Novelty Assessment Report

Paper: FlashRNN: Unlocking Parallel Training of Nonlinear RNNs for Large Language Models
PDF URL: https://openreview.net/pdf?id=mX8b64iUaa
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2026-01-01

Abstract

Recurrent Neural Networks (RNNs) laid the foundation for sequence modeling, but their intrinsic sequential nature restricts parallel computation, creating a fundamental barrier to scaling. This has led to the dominance of parallelizable architectures like Transformers and, more recently, State Space Models (SSMs). While SSMs achieve efficient parallelization through structured linear recurrences, this linearity constraint limits their expressive power and precludes modeling complex, nonlinear sequence-wise dependencies. To address this, we present FlashRNN, a framework that breaks the sequence-parallelization barrier for nonlinear RNNs. Building on prior work, we cast the sequence of nonlinear recurrence relationships as a single system of equations, which we solve in parallel using Newton's iterations combined with custom parallel reductions. Our implementation achieves speedups of up to $665\times$ over na\"ive sequential application, allowing training nonlinear RNNs at unprecedented scales. To showcase this, we apply FlashRNN to adaptations of LSTM and GRU architectures, successfully training models of 7B parameters that attain perplexity comparable to similarly-sized Transformers and Mamba2 architectures. To accelerate research in efficient sequence modeling, we release the FlashRNN codebase as an open-source framework for automatic training-parallelization of nonlinear RNNs, enabling researchers and practitioners to explore new nonlinear RNN models at scale.

Core Task Landscape

This paper addresses: **parallel training of nonlinear recurrent neural networks**

A total of **35 papers** were analyzed and organized into a taxonomy with **15 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Parallelization Frameworks and Algorithms**
- **Novel Nonlinear RNN Architectures**
- **Applications and Domain-Specific Architectures**

Complete Taxonomy Tree

- parallel training of nonlinear recurrent neural networks Survey Taxonomy
- Parallelization Frameworks and Algorithms
  - Sequence-Level Parallelization via Fixed-Point Formulations ★ (3 papers)
  - [0] FlashRNN: Unlocking Parallel Training of Nonlinear RNNs for Large Language Models (Anon et al., 2026) View paper
  - [2] Pararnn: Unlocking parallel training of nonlinear rnns for large language models (Danieli, 2025) View paper
  - [31] Towards Scalable and Stable Parallelization of Nonlinear RNNs (Xavier GonzÃ¡lez, 2024) View paper
  - Non-Iterative and Extreme Learning Machine Approaches (1 papers)
  - [1] An optimized parallel implementation of non-iteratively trained recurrent neural networks (Julia El Zini, 2021) View paper
  - Data-Parallel and Distributed Training Strategies (5 papers)
  - [7] Parallel implementations of recurrent neural network learning (Andrej Dobnikar, 2009) View paper
  - [21] Empirical investigation of stale value tolerance on parallel RNN training (Joo Hwan Lee, 2019) View paper
  - [26] Concurrent asynchronous learning algorithms for massively parallel recurrent neural networks (C. Wu, 1992) View paper
  - [33] Exponential Moving Average Model in Parallel Speech Recognition Training (Tian Xu, 2022) View paper
  - [35] Distributed Supervised Learning using Neural Networks (Scardapane, 2016) View paper
  - Optimization and Efficiency Enhancements (3 papers)
  - [17] Accelerating recurrent neural network training using sequence bucketing and multi-GPU data parallelization (Khomenko, 2016) View paper
  - [32] Single stream parallelization of generalized LSTM-like RNNs on a GPU (Hwang, 2022) View paper
  - [34] Training recurrent neural networks for dynamic system identification using parallel tabu search algorithm (DerviÅ KaraboÄa, 1997) View paper
- Novel Nonlinear RNN Architectures
  - Minimal and Convolutional RNN Variants (2 papers)
  - [3] Comba: Improving Nonlinear RNNs with Closed-loop Control (J Hu, 2025) View paper
  - [22] Minimal Convolutional RNNs Accelerate Spatiotemporal Learning (CoÅku Can Horuz, 2025) View paper
  - Implicit and Fixed-Point RNN Models (1 papers)
  - [25] Implicit Language Models are RNNs: Balancing Parallelization and Expressivity (SchÃ¶ne, 2025) View paper
  - Hybrid and Fusion Architectures (3 papers)
  - [4] Hybrid series/parallel all-nonlinear dynamic-static neural networks: development, training, and application to chemical processes (Angan Mukherjee, 2023) View paper
  - [14] A Novel Parallel Recurrent Fusion Network for Stock Market Forecasting (Meet Brijwani, 2025) View paper

