

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, 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, VGG¹, Resnet[1], and YOLO, 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)² 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:

References

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

¹<https://arxiv.org/abs/1409.1556>

²<https://arxiv.org/pdf/1511.08458.pdf>