MLP-KAN: UNIFYING DEEP REPRESENTATION AND FUNCTION LEARNING

Yunhong He* **Yifeng Xie*** **Zhengqing Yuan**[2] **Lichao Sun**[†1]

[1]Lehigh University [2]University of Notre Dame

ABSTRACT

Recent advancements in both representation learning and function learning have demonstrated substantial promise across diverse domains of artificial intelligence. However, the effective integration of these paradigms poses a significant challenge, particularly in cases where users must manually decide whether to apply a representation learning or function learning model based on dataset characteristics. To address this issue, we introduce MLP-KAN, a unified method designed to eliminate the need for manual model selection. By integrating Multi-Layer Perceptrons (MLPs) for representation learning and Kolmogorov-Arnold Networks (KANs) for function learning within a Mixture-of-Experts (MoE) architecture, MLP-KAN dynamically adapts to the specific characteristics of the task at hand, ensuring optimal performance. Embedded within a transformer-based framework, our work achieves remarkable results on four widely-used datasets across diverse domains. Extensive experimental evaluation demonstrates its superior versatility, delivering competitive performance across both deep representation and function learning tasks. These findings highlight the potential of MLP-KAN to simplify the model selection process, offering a comprehensive, adaptable solution across various domains. Our code and weights are available at `https://github.com/DLYuanGod/MLP-KAN`.

1 INTRODUCTION

In recent years, deep learning has evolved from early neural network concepts to sophisticated architectures, such as transformer networks (Vaswani, 2017), driven by advancements in computational resources and the availability of large datasets, thereby achieving remarkable performance across diverse applications. Along with important technological breakthroughs, representation learning (OpenAI, 2023a; Anthropic, 2024; OpenAI, 2023b; Touvron et al., 2023) and function learning (Narayan et al., 1996; Zhang et al., 2022; Wu et al., 2005) moments of prominence and have been extensively explored and utilized in various research and application tasks related to data and learning nowadays. At the same time, the focus of function learning research has shifted from simple function fitting to deep learning (Cuomo et al., 2022; Cai et al., 2021), which excels in tasks requiring precise function approximation and has seen new advancements, particularly in its applicability to univariate function tasks. The key difference between representation learning and function learning lies in their objectives: representation learning aims to extract features from data to understand its underlying structure (Bengio et al., 2013), while function learning focuses on creating direct mappings between inputs and outputs, making it more suited for tasks requiring precise functional relationships (Zupan et al., 1997).

In this paper, we introduce MLP-KAN, a novel framework that unifies two distinct learning approaches into a cohesive system, utilizing the Mixture of Experts (MoE) methodology Jiang et al. (2023). Within the architecture of MLP-KAN, Multi-Layer Perceptrons (MLP) (Rumelhart et al., 1986) function as representation experts, while Kernel Attention Networks (KAN) (Liu et al., 2024) are designated as function experts. The MoE mechanism efficiently routes inputs to the appropriate expert, significantly enhancing both efficiency and performance across a diverse range of tasks.

*Yunhong and Yifeng are independent undergraduate students, remotely work with Lichao Sun.
†Lichao Sun is corresponding author: lis221@lehigh.edu

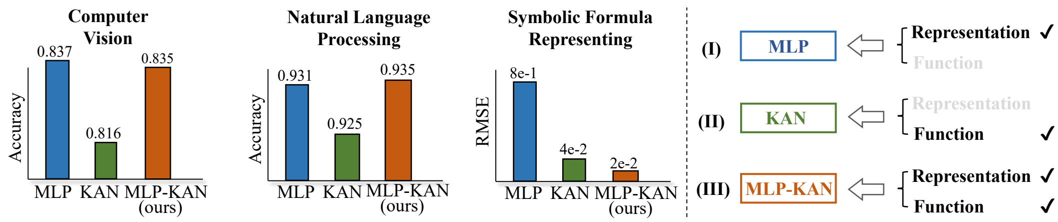

Figure 1: The comparison between the MLP, KAN, and our proposed MLP-KAN. In the domains of Computer Vision and Natural Language Processing, the goal is to achieve the highest accuracy possible. In contrast, for the Symbolic Formula Representation task, the objective is to minimize the root mean square error (RMSE). The numbers are the average values of the experimental results. MLP-KAN effectively combines the strengths of both, ensuring strong performance in representation and function learning, and eliminating the need for task-specific model selection.

MLP-KAN was developed to address the problem users encounter when determining whether to apply representation learning or function learning models across diverse datasets. By integrating MLPs and KANs within a mixture-of-experts framework, this architecture dynamically adapts to the specific characteristics of the task, ensuring optimal performance without requiring manual model selection. The main challenge in our method is effectively integrating MLPs and KANs, ensuring the right model is selected for each task without compromising performance. In additional, balancing the differing training needs of representation and function learning while maintaining efficiency across diverse datasets is complex. The main challenge in our method is effectively integrating MLPs and KANs, ensuring the right model is selected for each task without compromising performance, as shown in Figure 1. In additional, balancing the differing training needs of representation and function learning while maintaining efficiency across diverse datasets is complex.

To address the challenge of effectively integrating MLPs and KANs within the MoE framework, we utilized a soft MoE approach. This method enables dynamic and flexible routing between MLPs for representation learning and KANs for function learning. By incorporating this MoE system within a transformer framework, the model can seamlessly perform deep representation learning or deep function learning, adapting to the specific nature of the task at hand while maintaining efficiency across diverse datasets.

The main contributions of this work are as follows:

- We present MLP-KAN, a unified framework that synergizes MLP for representation learning with KAN for function learning. This novel architecture leverages a MoE mechanism to dynamically route tasks between representation and function experts, addressing the challenge of selecting the appropriate learning paradigm for diverse datasets.
- We propose a flexible and versatile model by integrating MLP-KAN within the transformer architecture, enabling efficient performance across both representation and function learning tasks. This integration enhances model capability and improves performance across a broad range of tasks, including computer vision, natural language processing, and symbolic formula representation.
- We perform extensive experimental evaluations, demonstrating that MLP-KAN consistently outperforms or matches state-of-the-art models such as MLP and KAN on widely recognized benchmarks, including computer vision,nature language processing, and functional dataset. Our approach achieves superior accuracy in representation learning tasks and lower RMSE in function learning tasks, underscoring its universal applicability across diverse domains.

2 Related Work

Deep Representation Learning. Deep representation learning has gained significant attention due to its ability to automatically discover hierarchical feature representations from raw data (Butepage et al., 2017; Zhong et al., 2016; Long et al., 2018), outperforming traditional hand-crafted feature