- [18] A parallel-fusion RNN-LSTM architecture for image caption generation (Minsi Wang, 2016) View paper
- Self-Organizing and Adaptive RNN Structures (2 papers)
- [15] Nonlinear model predictive control based on a self-organizing recurrent neural network (Hong-gui Han, 2015) View paper
- [28] Design of an SCRFNN-based nonlinear channel equaliser (Ruiâ□□Chang Lin, 2005) View paper
- Specialized Memory and Gating Mechanisms (1 papers)
- [30] LMUFormer: Low Complexity Yet Powerful Spiking Model With Legendre Memory Units (Liu, 2024) View paper

• Applications and Domain-Specific Architectures
- Control and Optimization Applications (8 papers)
- [6] Safety-Certified Multi-Target Circumnavigation With Autonomous Surface Vehicles via Neurodynamics-Driven Distributed Optimization (Yue Jiang, 2024) View paper
- [8] A recurrent neural network-based identification of complex nonlinear dynamical systems: a novel structure, stability analysis and a comparative study (R. Shobana, 2023) View paper
- [9] Multiple Mittagâ□□Leffler Stability of Almost Periodic Solutions for Fractional-Order Delayed Neural Networks: Distributed Optimization Approach (Chenxi Song, 2023) View paper
- [10] Machine learningâ□□based distributed model predictive control of nonlinear processes (Scarlett Chen, 2020) View paper
- [11] Optimal adaptive output regulation of uncertain nonlinear discrete-time systems using lifelong concurrent learning (R. Moghadam, 2022) View paper
- [12] Graph Neural Network-Based Distributed Optimal Control for Linear Networked Systems: An Online Distributed Training Approach (Song, 2025) View paper
- [19] Stabilization of nonlinear nonminimum phase systems: adaptive parallel approach using recurrent fuzzy neural network (C.-H. Lee, 2004) View paper
- [20] Nonlinear model-predictive control for industrial processes: An application to wastewater treatment process (Honggui Han, 2013) View paper
- Time-Series Forecasting and Prediction (1 papers)
- [27] Score Prediction of Sports Events Based on Parallel Self-Organizing Nonlinear Neural Network. (Junyao Ling, 2022) View paper
- Signal Processing and Equalization (1 papers)
- [29] A Nonlinear Concurrent Butterfly Equalizer (K. S. Mayer, 2024) View paper
- Computer Vision and Multimodal Tasks (2 papers)
- [5] Parallel sequence classification using recurrent neural networks and alignment (Federico Raue, 2015) View paper
- [16] Facial Landmark Detection (Romain BELMONTE, 2022) View paper
- Process Monitoring and Feature Learning (1 papers)
- [13] A novel spatiotemporal process feature learning method based on the pseudo-siamese network for complex chemical process concurrent condition monitoring (Yuemei Xu, 2022) View paper
- General Neural Network Theory and Surveys (2 papers)
- [23] Intelligent computational algorithms based on neural networks: a survey (Min Yang, 2025) View paper
- [24] A modified particle swarm optimization-based dynamic recurrent neural network for identifying and controlling nonlinear systems (H. Ge, 2007) View paper

Narrative

Core task: parallel training of nonlinear recurrent neural networks. The field organizes around three main branches that reflect distinct research emphases. The Parallelization Frameworks and Algorithms branch focuses on computational strategies to accelerate RNN training, including data-parallel methods, sequence-level parallelization via fixed-point formulations, and hardware-aware optimizations that exploit GPU or distributed architectures. The Novel Nonlinear RNN Architectures branch explores new model designs—such as minimal convolutional variants, self-organizing structures, and hybrid series-parallel topologies—that balance expressiveness with trainability. The Applications and Domain-Specific Architectures branch addresses how RNNs are tailored to particular domains, from control and system identification to vision and language tasks, often incorporating domain constraints or specialized loss functions. Together, these branches capture the interplay between algorithmic innovation, architectural design, and practical deployment.

A particularly active line of work within parallelization explores sequence-level methods that reformulate recurrent dependencies as fixed-point problems, enabling greater concurrency during training. FlashRNN[0] sits squarely in this cluster, proposing a fixed-point formulation that allows parallel computation across time steps. It shares conceptual ground with Pararnn[2], which also targets sequence-level parallelism, and contrasts with more traditional data-parallel approaches like Optimized Parallel RNN[1] or hardware-centric schemes such as Single Stream GPU[32]. Meanwhile, works like Scalable Parallel RNNs[31] emphasize scalability across distributed systems, highlighting trade-offs between communication overhead and per-device computation. The central tension across these efforts is whether to parallelize over sequences, layers, or data batches, and how to manage the inherent sequential dependencies of recurrence without sacrificing model fidelity or convergence guarantees.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. Pararnn: Unlocking parallel training of nonlinear rnns for large language models

Authors: Danieli, Federico, Rodriguez, Pau, Federico Danieli, et al. (15 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Recurrent Neural Networks (RNNs) laid the foundation for sequence modeling, but their intrinsic sequential nature restricts parallel computation, creating a fundamental barrier to scaling. This has led to the dominance of parallelizable architectures like Transformers and, more recently, State Space Models (SSMs). While SSMs achieve efficient parallelization through structured linear recurrences, this linearity constraint limits their expressive power and precludes modeling complex, nonlinear se...

