Furthermore, we illustrate how our complete model benefits from representing view-dependent emitted radiance and from passing our input coordinates through a high-frequency positional encoding. Removing the positional encoding significantly reduces the model's ability to represent high-frequency geometry and texture, resulting in an oversmoothed appearance. [15]

2.4.4 Other relevant investigations:

Neural rendering:

Different neural rendering approaches have been proposed to synthesize novel sees of a scene with a given set of photos. NeRF models the scene as a brilliance field Propelled by NeRF, followup works expand it to realize speedier induction and energetic scenes , and accomplish reflection decay . A few other neural representations have been proposed to demonstrate meso-scale surfaces. utilize neural bidirectional surface capacities (BTFs) to model known surface with propose to memorize a complex shape as a combination of a smooth low-frequency marked remove work (SDF) and a nonstop high-frequency marked separate unequivocally speaks to the surface in a neural representation through UV parameterization to bolster surface altering and mapping. In any case, such 2D parameterization accept the target protest can be smoothly mapped to a 2D parameter space, which isn't appropriate for most surfaces with meso-structure. proposed a mesh-based neural certain representation to unravel the shape and appearance. With geometry and surface highlights characterized on vertices, it accomplishes the geometry and texture editing of the neural understood field. All things considered, NeuMesh utilizes anticipated SDF instead of densities in volume rendering, which cannot be characterized on non-watertight meso-structure. meso-structure. Other than, the work putting away encodings closely fits the target surface, and as a result the meso-structure isn't learned as surface properties. the plausibility to show the surface with meso-structures through NeRF. The model is prepared on engineered datasets with rendering comes about of patches in a bounding box on a plane beneath known lighting conditions. Surfaces are mapped to the shapes by over and over putting the reproduced bounding box on surfaces. our approach targets NeRF surface amalgamation, which at the same time learns the Phong reflection coefficients, meso-structure and lighting conditions from real-world objects with surfaces The objective of surface blend is to synthesize a unused surface that shows up to be produced by the same fundamental prepare . The spearheading work by continuously develops the synthesized locale by relegating pixels one by one. The assignment is decided by neighborhood similitude. Taking after this thought, a settled neighborhood is utilized in to dodge non-uniform design conveyance. Patch-based strategy proposes to mix the covered locales between patches. The works ; the covered locales by means of energetic programming and chart cut, find perfect districts adhering to the union imperatives. proposed an elective approach by surface optimization. In expansion to conventional coordinating and optimization strategies, neural systems are too presented in surface union. show a data-driven approach to generating.

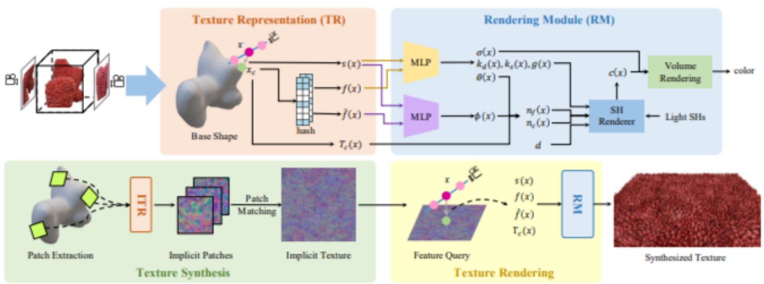


Figure 2.6: Enhanced Scene Modeling and Rendering from Multi-View Images using Disentangled Representations and Implicit Textures

Given a set of multi-view images, we first estimate its base shape. Based on it, we model the scene with a disentangled representation of the base shape and NeRF texture with meso-structure. The query point x is projected onto the base shape as footpoint x_c. Latent features $f(x), \hat{f}(x)$ representing textures are fetched by feeding x_c to hash grids. Along with matrices of local tangent space $T_c(x)$, latent features $f(x), \hat{f}(x)$, and SDF value $s(x)$ are fed into the rendering module (RM). The density $\sigma(x)$, coefficients of Phong shading model $k_d(x), k_s(x), g(x)$, elevation and azimuth angles of the fine normal $\theta(x), \phi(x)$ are predicted based on the input features and SDF. The color $c(x)$ of the query point x is calculated by Spherical Harmonic (SH) rendering based on the coarse and fine normals $n_c(x), n_f(x)$, viewing direction d, shading coefficients $k_d(x), k_s(x), g(x)$ and lighting SHs. Based on the implicit texture representation (ITR), we extract implicit patches from the base shape and synthesize texture by an implicit patch matching algorithm. By querying $f(x), \hat{f}(x)$ and $T_c(x)$ from the synthesized implicit textures, we are able to render the appearance of the synthesized texture.

