UE22CS320A – Capstone Project Phase – 1

Project Progress Review #2

Project Title : Exploring the Applicability of the Lottery Ticket Hypothesis to Diffusion Models

Project ID : PW25_COP_01

Project Guide : Mr. Prakasha C O

Project Team with SRN : PW_25_COP_01_216_240_245_310

Agenda

1. Introduction
2. Problem Statement
3. Feasibility study and Applications/Use cases
4. Suggestions from Review - 1
5. Constraints / Dependencies / Assumptions / Risks
6. Functional Requirements
7. Non-Functional Requirements
8. Literature Survey
9. Summary of Literature Survey
10. Conclusion
11. References

Introduction

- In this capstone project, we aim to explore the application of the Lottery Ticket Hypothesis (LTH) to diffusion models, a type of generative model that has gained prominence in areas such as image generation and denoising tasks.

- The LTH, introduced by Frankle and Carbin in 2018, posits that within large, pre-trained neural networks, there exist smaller sub-networks—referred to as "winning tickets"—that can be retrained from scratch to achieve comparable, or sometimes even better, performance relative to the full, original network.[1]

Introduction

What are diffusion models?

Diffusion Models are a class of generative models that generate data (e.g., images) by reversing a gradual process of adding noise to the data.

Training Process:

- The model learns to add noise to data (e.g., images) step by step.
- Then, it learns to reverse the process, denoising the corrupted data to recover the original, clean data.

Generation:

- Starting from random noise, the model gradually refines the noise until it generates realistic data.

Applications:

- Used in tasks like image synthesis, audio generation, and video creation, achieving state-of-the-art results in text-to-image models (e.g., DALL-E)

Introduction

Research Gap & Motivation

While the Lottery Ticket Hypothesis (LTH) has been extensively studied its applicability to diffusion models, remains largely unexplored.

Most existing research on LTH focuses on discriminative models leaving a gap in understanding how LTH can be leveraged in the context of generative processes that involve iterative transformations, such as diffusion models [2].

Key Challenge

Diffusion models, despite their impressive performance in generating high-quality samples, are computationally intensive due to the long chain of reversing the diffusion process.

Problem Statement

The main research question is: How effectively does the Lottery Ticket Hypothesis (LTH) apply to diffusion models?

Sub-questions include:

1. How often do "winning tickets" exist in diffusion models, and can they be identified through pruning and reinitialization?
2. What are the efficiency gains observed in terms of computational cost and memory usage when using LTH on diffusion models?
3. Can LTH be generalized across different diffusion model architectures (e.g., DDPM vs. Latent Diffusion Models)?
4. How does the performance of pruned diffusion models (winning tickets) compare to full models in terms of robustness to noisy inputs and adversarial attacks?
5. Do "winning tickets" transfer across diffusion networks of the same family?

Feasibility Study and Applications/Use Cases

Applications and Use Cases

1. **Model Compression:** LTH can lead to lighter diffusion models, enabling deployment on resource-constrained devices.
2. **Improved Training Efficiency:** Identifying winning tickets could reduce training time and resource requirements for diffusion models.
3. **Generative Art and Content Creation:** Efficient models derived from this research can help creators generate high-quality visuals with fewer resources.
4. **Natural Language Processing (NLP):** Findings could enhance the efficiency of language models in tasks like text generation and sentiment analysis.

Feasibility Study and Applications/Use Cases

Possible Shortcomings/Challenges

1. **Model Complexity**: Diffusion models can be computationally intensive, limiting experimental scope and requiring powerful hardware.
2. **Overfitting**: There's a risk of overfitting, necessitating careful regularization and validation techniques to ensure generalization.
3. **Generalization Across Tasks**: The sparsity identified by LTH may not generalize well across different tasks or datasets.
4. **Parameter Selection**: Identifying "winning tickets" can be complex, requiring extensive experimentation with pruning techniques.
5. **Interpretability**: Winning tickets in diffusion models may be difficult to interpret, complicating the understanding of their contributions.

Suggestions from Review - 1

Mentor Notes:

Initial Objective: Mechanistic Interpretability in Diffusion Models

- Our original objective was to explore architectural modifications aimed at improving model interpretability by eliminating superposition. By building superposition-free models, we hoped to make models more interpretable.

- However, further research showed that models without superposition still produce polysemantic neurons.Neurons tend to represent multiple features ambiguously , leading us to conclude that architectural changes alone are fundamentally non-viable for creating interpretable models.[9]

- **Redirection to LTH on Diffusion Models**

  Based on this, we decided to shift our focus to a more feasible direction—applying the Lottery Ticket Hypothesis (LTH) to diffusion models. research and developments, including recent work from March 2024, highlight the greater feasibility and scope of this area . [2]

Constraints / Dependencies / Assumptions / Risks

Constraints

Computational Resources: Requires high-performance hardware for model training and pruning.

Time Limitations: Limited time for conducting extensive experiments.

Dependencies

Software/Hardware: Specific versions of deep learning libraries (e.g., PyTorch) and access to GPUs/TPUs.

Assumptions

Generalizability: Results will apply to various diffusion model architectures.

Data Consistency: Assumes datasets (e.g., CIFAR-10, MNIST) are representative.

Risks

Overfitting: Risk of pruned networks overfitting.

Hardware/Software Failure: Potential hardware or library compatibility issues.

Performance: Pruned models might not retain sufficient performance.

Functional Requirements

Baseline Diffusion Models

- Implement standard architectures and train on diverse datasets.
- Conduct hyperparameter tuning for optimal performance.

Tailored Pruning Algorithms

- Develop algorithms to identify and remove unnecessary parameters.
- Utilize weight, structured, and dynamic pruning techniques.

Reinitialization Strategies

- Implement methods to reinitialize weights of pruned sub-networks.
- Explore random and winning ticket-based reinitialization.

Computational Efficiency Measurement

- Compare efficiency of baseline and pruned models.

Winning Ticket Analysis

- Investigate occurrence, stability, and generalization of winning tickets.
- Develop visualization methods for interpretability.

