Comparative Analysis of Machine Learning Models: Alexnet, VGG, Resnet, YOLO

Pham Duc An 10422002
Tran Hai Duong 10422021
Vo Thi Hong Ha 10421015
Nguyen Hoang Anh Khoa 10422037
Truong Hao Nhien 10422062
Nguyen Song Thien Phuc 10422067
Bui Duc Xuan 10422085

{@student.vgu.edu.vn}

Abstract

In this project, we conducted a comprehensive comparative analysis of prominent machine learning models, namely Alexnet, VGG, Resnet, and YOLO, with a focus on their efficacy in image recognition. Leveraging a curated dataset representative of diverse real-world scenarios with CIFAR-10, our study delved into the nuances of each model's architecture, training process, and computational requirements. Through rigorous evaluation using metrics such as accuracy, precision, and recall, our results reveal nuanced performance distinctions. Notably, Resnet demonstrated superior accuracy, VGG excelled in feature extraction, YOLO showcased real-time efficiency, and Alexnet exhibited a stable performance. These findings provide valuable insights for practitioners and researchers seeking to optimize model selection for specific applications, shedding light on the trade-offs between accuracy, computational cost, and real-time processing capabilities. Project's detailed code are provided at `https://github.com/nhientruong04/LIA-introCS-proj`.

1. Introduction

Human cognitive processes, mirroring the intricacies of an advanced supercomputer[1], rely on the nuanced interaction of neurons to perceive diverse stimuli such as digits, numbers, words, and images. This cognitive evolution spans from its early stages to the present era, with a notable milestone being the introduction of not only Generative AI but also other groundbreaking advancements in artificial intelligence.

In an era defined by the rapid advancement of artificial intelligence, the field of machine learning and deep learning stands as a beacon of innovation, transforming the way computers perceive and interact with their surroundings. This project delves into the intricate world of these cutting-edge technologies, specifically focusing on computer vision—a domain crucial for tasks ranging from image recognition to object detection. With the four chosen prominent models-Alexnet[4], VGG[5], Resnet[2], and YOLO[3], we are seeking to unravel the complexities and reveal the engine behind these widely recognized models.

The comprehensive AI taxonomy proposed by IBM outlines seven distinct types, with Generative AI representing the initial stride in the AI continuum. Within this evolving landscape, Convolutional Neural Networks (CNNs)[1] have garnered particular attention and proven to be a pivotal model. CNNs stand out for their remarkable application in various domains, excelling in tasks such as image classification, object detection, and pattern recognition.

In the realm of machine learning (ML), CNNs have risen to prominence, offering a competitive edge over traditional regression and statistical models, particularly in tasks requiring image analysis. Their efficacy is underscored by their ability to automatically learn hierarchical features from data, making them well-suited for complex visual tasks.

This paradigm shift exemplifies the dynamic nature of AI, where models like CNNs, designed to emulate the human visual system, have become indispensable tools in addressing intricate challenges across diverse disciplines.

The rest of this paper is organized as follows:

Literature review Before delving into the specific machine learning models, it is crucial to contextualize this study within the existing body of knowledge. The literature review section provides a comprehensive overview of relevant research, identifying key advancements, methodolo-

[1] `https://arxiv.org/pdf/1511.08458.pdf`

gies, and challenges in the field of image recognition and machine learning. By synthesizing existing literature, we aim to establish a foundation for understanding the evolution of these models and highlight gaps that our study seeks to address.

Models The following section explores four machine learning models-Alexnet, VGGNet, ResNet, and YOLO. We delve into their architectures, training nuances, and some key highlight that makes each of the models different.

Challenges Despite the remarkable strides made in the development of machine learning models, challenges persist. In this section, we identify and discuss key challenges encountered in the deployment and optimization of these models, offering insights into areas that demand further attention.

Experiments The experiment section outlines the experimental setup, including the dataset, evaluation metrics, and implementation details. We then present the results of our experiments, highlighting the nuances of each model's performance.

Conclusion Summarizes findings, highlighting nuanced performance distinctions, and discusses implications for practitioners. Outlines potential avenues for future research.

References

[1] M. A. Hassan and Q. M. Rizvi. Computer vs human brain: An analytical approach and overview. *Computer*, 6:580–583, 2019.

[2] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition, 2015.

[3] G. Jocher. Yolov5 by ultralytics, 2020.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, may 2017.

[5] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.