surface through optimizing the Gram network of inactive highlights extricated by Follow-up works prepare feed-forward convolutional systems to supplant the time-consuming optimization Generative antagonistic systems (GANs) are too broadly utilized for surface amalgamation prepare a GAN to twofold the spatial degree of surface squares, empowering the show to synthesize non-stationary surface. utilize the Gram network delivered by the discriminator in adversarial misfortune to make strides the quality of synthesized surface. propose a Mesh-CNN based GAN engineering to synthesize geometric surfaces. progress the quality of GAN comes about. utilize occasional implanting as input and supplant the convolution layer with a Multi-Layer Perceptron (MLP) to demonstrate understood areas. [7]

Method:

We show a strategy to capture, show, synthesize and apply NeRF surfaces with meso-structure from real-world multi-view pictures. The outline of our pipeline is appeared in Fig. multi-view pictures of the scene, our show learns to unravel meso-structure surfaces and the basic base shape.By testing the certain patches of idle highlights on the base shape and utilizing them to synthesize a bigger surface outline, we are ready to beautify an subjective given work with the synthesized result. the taking after, we are going present surface representation in Sec.3.1, surface amalgamation in Sec.3.2, and show optimization in Sec.3.3. [7]

2.5 Conclusion :

AI technologies are revolutionizing agriculture by enabling farmers to efficiently analyze land, soil health, and crop conditions. Vertical cropping, for example, reduces water usage, maximizes land efficiency, and can be implemented in urban areas within buildings, addressing challenges such as labor shortages. AIbased predictive models also recommend suitable pesticides, crops, and planting locations preemptively, mitigating the risk of widespread disease outbreaks. [16]

Chapter 3

BACKGROUND EXTRACTION WITH ARTIFICIAL INTELLIGENCE

3.1 Introduction

Current technologies face some challenges in background removal for complex objects such as trees, a problem we encountered with NeRF (Neural Radiance Fields) technology. Trees are characterized by their fine details and complex structure, including multiple leaves and branches, making it difficult to accurately separate them from the background using traditional methods. Given the large number of trees and their complexity, manual processing is impractical. Therefore, we have turned to using artificial intelligence (AI) techniques to facilitate the process, speed it up, and ensure its effectiveness. These techniques rely on deep learning and neural networks to improve the accuracy of background removal and reduce the need for manual intervention, thereby enhancing the quality of results and increasing the overall efficiency of the process.

3.2 Action steps

3.2.1 Background extraction

Background extraction is an essential computer vision technique used to distinguish the foreground object from the background in an image. This process facilitates numerous applications, including object detection, image segmentation, and video editing. Recent advancements in deep learning have greatly enhanced the precision and efficiency of background extraction algorithms.[8]

3.2.2 Mention all the metods:

- **Photoshop:** Preparing images for utilization as three-dimensional object sculptures in 3D software is a task undertaken in Photoshop. The Quick Mask mode in Photoshop proves to be a robust and adaptable technique for creating precise selections.

(a) Before adjustment

(b) After adjustment

Figure 3.1: An example of background removal in an PHOTOSHOP

- **AI tools:**

Clipping Magic is an online tool created for fast and efficient image background removal. It is popular among both professionals and hobbyists who need to isolate subjects in images without extensive manual editing. Here are some of its key features:

Automatic background removal:

Utilizes advanced AI algorithms to automatically detect and remove backgrounds from images.

(a) Before adjustment (b) After adjustment

Figure 3.2: An example of background removal in an AI Tools

3.2.3 Problems

When using AI tools for background removal, we face several issues that affect the quality and efficiency of the work.

1. These tools are very time-consuming, especially when the images are complex and contain fine details.

2. The process can be tedious when done manually, such as when using programs like Photoshop, as it requires careful and continuous adjustments to ensure a satisfactory result.

