Hyperparameter Optimization for Convolutional Neural Networks using the Salp Swarm Algorithm

Entesar H. Abdulsaed[1], Maytham Alabbas[1*], Raidah S. Khudeyer[2]
[1] Department of Computer Science, College of Computer Science and Information Technology, University of Basrah, Basrah, Iraq.
[2] Department of Computer Information Systems, College of Computer Science and Information Technology, University of Basrah, Basrah, Iraq.
Email: hopidhadha@gmail.com, ma@uobasrah.edu.iq, raidah.khudayer@uobasrah.edu.iq
[*] Correspondence

Keywords: deep learning, convolutional neural networks, salp swarm algorithm, hyperparameters optimization

Received: November 22, 2023

Convolutional neural networks (CNNs) have exceptionally performed across various computer vision tasks. However, their effectiveness depends heavily on the careful selection of hyperparameters. Optimizing these hyperparameters can be challenging and time-consuming, especially when working with large datasets and complex network architectures. In response, we propose a novel approach for hyperparameter optimization in CNNs using the Salp Swarm Algorithm (SSA). Based on the natural behavior of mollusks, SSA mimics the collective intelligence that governs feeding and navigation. Taking advantage of SSA's unique properties, our research thoroughly explores the hyperparameter space. This exploration aims to identify the algorithm that maximizes CNNs performance. This paper presents the architecture of the SSA-based framework for hyperparameter optimization and compares it to other established optimization techniques, such as Particle Swarm Optimization (PSO) and Genetic Algorithm (GA). We also present experimental results using the MNIST and fashion MNIST datasets, achieving an impressive classification accuracy of 99.46% for MNIST and 94.53% for fashion-MNIST. This case study not only contributes to the fields of deep learning and hyperparameter optimization by demonstrating the effectiveness of SSA in optimizing CNNs, but it also provides benefits to researchers and practitioners who are looking for optimal hyperparameter configurations for CNNs in a variety of computer vision applications. We also evaluate the scalability and robustness of our proposed method in the context of different CNNs structures. The insights we gained highlight SSA's potential for addressing challenges related to hyperparameter optimization.

Povzetek: Članek predstavlja optimizacijo hiperparametrov v konvolucijskih nevronskih mrežah s pomočjo algoritma Salp Swarm, ki izboljša učinkovitost in natančnost.

1 Introduction

Deep learning has emerged as a powerful and versatile field within the broader domain of machine learning [1]. It has revolutionized various domains, such as computer vision, natural language processing, and speech recognition [2, 3]. One of the fundamental techniques used in deep learning is convolutional neural networks (CNNs), which have demonstrated exceptional performance in image recognition, object detection, and classification tasks.

CNNs are designed to process grid-like data, such as images, by capturing spatial and hierarchical relationships between different features. Their architecture consists of multiple layers, including convolutional (Conv.), pooling, and fully connected layers (FC layers), allowing them to automatically extract meaningful features from input data [4]. This inherent capability makes CNNs highly effective in analyzing visual data and extracting intricate patterns that may not be discernible to the human eye [5].

While CNNs offer numerous advantages, including their ability to handle large amounts of data, learn complex representations, and achieve state-of-the-art performance in various tasks, they also possess specific weaknesses [6]. One of the critical challenges in utilizing CNNs effectively is selecting appropriate hyperparameters [7]. Hyperparameters are the configuration settings that control the behavior and performance of a CNNs model, such as learning rate, batch size, dropout rate, and kernel size [8].

The optimal selection of hyperparameters significantly impacts CNN models' performance and convergence speed. However, choosing the right combination of hyperparameters is a challenging and time-consuming task. Traditional methods, such as grid search, random search, and Bayesian optimization, suffer from computational inefficiency and may not explore the entire hyperparameter space effectively. Therefore, there is a need for advanced techniques that can efficiently and effectively optimize hyperparameters for CNNs [9].