Non – Functional Requirements

Scalability: The pruning and evaluation framework should be scalable to handle various diffusion model architectures

Performance: Pruned models should maintain sample quality within an acceptable threshold compared to full models

Efficiency: The pruning process should result in significant reductions in computational cost and memory usage

Reproducibility: Ensure all experiments and results are reproducible by providing necessary code and configurations

Literature Survey(Survey 1)

The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks

Authors:
Jonathan Frankle, Michael Carbin (MIT CSAIL)

Published:
2019

Objective:
This paper introduces the **Lottery Ticket Hypothesis**, which proposes that within dense, randomly initialized neural networks, there exist small **subnetworks ("winning tickets")**. These subnetworks can be trained in isolation to achieve the same performance as the original network in **fewer iterations**. The method outlined in the paper helps identify these winning subnetworks, showing that they can reduce the size of the network by up to **90%** without sacrificing performance.

Link to Paper:
arxiv.org/abs/1803.03635

Literature Survey(Survey 1)

Introduction to Lottery Ticket Hypothesis (LTH)

- **Core Idea**: The Lottery Ticket Hypothesis (LTH), introduced by Frankle and Carbin in 2019, posits that large, dense neural networks contain much smaller, sparse subnetworks ("winning tickets") that can be trained independently to achieve similar performance to the full network. These subnetworks emerge through a process of pruning unimportant weights after an initial training cycle.[1]
- **Research Motivation**: Traditional neural networks are overparameterized, and pruning techniques reduce the network size post-training. LTH aims to find sparse networks that could be trained from the start, reducing computational costs from both training and inference without sacrificing accuracy.
- **Key Question**: Why train a large, dense network only to prune it later when a smaller network could suffice? LTH offers a potential solution by identifying these "winning tickets" at initialization.

The figure demonstrates how pruning eliminates unimportant connections, reducing a dense neural network to a smaller, more efficient subnetwork. This is key to understanding how "winning tickets" are identified.

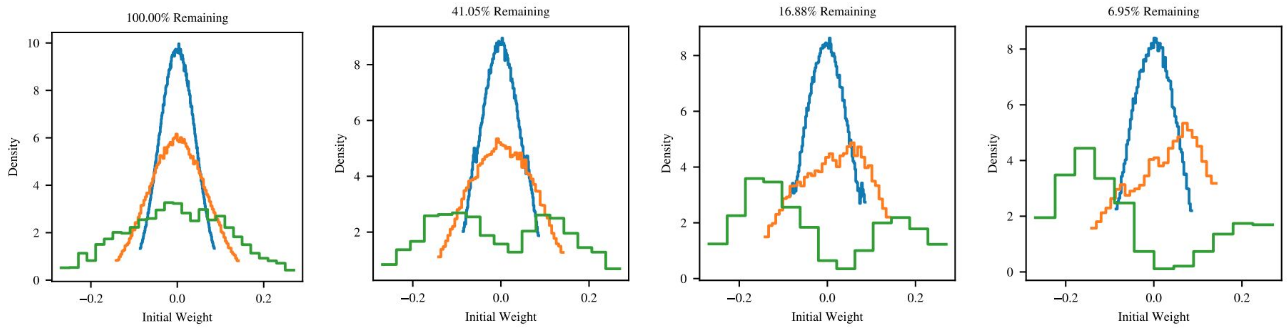


Figure 15: The distribution of initializations in winning tickets pruned to the levels specified in the titles of each plot. The blue, orange, and green lines show the distributions for the first hidden layer, second hidden layer, and output layer of the Lenet architecture for MNIST when trained with the adam optimizer and the hyperparameters used in 2. The distributions have been normalized so that the area under each curve is 1.

Literature Survey(Survey 1)

Key Techniques and Findings

- **Iterative Pruning Process**: The authors used iterative pruning, a method where small percentages of network weights are removed in several stages. After each pruning round, the remaining weights are reset to their original values, forming the "winning ticket." This process is repeated multiple times to refine the network.
- **Experiments Conducted**: The hypothesis was tested on MNIST (fully connected networks) and CIFAR-10 (convolutional networks). The experiments showed that subnetworks pruned to as little as 10-20% of the original size could perform on par with the original networks in terms of accuracy.
- **Performance Benefits**: These pruned networks ("winning tickets") not only retained accuracy but also learned faster, converging in fewer iterations compared to their unpruned counterparts, illustrating the potential to reduce computational costs without loss of performance.

This figure shows the test accuracy of pruned and unpruned networks over time, highlighting how winning tickets found through iterative pruning converge faster and retain performance compared to the full network.