3. The results are often inaccurate, as the tool can leave unwanted traces or fail to recognize edges clearly, affecting the quality of the final image.

4. These tools require powerful hardware processing, which means high-performance computers with significant processing power are needed to ensure a smooth and efficient workflow.

3.3 Tools and datasets

Tools:

3.3.1 Roboflow

Roboflow is a platform for managing, preprocessing, and augmenting computer vision datasets. It supports data upload, organization, preprocessing, augmentation, and annotation. It integrates with machine learning frameworks for model training and deployment, supports collaboration, and offers APIs for integration.

3.3.2 Google colab

To retrain the yolov8 and sam models on the database classes, we used the free google Colab environment as well as opencv libraries in python language.

Google colab: Google Colab is a free service from Google that provides a notebook environment that runs in the cloud. Colab supports popular machine learning libraries and allows editing documents as in Google Docs. It offers three types of operation: Central Processing Units (CPUs), Graphics Processing Units (GPUs), and Tensor Processing Units (TPUs), with a maximum of 12 hours of continuous operation. Multiple instances of CPU, GPU, and TPU can run at the same time, but the resources are shared. We used Colab resources with 12.69 GB RAM, 107.72 GB disk space, Tesla K80 GPU, CUDA version 11000, and cuDNN 7.6.5.

3.3.3 Python

Python is a versatile and very popular high-level programming language. The Python programming language (most recently Python 3) is used for web development, machine learning applications, and all the cutting-edge technologies in the software industry.

Datasets

3.3.4 Choosing the database:

Selecting the database was not an easy task. Initially, we attempted to build a database based on basic criteria in object detection. This effort resulted in a collection of 180 images, requiring careful consideration of the locations and angles of the selected trees.

1. Dataset preparation:

- **Upload images:**Upload your images to Roboflow.
- **Annotate backgrounds:** Use Roboflow's annotation tools to label the background in your images. This can be done manually or using pre-trained models.

2.Model integration:

- **Choose a model:** Integrate with machine learning frameworks like TensorFlow, PyTorch, or specialized segmentation models (e.g., U-Net) to train a model for background extraction.
- **Train the Model:**Use the annotated and augmented dataset to train your model within the chosen framework.

3.Model deployment:

- **Deploy the model:** Once trained, deploy the model using Roboflow's deployment options. This allows you to use the model to automatically extract backgrounds from new images.

By following these steps, you can effectively use Roboflow for background extraction in computer vision tasks.

3.3.5 Splitting data

The Training dataset

The training set constitutes the largest portion of your dataset, reserved specifically for training your model. During training, the model learns

from this dataset, and subsequent inference on these images should be approached cautiously, as the model may have memorized the correct outputs. It's recommended to allocate around
:70 % [4]

The validation dataset

The validation set is a distinct subset of your dataset utilized during training to assess the model's performance on unseen images. Throughout training, it's common to monitor validation metrics, such as validation mAP or loss, after each training epoch. These metrics provide insight into the model's progress and help determine when it achieves optimal performance on the validation set. It's advisable to set aside approximately
:20 % [4]

The test dataset

Following the conclusion of all training experiments, insights into the model's performance on the validation set may influence the model's design, potentially leading to overfitting. To mitigate this, a separate, completely untouched portion of the data—known as the test set—is essential. Allocating around :10 % [4]

How train, validation, and test relate to preprocessing and Augmentation

The train, validation, and test concept significantly impacts data preprocessing and augmentation strategies for preparing your dataset for training and model deployment. Preprocessing involves standardizing your dataset across all three splits through image transformations such as cropping or grayscale conversion. These preprocessing steps are applied uniformly to the training, validation, and test datasets. [4]

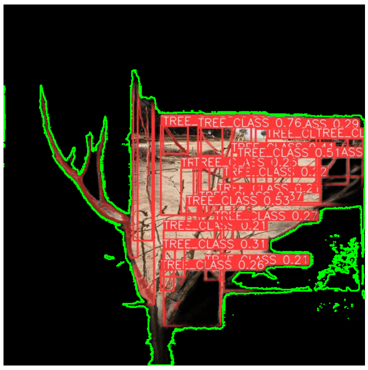
3.3.6 Training models:

Training YOLOv8 and SAM models on Google Colab requires setting up the working environment by installing the necessary libraries such as TensorFlow and PyTorch, loading and browsing the data, setting the model and training parameters, executing the training process by monitoring its progress and evaluating the performance using accuracy and loss criteria, and finally saving the model for future use.