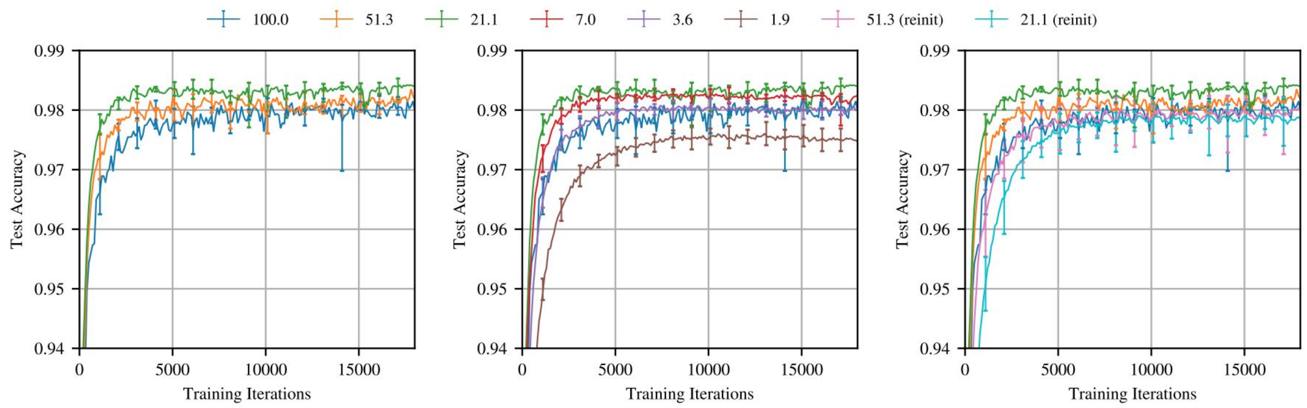


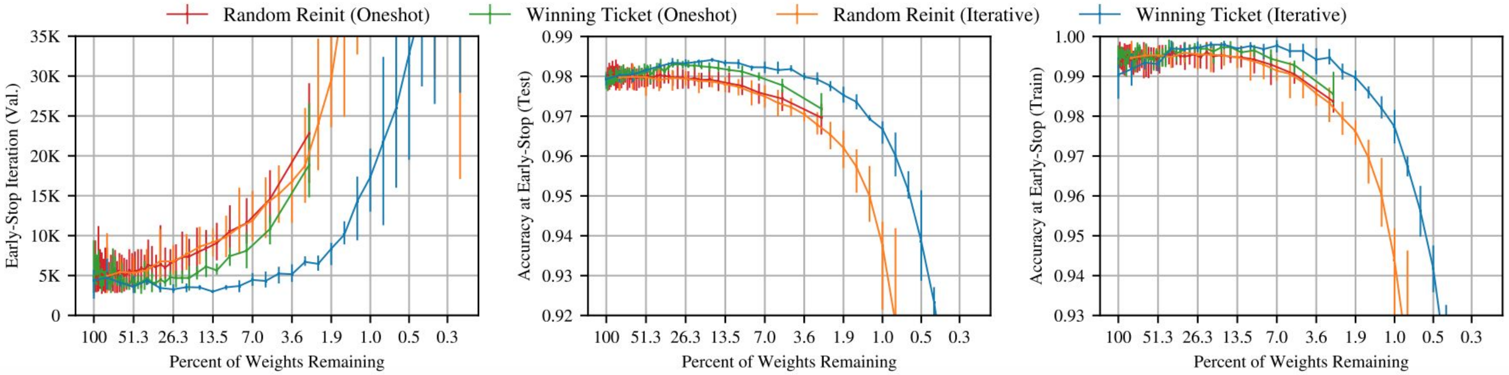
Figure 3: Test accuracy on Lenet (iterative pruning) as training proceeds. Each curve is the average of five trials. Labels are P_m—the fraction of weights remaining in the network after pruning. Error bars are the minimum and maximum of any trial.

Literature Survey(Survey 1)

Comparative Analysis & Key Results

- **LTH vs. Conventional Pruning**: Unlike conventional pruning methods, which retrain networks after pruning, LTH subnetworks are identified through the original training process itself. The winning tickets are reset to their original initialized values, not randomly reinitialized, as random reinitialization leads to poor performance. This shows that their initial conditions (initial weights) play a crucial role.
- **Generalization**: Another significant finding is that pruned subnetworks often generalize better. Despite their smaller size, winning tickets often exhibited higher test accuracy, suggesting that the original network might be overfitting, while the pruned subnetwork captures a more optimal balance between capacity and simplicity.
- **Further Insights**: The hypothesis extends beyond mere size reduction; it challenges conventional wisdom by suggesting that dense networks contain smaller, better-optimized subnetworks that could be trained from scratch with fewer parameters while retaining or improving accuracy.

- The figure highlights the difference between networks pruned and reset to their initial weights versus those reinitialized randomly. Winning tickets clearly outperform reinitialized networks, demonstrating the importance of initialization in achieving high performance.



(a) Early-stopping iteration and accuracy for all pruning methods.

Literature Survey(Survey 1)

Broader Implications and Future Research

- **Implications for Neural Network Design**: LTH opens the door for designing more efficient networks from the start. If we can identify winning tickets early in the training process, we could avoid the resource-intensive training of large networks and instead focus on smaller, sparse architectures from the beginning.
- **Efficiency Gains**: The application of LTH has the potential to significantly improve the efficiency of deep learning models in resource-constrained environments, such as mobile devices, by reducing memory and computational requirements during inference.
- **Future Research Directions**: Future work could explore the applicability of LTH across different network architectures, like diffusion models, transformers, and larger datasets such as ImageNet. Additionally, understanding the properties that make certain initializations more favorable (winning tickets) can lead to new initialization techniques or regularization methods that enhance network training.
- **Broader Impact**: LTH could help in reducing energy consumption, and possibly even improving theoretical models of how neural networks learn and generalize. Moreover, combining LTH with techniques like dropout or other regularization strategies may yield even more efficient networks.

This figure shows the application of iterative pruning to VGG-19, a more complex architecture. It demonstrates that LTH applies not just to basic networks but also to deeper, more sophisticated models, highlighting its broad applicability.

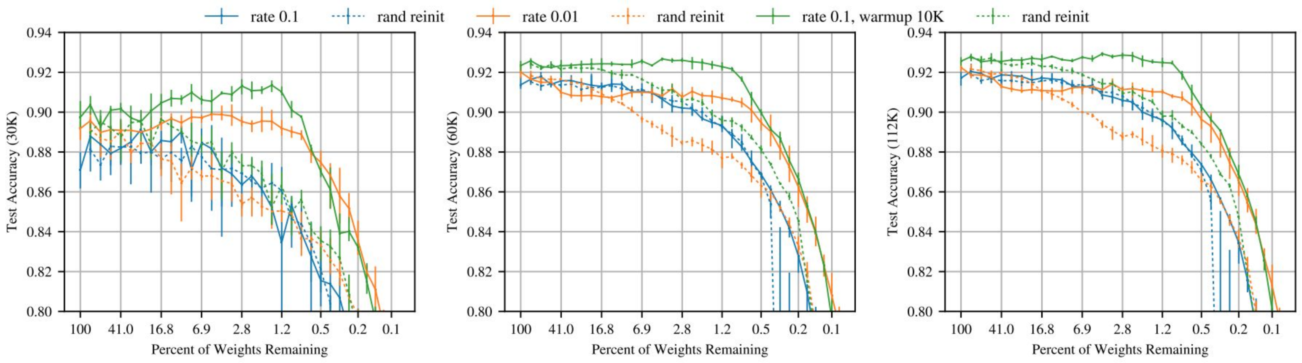


Figure 7: Test accuracy (at 30K, 60K, and 112K iterations) of VGG-19 when iteratively pruned.

Literature Survey(Survey 2)

Title:
The Elastic Lottery Ticket Hypothesis

Authors:
Xiaohan Chen, Yu Cheng, Shuohang Wang, Zhe Gan, Jingjing Liu, Zhangyang Wang

Published:
2021

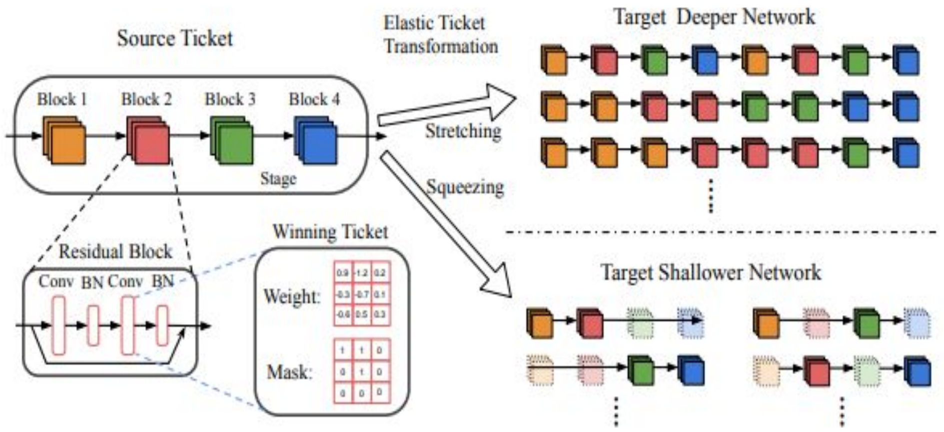
Objective:
This paper introduces the **Elastic Lottery Ticket Hypothesis**, which suggests that once a **winning ticket** is found for a specific model, the same ticket can be **"shrunk" or "stretched"** to find winning tickets for other models within the same family. This approach significantly reduces **computational time and cost**, making these models more accessible to a wider audience.

Link to Paper:
https://arxiv.org/abs/2103.16547

Literature Survey(Survey 2)

- **Elastic Lottery Ticket Hypothesis (E-LTH):** extends the **Lottery Ticket Hypothesis (LTH)** by proposing that winning tickets (sparse subnetworks found through pruning) can be **transferred across different neural network architectures**. Specifically, E-LTH suggests that a winning ticket found for one network can be stretched or squeezed to match another network within the same architecture family (e.g., ResNet, VGG). This enables reusing a winning ticket without having to undergo the expensive process of iterative pruning for each architecture separately. [3]

- **Elastic Ticket Transformation (ETT):** This involves replicating or removing specific layers from a winning ticket to match the target architecture. The authors conduct experiments using networks such as ResNets and VGG on CIFAR-10 and ImageNet, showing that transformed tickets maintain comparable performance to those found through more computationally expensive methods like Iterative Magnitude Pruning (IMP).



Literature Survey(Survey 2)

Iterative Magnitude Pruning (IMP):

IMP is a method used to identify winning tickets by repeatedly pruning weights based on their magnitudes. The process involves:

1. Training the full network.
2. Pruning a fraction of the smallest magnitude weights.
3. Resetting the remaining weights to their initial values and retraining the network.
4. This cycle is repeated until the desired sparsity is achieved.

Advantages of IMP:

- **High Performance:** IMP typically results in winning tickets that perform similarly to the original network, even with significantly fewer parameters.
- **Reliable Pruning:** It preserves the most important weights, allowing the network to retain essential features for good performance.
- **IMP** was one of the first models to be proposed to implement the lottery ticket hypothesis. To this day, no other model has provided the level of accuracy that has been seen in IMP models.

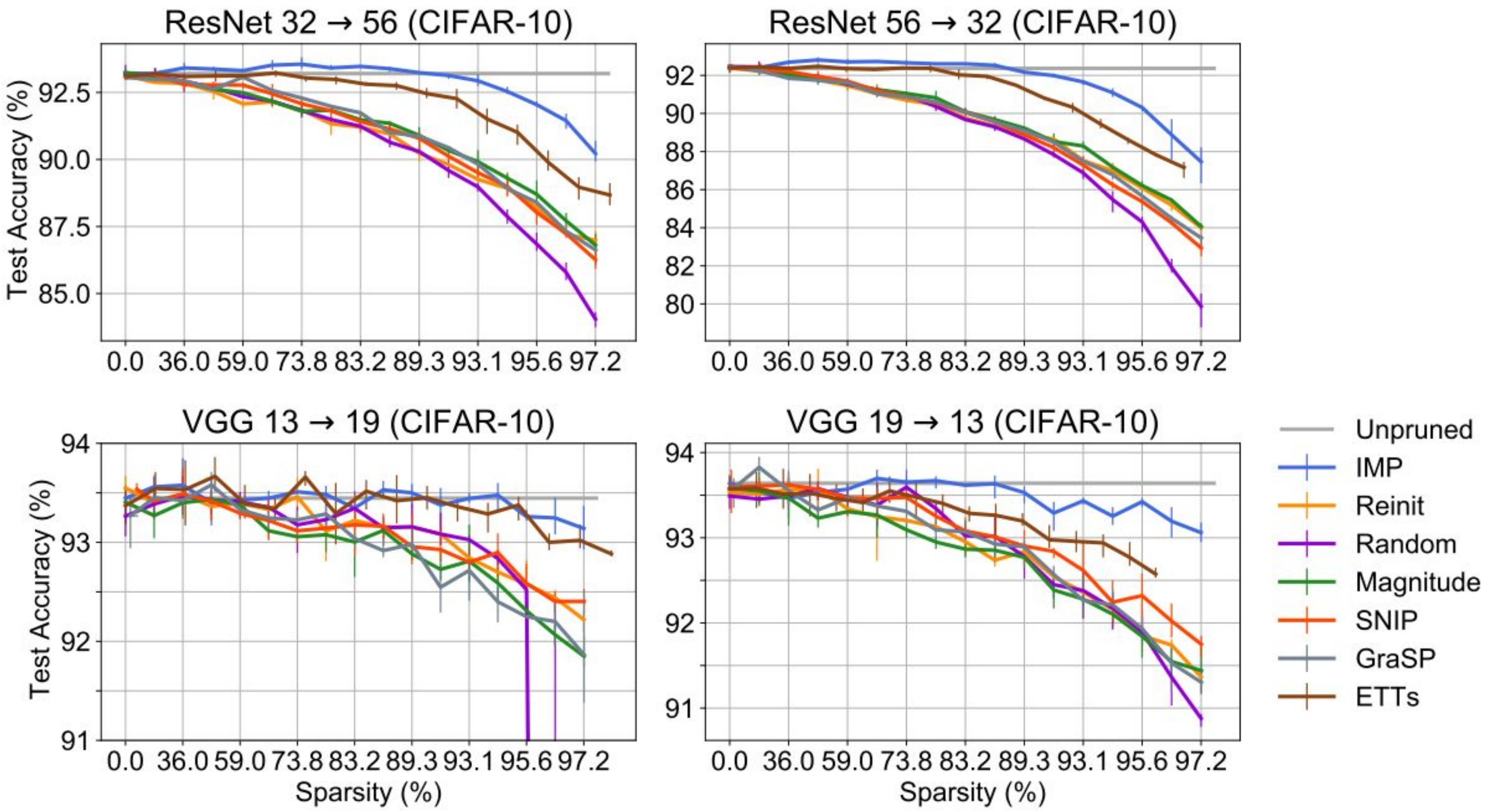
Disadvantages of IMP:

- While highly accurate, IMP is one of the most inefficient and costly models in the market. Dynamic Sparse Training models like RigL are much more efficient and sometimes even approach IMP levels of accuracy.

Literature Survey(Survey 2)

To show that the new network generated by ETTs is a winning ticket, we compare the performance of ETTs with the following baselines when transforming between ResNet-32 and 56 and between VGG-13 and 19.

- These graphs are sparsity vs accuracy graphs which test out to see how models hold out when more and more of the parameters are pruned.
- We can see from them that while ETT is not as accurate as IMP itself, it approaches the same levels, while being much more efficient, both computationally and cost wise.
- Overall ETTs seem to work well in ResNet, but see some fall in accuracy at high sparsity in VGG models.



Literature Survey(Survey 2)

Use Cases for E-LTH

- **Cross-Platform Model Deployment**: E-LTH allows models to adapt to various hardware platforms by stretching or shrinking subnetworks without retraining.
- **Neural Network Compression**: E-LTH enables efficient pruning, producing smaller models that maintain high accuracy in resource-constrained environments.
- **Transfer Learning**: E-LTH facilitates transferring sparse subnetworks across different architectures, speeding up fine-tuning for related tasks.
- **Dynamic Model Scaling**: E-LTH supports adjusting model sizes in real-time based on available computational resources, ensuring optimal performance.
- **Cloud-to-Edge Adaptation**: E-LTH enables deploying a single sparse network on both powerful cloud servers and low-power edge devices with minimal performance loss.

Future prospect for E-LTH

- In the future, **E-LTH** could revolutionize model compression and deployment by enabling more **adaptive and scalable neural networks** across diverse architectures and hardware platforms.
- Research could further explore applying E-LTH to a wider range of **tasks and architectures**, including transformers and large-scale models.

Literature Survey(Survey 3)

Dual Lottery Ticket Hypothesis (DLTH)

Authors:
Yue Bai, Huan Wang, Zhiqiang Tao, Kunpeng Li, Yun Fu

Objective:
The **Dual Lottery Ticket Hypothesis (DLTH)** builds on the foundational principles of the original Lottery Ticket Hypothesis (LTH). It posits that not only can small subnetworks, referred to as **"winning tickets"**, be identified within a densely connected neural network, but also that **any randomly selected subnetwork from a randomly initialized dense network can be transformed into a trainable condition.** This transformation enables these subnetworks to achieve performance levels comparable to those of the winning tickets identified by LTH. By demonstrating that the potential for effective training extends beyond specific winning tickets, DLTH significantly broadens the landscape of model optimization.

Literature Survey(Survey 3)

Dual Lottery Ticket Hypothesis (DLTH)

- The Dual Lottery Ticket Hypothesis (DLTH) builds upon the foundational principles of the original Lottery Ticket Hypothesis (LTH). It posits that within a densely connected neural network, not only can small subnetworks, termed "winning tickets," be identified, but any randomly selected subnetwork can also be transformed into a trainable entity that achieves commendable performance. This transformation suggests that the landscape of model optimization can be broadened beyond the constraints of specific winning tickets.[4]

- Introduces Random Sparse Network Transformation (RST) to transform a randomly selected subnetwork into a trainable condition. RST simply introduces a gradually increased regularization term to achieve information extrusion from extra weights (which are set to be masked) to the target sparse structure.

- Provides a more controllable way to select and train sparse networks, paving the way for future research in neural network pruning and optimization.

Literature Survey

DLTH vs. LTH: A Comparative Perspective

- LTH Recap: The Lottery Ticket Hypothesis (LTH) posits that within a dense network, a small subnetwork (the "winning ticket") exists, which, when trained in isolation, performs comparably to the full model. However, identifying this subnetwork requires pruning a pre-trained network, making LTH resource-intensive.

- DLTH's Novel Approach: DLTH extends this idea by proposing that instead of searching for specific "winning tickets," any randomly selected subnetwork can be transformed into a trainable condition. This transformation is achieved through the Random Sparse Network Transformation (RST) method.