3.4 Result and discussion

In this section we will test/verify the performance of the retrained yolov8 model on subclasses, based on selected data, and we obtained good results even in complex scenes.

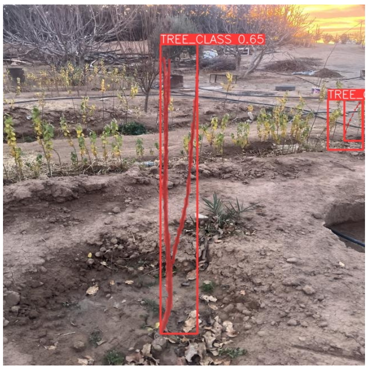
Some of the results can be seen in the following figure (3.13):

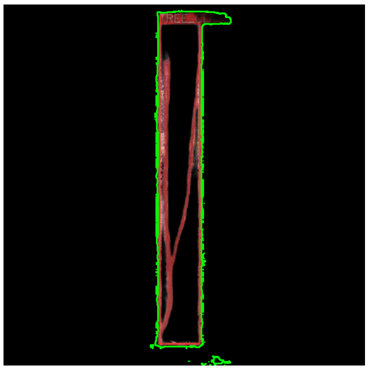


background with segmented areas

Original photo

segmented image

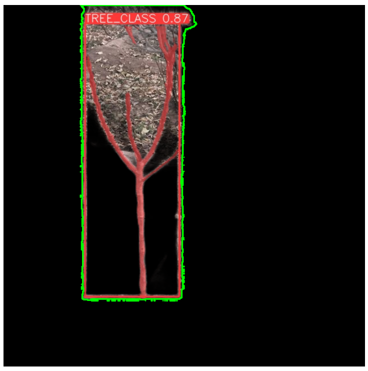




background with segmented areas

Original photo

segmented image



background with segmented areas

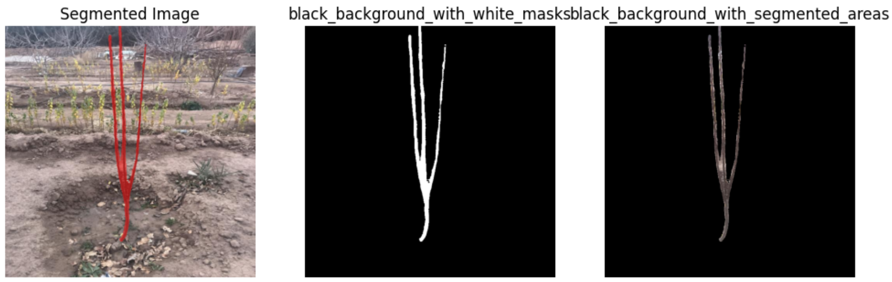
Original photo

segmented image

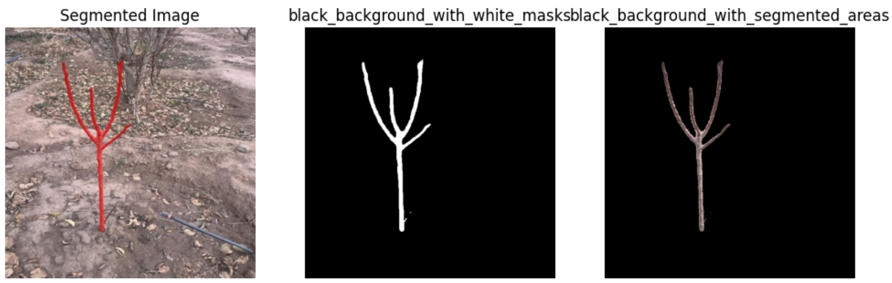
Figure 3.13: Performance test results of the retrained YOLOv8 model on subclasses

In this section, we will evaluate the performance of the retrained **SAM** model on subclasses using selected data. We achieved good results even in complex scenes. Some of these results are displayed in Figure (3.14) below:

(a) Class 1 tree



(b) Class 2 tree



(c) Class 3 tree

Figure 3.14: Performance test results of the retrained SAM model on subclasses

3.4.1 Model yolov8 accuracy and loss :

This presentation will review the accuracy and loss curves of the YOLO model, to understand how to optimize the performance of the model in the training and evaluation phases figure(3.15)

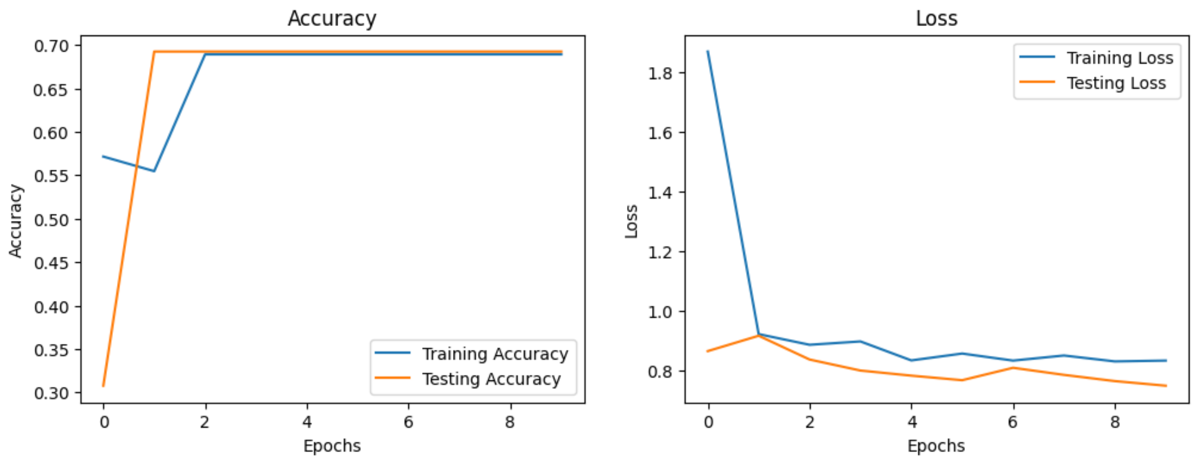


Figure 3.15: Accuracy and Loss Ratios Curve of the YOLOv8 Model

Discussion :

The performance graphs of the YOLO v8 model illustrate the accuracy and loss curves for both training and testing data across epochs.

In the first graph, which shows accuracy, we observe that the training accuracy starts at around 0.55 and rapidly increases to about 0.70 after the third epoch, remaining stable at this value until the tenth epoch. On the other hand, testing accuracy starts at around 0.30 and sharply increases to approximately 0.68 after the second epoch, then stabilizes close to the training accuracy until the tenth epoch. This indicates that the model learns quickly and reaches early performance stability for both training and testing data.

In the second graph, which depicts loss, we see that the training loss starts at around 1.8 and rapidly decreases to below 1.0 after the first epoch, then continues to decrease slowly to about 0.8 by the tenth epoch. The testing loss starts at around 0.8 and gradually decreases, also stabilizing around 0.8 after several epochs. This trend reflects the model's ability to effectively reduce error during training.

Overall, these curves indicate that the model learns quickly and achieves good performance stability after a few epochs. The lack of a significant gap between the training and testing accuracy and loss curves suggests that the model is not highly prone to overfitting, reflecting a good balance in performance.

3.4.2 Model sam accuracy and loss :

This presentation will review the accuracy and loss curves of the SAM model, to understand how to optimize the performance of the model in the training and evaluation phases figure(3.16)