-

Literature Survey (Survey 3)

The Random Sparse Network Transformation (RST) Method

Overview of RST: The RST method is a key component of DLTH. It involves selecting a subnetwork randomly from a dense, randomly initialized neural network and applying a gradually increasing regularization term to the weights outside the subnetwork (those set to be masked). This process extrudes useful information from the masked weights into the subnetwork.

Step-by-Step Process:

- **Random Subnetwork Selection:** A subnetwork is chosen at random from the dense network.
- **Regularization:** A regularization term is applied to the masked weights, forcing them to contribute information to the unmasked subnetwork.

Literature Survey (Survey 3)

A regularization term is used to realize information extrusion.Given a randomly initialized dense network f(x; θ) with a randomly selected subnetwork represented by mask m ∈ {0, 1}, the information extrusion can be achieved as optimizing following loss:

$$\mathcal{L}_R = \mathcal{L}(f(x;\theta), \mathcal{D}) + \frac{1}{2}\lambda\|\theta^*\|_2^2,$$

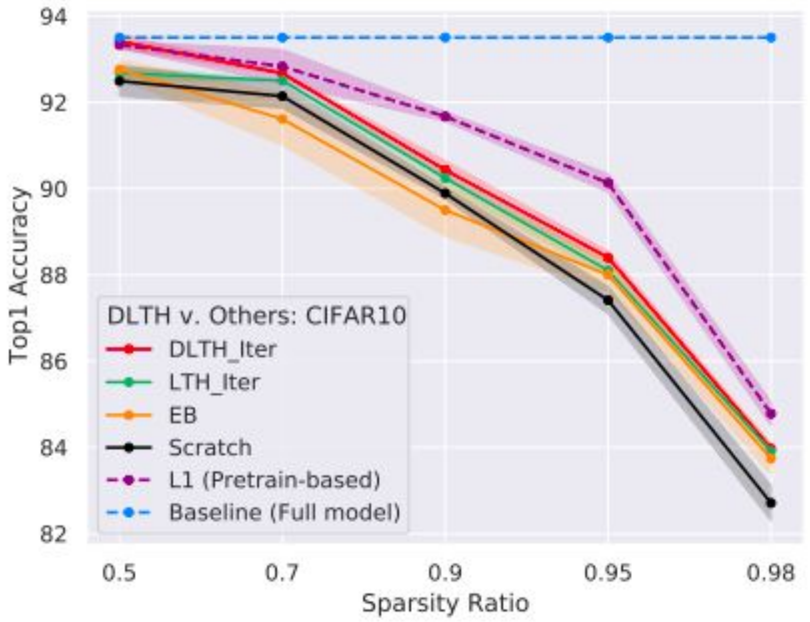
where

- loss LR contains a regular training loss L on given data D
- L2 regularization term added on θ ∗ .
- θ∗ is the part of parameter θ which being masked by m =0.
- unmasked weights, referred as θ ∗
- λ, trade-off parameter is set to gradually increase from a small value

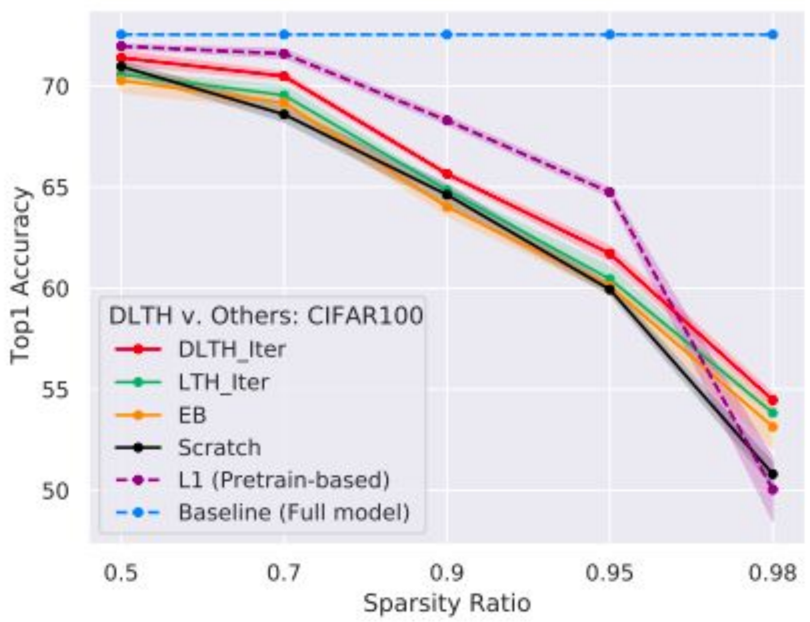
Literature Survey DLTH – Experimental Results

Empirical studies demonstrate that the DLTH framework not only matches but often exceeds the performance benchmarks established by LTH, particularly at high sparsity levels—up to 98%.

The sparse structure is unpredictable in LTH. DLTH allows the sparse structure being controlled by ourselves.



(a) DLTH v. Others on CIFAR10.



(b) DLTH v. Others on CIFAR100.

Literature Survey(Survey 4)

Successfully Applying Lottery Ticket Hypothesis to Diffusion Model

Authors:
Chao Jiang, Bo Hui, Bohan Liu, Da Yan

Published:
2023

Objective:
The objective of this paper is to apply the **Lottery Ticket Hypothesis (LTH)** to **Denoising Diffusion Probabilistic Models (DDPMs)**, with the goal of identifying sparse, trainable subnetworks (or "winning tickets") that can significantly reduce the model's size and computational cost, while maintaining the same high-quality image generation performance as the original, unpruned model. This is aimed at improving the efficiency and scalability of diffusion models, making them more practical for resource-constrained environments.

Link to Paper:
https://arxiv.org/abs/2310.18823

Literature Survey(Survey 4)

Lottery Ticket Hypothesis for diffusion model

Despite the success of diffusion models, their training and inference are computationally expensive due to the extended reverse process. The Lottery Ticket Hypothesis (LTH) suggests that sparse subnetworks, or "winning tickets," can match the performance of the full model when trained in isolation. For the first time, we apply LTH to diffusion models, finding subnetworks with 90%-99% sparsity that maintain performance on benchmarks like CIFAR-10, CIFAR-100, and MNIST. Unlike previous works that assume uniform sparsity across layers, we observe that the similarity between winning tickets varies across the model, with upstream layers being more similar than downstream ones. Therefore, we propose identifying winning tickets with varying sparsity across layers. Experiments show that this approach reduces memory usage and computational costs (FLOPs) while preserving model performance.[5]

Key Objectives:

1. Identify sparse subnetworks in diffusion models that retain high-quality image generation while reducing computational demands.
2. Prune networks iteratively, resetting remaining weights to their original values, to find the winning ticket within the full diffusion model.

Literature Survey(Survey 4)

Key Techniques and Findings

- **Lottery Ticket Hypothesis in Diffusion Models**:
  - It was applied for the first time to **Denoising Diffusion Probabilistic Models (DDPMs)**.
  - It aimed to find sparse, trainable subnetworks (winning tickets) that reduce computational costs without sacrificing quality.
- **Iterative Pruning Strategy**:
  - The network weights are pruned iteratively, resetting remaining ones to original values. Pruning up to **99%** of the network, maintaining image generation performance.
- **Layer-wise Sparsity**:
  - The pruned networks have varying sparsity across layers, with higher sparsity in later layers to improve efficiency while preserving key features in early layers.
- **Performance Results**:
  - The achieved comparable image generation quality with **90–99% sparsity**.
  - There is significant reduction in computational cost while retaining visual fidelity.

Literature Survey(Survey 4)

Comparative Analysis & Key Results

1. Pruned vs. Unpruned Models

- **Pruning Ratio**: Networks pruned up to **99%**.
- **Performance**: Pruned networks performed **on par** with unpruned models in terms of image generation quality.

2. Efficiency Gains

- **Computational Cost**: Significant reduction in memory usage and training time due to model sparsity.
- **Faster Convergence**: Pruned networks required fewer iterations to converge compared to full models.

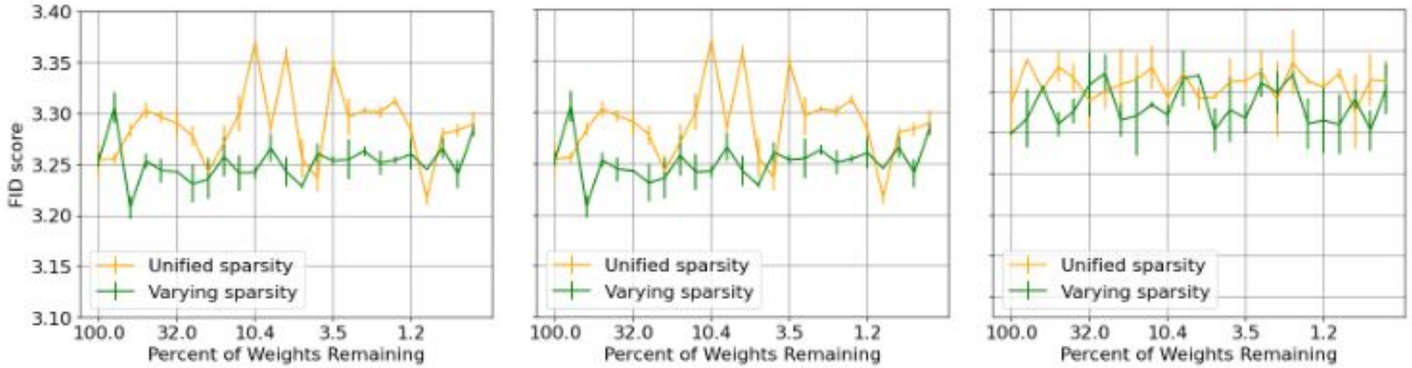
3. Layer-Wise Sparsity

- **Adaptive Pruning**: Lower sparsity in early layers (preserving more weights) and higher sparsity in later layers optimized performance.
- **Effectiveness**: This approach allowed for minimal quality loss despite extensive pruning.

4. Image Quality Consistency

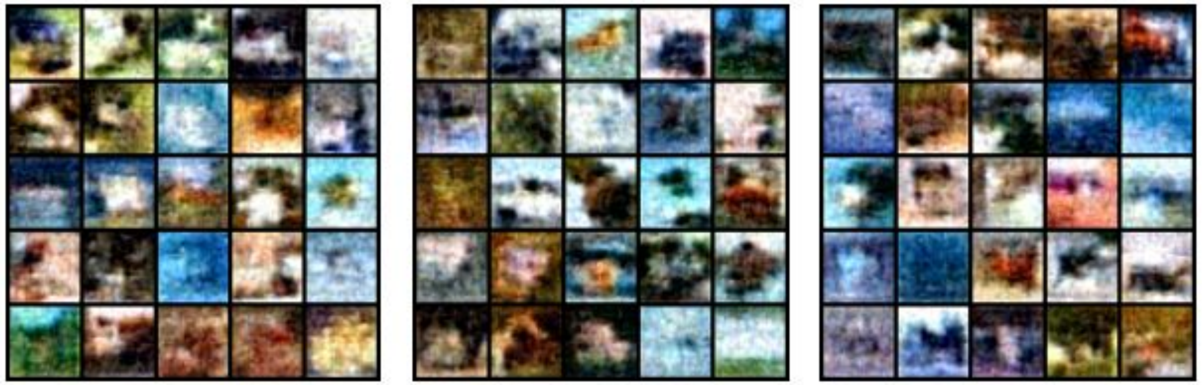
- Maintained high-quality image generation even at **99% sparsity**.
- Comparable results on **CIFAR-10** and **MNIST** datasets.