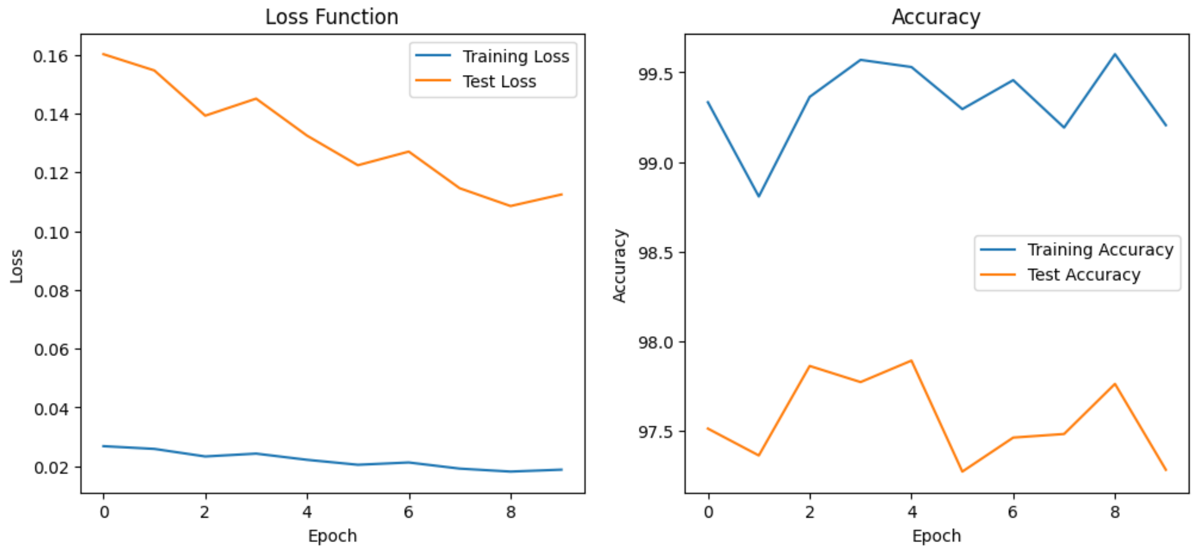


Figure 3.16: Accuracy and Loss Ratios Curve of the SAM Model

Discussion:

From the provided chart, we can observe the development of the model's performance over ten training epochs by monitoring the loss functions and accuracy for both training and test data.

On the left side of the chart, the loss curve shows a gradual decrease for both the training data (in blue) and the test data (in orange). The training data loss starts at around 0.02 and remains relatively low throughout the epochs, indicating that the model is learning effectively from the training data. The test data loss starts at a higher value (0.16) and gradually decreases to around 0.1, reflecting an improvement in the model's performance on unseen data.

On the right side, the accuracy curve shows that the training data accuracy starts high99 % and remains relatively stable with minor fluctuations, suggesting that the model generalizes well on the training data. However, the test data accuracy starts at a lower level 97.5 % and exhibits fluctuations across the epochs without significant improvement, which might indicate some challenges in generalizing to the test data.

Overall, these charts illustrate that the model performs well on the training data

3.5 Comparison of results

When comparing the two sets of curves, the model in the first image shows a rapid improvement in accuracy and reduction in loss during the initial epochs but quickly reaches a plateau, with accuracy stabilizing around 0.7 after the third epoch, indicating limited further improvement and potential learning saturation. Conversely, the model in the second image exhibits a better overall performance, with a low and stable training loss ($\sim$0.02) and a consistent decline in test loss from 0.16 to about 0.10, suggesting continual improvement. Training accuracy remains high (around 99 %) with minor fluctuations, while test accuracy ranges between 97 % and 98 %, showing higher stability and better generalization despite some variability. Overall, the second model demonstrates more sustained improvement and superior generalization on test data compared to the first model, which shows limited performance gains after a few epochs.

3.6 Conclusion

In this work, we trained two different models on our dataset to remove the background: YOLOv8 and Segment anything The results showed that the YOLOv8 model was better at image segmentation, but we had an issue with the accuracy of the masks due to its prediction frames. As a result, we chose the Segment anything model as the best option to accurately remove the background. We also noticed that optimizing the Segment anything model significantly increases the accuracy of the results.

GENERAL CONCLUSION

Intelligence technologies are the cornerstone of today's technological development, which have had a significant impact on a vast field including image processing and background removal Background image extraction from images containing trees and other complex objects is a notable challenge due to information difficulty and accuracy required so Special attention was paid to platform and deep learning methods and Initially, the topic discusses the challenges associated with the traditional methods of peristalsis, including the difficulty in separating solids from the periphery, and the technical issues involved The study shows that the traditional methods rarely works well and requires a great deal of manual handling, making it unsuitable for scale control. The thesis focuses on the Roboflow platform as a modern tool for providing an efficient solution for managing and processing image data sets. This platform enables users to upload, organize and enhance image data, and provides efficient data presentation tools. In addition, the topic explores how deep learning and transfer learning techniques can be used to enhance modeling performance and reduce the need for manual intervention. Transfer learning is particularly useful in data-scarce cases, as it improves performance by pre-trained models and enables even higher results with limited data sets The technical challenges associated with the use of AI tools for background extraction have also been addressed, such as the long processing time of complex models, possible accuracy issues, the need for powerful hardware and to ensure efficiency and effectiveness that is. The theme also highlights other tools such as Photoshop and various AI-based tools that can be used to remove backgrounds. Additionally, it uses Google Colab, a platform that provides an integrated development environment while facilitating model testing and data analysis to manage data preprocessing processes and implementation details Based on the findings of this study, it is clear that the application of AI techniques in background filtering represents a major step towards improving imaging processing These techniques contribute to more accurate and effective results and reduces the manual effort required. With continued developments in this area, it is expected that there will be significant improvements in the available tools and techniques, leading to wider and more effective applications in the future Thus, investment in AI technologies, Continuous R&D in this area is key to achieve tangible progress and provide innovative solutions to the existing challenges Adoption of this technology in industry variety will help improve efficiency and achieve more accurate and effective results

Bibliography

[1] Asad Abdi. Three types of machine learning algorithms, 11 2016.

[2] Mona Ashok, Rohit Madan, Anton Joha, and Uthayasankar Sivarajah. Ethical framework for artificial intelligence and digital technologies. *International Journal of Information Management*, 62:102433, 2022.

[3] Ms Hetal Bhaidasna and Mr Zubin Bhaidasna. Object detection using machine learning: A comprehensive. 2023.

[4] Roboflow Blog. Train, validation, test split for machine learning. https://blog.roboflow.com/train-test-split/, 2020.

[5] Wei-Lun Chao. Machine learning tutorial. *Digital Image and Signal Processing*, 2011.

[6] Floriana Ciaglia, Francesco Saverio Zuppichini, Paul Guerrie, Mark McQuade, and Jacob Solawetz. Roboflow 100: A rich, multi-domain object detection benchmark. *arXiv preprint arXiv:2211.13523*, 2022.

[7] Yi-Hua Huang, Yan-Pei Cao, Yu-Kun Lai, Ying Shan, and Lin Gao. Nerf-texture: Texture synthesis with neural radiance fields. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–10, 2023.

[8] Thien Huynh-The, Oresti Banos, Sungyoung Lee, Byeong Ho Kang, Eun-Soo Kim, and Thuong Le-Tien. Nic: A robust background extraction algorithm for foreground detection in

dynamic scenes. *IEEE transactions on circuits and systems for video technology*, 27(7):1478–1490, 2016.

[9] John D Kelleher. *Deep learning*. MIT press, 2019.

[10] Andrej Krenker, Janez Bešter, and Andrej Kos. Introduction to the artificial neural networks. *Artificial Neural Networks: Methodological Advances and Biomedical Applications. InTech*, pages 1–18, 2011.

[11] Nikita Laptev, Vladislav Laptev, Olga Gerget, and Dmitriy Kolpashchikov. Integrating traditional machine learning and neural networks for image processing. In - , volume 31, pages 896–904, 2021.

[12] Batta Mahesh. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9(1):381–386, 2020.

[13] Mendix. What are the different types of ai models? https://www.mendix.com/blog/what-are-the-different-types-of-ai-models/, 2024. Accessed: 2024-06-16.

[14] FETHIA MESBAH. Détection d'objets par deep neural network à l'aide du modèle yolo en temps réel. 2021.

[15] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

[16] Deepak G Panpatte. Artificial intelligence in agriculture: An emerging era of research. *Anand Agricultural University*, pages 1–8, 2018.

[17] Iqbal H Sarker. Ai-based modeling: techniques, applications and research issues towards automation, intelligent and

smart systems. *SN Computer Science*, 3(2):158, 2022.

[18] Tanha Talaviya, Dhara Shah, Nivedita Patel, Hiteshri Yagnik, and Manan Shah. Artificial intelligence in agriculture. 2020.

[19] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3:1–40, 2016.

[20] Patrick Henry Winston. *Artificial intelligence.* Addison-Wesley Longman Publishing Co., Inc., 1984.