Literature Survey(Survey 4)



(a) CIFAR-10 (b) CIFAR-100 (c) MNIST

Figure 2: Performance of DDPM w.r.t. sparsity of Unet



(a) Original DDPM (b) Sub-network (Sparsity: 67.2% (c) Winning ticket (Sparsity: 99.4%))

(a) Original DDPM

(b) Winning ticket (Sparsity: 99.4%)

Literature Survey(Survey 4)

PES UNIVERSITY

Implications

- **Efficient Model Deployment**: The successful pruning of diffusion models enables deploying high-performing generative models on **resource-constrained devices** (e.g., mobile, IoT).
- **Scalability**: The technique offers potential for scaling diffusion models in real-world applications where computational resources are limited.
- **Energy Savings**: Reduced computational overhead translates into significant energy savings in training and inference, making large-scale AI applications more sustainable.

Future Research Directions

- **Pruning in Different Types of Diffusion Models:** Explore the application of the Lottery Ticket Hypothesis to various diffusion models (e.g., Variational Diffusion Models, Latent Diffusion Models, Score-Based Diffusion Models).
- **Advanced Pruning Strategies**: Investigate more sophisticated layer-wise pruning strategies to enhance **fine-grained control** over sparsity.
- **Real-World Applications**: Further study the application of pruned diffusion models in areas such as **real-time image synthesis, video generation, and multimodal tasks**.
- **Transfer Learning with Pruned Models**: Research how pruned diffusion models perform in **transfer learning** scenarios and across different datasets.

Summary of Literature Survey

Summary of Literature Review

Project Title: Exploring the Applicability of the Lottery Ticket Hypothesis to Diffusion Models

Team Members: Gauthama Siddharth, Hrishikesh V, Jagrath P Patel, M Barath Vikraman
Key Insights:
Survey1: Focuses on providing a better understanding of the Lottery Ticket Hypothesis and its original pruning models, which acts as the foundation for all further research topics.
Survey2: Provides improved pruning methods while also proposing reusability of similar winning tickets to sometimes completely avoid the pruning process itself.
Survey3: Suggests solutions to improve the efficiency and speed of pruning so that we can arrive at the winning tickets faster and avoid computational complexities.
Survey4: Elaborates the problem statement and proves the possibilities of the project by showing off the implementation of some of the recent features.

Summary of Literature Survey

The literature review reveals a clear progression in the understanding and application of lottery ticket hypothesis, from its original conception to its specialized adaptation for diffusion models. The original LTH, introduced by Frankle and Carbin, demonstrated that dense neural networks contain sparse subnetworks capable of training to similar accuracy when reset to their initial weights.

Methodological Evolution and Cross-Domain Applications

1. **Classical LTH to Elastic LTH**: The elastic variant demonstrated greater flexibility by allowing weight reinitialization within specific ranges, addressing the original hypothesis's limitation of strict initial weight preservation. This advancement showed that the "winning tickets" are more robust than initially thought.
2. **Dual LTH**: Extended the concept by identifying complementary subnetworks, suggesting that neural networks contain multiple viable sparse architectures. This revelation has important implications for model redundancy and optimization.

Any other information

Provide any other information you wish to add on.

Note: Changes can be made in the template, with the consent of the guide for inclusion of any other information.

Conclusion

- **Research Significance**: This project aims to explore the applicability of the Lottery Ticket Hypothesis (LTH) to diffusion models, revealing how sparse subnetworks can enhance model efficiency without sacrificing performance.
- **Key Objectives**:
  - Identify winning tickets within diffusion models, enabling significant computational savings.
  - Develop innovative pruning strategies to optimize resource utilization.
  - Assess the transferability and adaptability of pruned subnetworks across various architectures.
- **Potential Impact**: The findings can lead to more efficient deployment of generative models in resource-constrained environments, promoting sustainability in AI applications.
- **Future Directions**:
  - Expand the investigation to diverse diffusion model architectures.
  - Explore advanced layer-wise pruning strategies.
  - Investigate the effectiveness of pruned models in real-world applications, such as real-time image synthesis and video generation.

References

[1] J. Frankle and M. Carbin, "The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks," arXiv preprint arXiv:1803.03635, 2018. [Online]. Available: https://arxiv.org/abs/1803.03635

[2] B. Liu, Z. Zhang, P. He, Z. Wang, Y. Xiao, R. Ye, Y. Zhou, W.-S. Ku, and B. Hui, "A Survey of Lottery Ticket Hypothesis," arXiv preprint arXiv:2403.04861v1 [cs.LG], 2024. [Online]. Available: https://arxiv.org/abs/2403.04861v1

[3] X. Chen, Y. Cheng, S. Wang, Z. Gan, J. Liu, and Z. Wang, "The Elastic Lottery Ticket Hypothesis," arXiv preprint arXiv:2103.16547, submitted on Mar. 30, 2021; revised on Oct. 28, 2021. [Online]. Available: https://arxiv.org/abs/2103.16547.

[4] Y. Bai, H. Wang, Z. Tao, K. Li, and Y. Fu, "Dual Lottery Ticket Hypothesis," arXiv preprint arXiv:2203.04248, 2022. [Online]. Available: https://arxiv.org/abs/2203.04248.

[5] C. Jiang, B. Hui, B. Liu, and D. Yan, "Successfully Applying Lottery Ticket Hypothesis to Diffusion Model," arXiv preprint arXiv:2310.18823, 2023. [Online]. Available: https://arxiv.org/abs/2310.18823.

[6] Jiafeng Mao, Xueting Wang, and Kiyoharu Aizawa, "The Lottery Ticket Hypothesis in Denoising: Towards Semantic-Driven Initialization," arXiv preprint arXiv:2312.08872v4, 2023.[Online].Available: https://arxiv.org/pdf/2312.08872

References

[7] L. Yang et al., "Diffusion Models: A Comprehensive Survey of Methods and Applications," arXiv, 2022. [Online]. Available: https://arxiv.org/pdf/2209.00796.

[8] A. Elhage et al., "Toy Models of Superposition," Transformer Circuits Thread, 2022. [Online]. Available: https://transformer-circuits.pub/2022/toy_model/index.html.

[9] A. Bricken et al., "Towards Monosemanticity: Decomposing Language Models With Dictionary Learning," Transformer Circuit Thread, 2023. [Online]. Available:https://transformer-circuits.pub/2023/monosemantic-features/index.html

Thank You