⚠ Similarity Notice

This paper appears to be a variant or near-duplicate of the original FlashRNN paper, with the primary difference being the framework name (ParaRNN vs. FlashRNN). The abstract describes essentially identical technical contributions: casting nonlinear RNN recurrences as systems of equations solved via Newton's method with parallel reductions, achieving 665× speedups, training 7B parameter LSTM/GRU models, and releasing an open-source framework. The core methodology, experimental results, and claims are nearly identical, suggesting this is likely a resubmission or variant of the same work.

---

2. Towards Scalable and Stable Parallelization of Nonlinear RNNs

Authors: Xavier GonzÃ¡lez, Scott Linderman, Jimmy Smith, Andrew Warrington | **Year/Venue**: 2024 | **URL**: View paper

Abstract

Transformers and linear state space models can be evaluated in parallel on modern hardware, but evaluating nonlinear RNNs appears to be an inherently sequential problem. Recently, however, Lim et al. '24 developed an approach called DEER, which evaluates nonlinear RNNs in parallel by posing the states as the solution to a fixed-point problem. They derived a parallel form of Newton's method to solve the fixed-point problem and achieved significant speedups over sequential evaluation. However, the...

Relationship Analysis

Both papers belong to the same taxonomy category of sequence-level parallelization via fixed-point formulations, using Newton-based iterative methods to parallelize nonlinear RNN training. They share the core approach of casting recurrent dependencies as fixed-point problems and solving them with parallel Newton iterations, both building on similar foundational work (DEER/Lim et al. '24). The key differences are that the original paper (FlashRNN) focuses on custom CUDA implementations with parallel reduction algorithms and demonstrates large-scale language modeling at 7B parameters, while the candidate paper emphasizes algorithmic improvements through quasi-Newton approximations and Levenberg-Marquardt stabilization (ELK) to address computational complexity and numerical stability issues.

Contributions Analysis

Overall novelty summary. The paper introduces FlashRNN, a framework enabling parallel training of nonlinear RNNs by casting recurrence relationships as a system of equations solved via Newton's method and custom parallel reductions. It resides in the 'Sequence-Level Parallelization via Fixed-Point Formulations' leaf, which contains only three papers total. This leaf sits within the broader 'Parallelization Frameworks and Algorithms' branch, indicating a relatively sparse but well-defined research direction focused on iterative fixed-point methods rather than data-parallel or non-iterative strategies.

The taxonomy reveals neighboring leaves addressing complementary parallelization strategies: 'Data-Parallel and Distributed Training Strategies' (five papers) focuses on multi-worker synchronization, while 'Non-Iterative and Extreme Learning Machine Approaches' (one paper) eliminates backpropagation through time entirely. The 'Novel Nonlinear RNN Architectures' branch explores architectural innovations like minimal convolutional variants and hybrid fusion models, which often assume or enable parallelization but do not primarily contribute algorithmic frameworks. FlashRNN's fixed-point formulation bridges algorithmic parallelization with architectural adaptations (LSTM/GRU), connecting these two major branches.

Among thirty candidates examined, the FlashRNN framework itself (Contribution A) and adapted LSTM/GRU architectures (Contribution B) show no clear refutation across ten candidates each, suggesting limited direct overlap in the examined literature. However, the open-source PyTorch+CUDA library (Contribution C) encountered two refutable candidates among ten examined, indicating prior implementations or tools with overlapping functionality. The framework and architectural contributions appear more novel within this search scope, while the software artifact faces stronger prior work in the examined sample.

Based on the limited top-30 semantic search, FlashRNN occupies a sparsely populated methodological niche—sequence-level fixed-point parallelization—with only two sibling papers in its taxonomy leaf. The framework and architectural adaptations show stronger novelty signals than the software library component. This assessment reflects the examined candidate set and does not claim exhaustive coverage of all relevant prior work in parallel RNN training or open-source tooling.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: FlashRNN framework for parallel training of nonlinear RNNs

Description: The authors introduce FlashRNN, a framework that enables parallel training of nonlinear recurrent neural networks by casting the sequence of nonlinear recurrence relationships as a system of equations solved using Newton iterations combined with custom parallel reductions. This overcomes the traditional sequential computation barrier that has limited RNN scalability.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Comba: Improving Nonlinear RNNs with Closed-loop Control

URL: View paper

Brief Assessment

Comba[3] focuses on a specific nonlinear RNN variant with delta learning rule for memory management, not a general framework for parallelizing arbitrary nonlinear RNNs through Newton iterations and parallel reductions as in the original paper.

---

2. An optimized parallel implementation of non-iteratively trained recurrent neural networks

URL: View paper

Brief Assessment

Optimized Parallel RNN[1] focuses on parallelizing non-iterative ELM-based training algorithms for RNNs, not on enabling parallel training of nonlinear RNNs through Newton iterations and parallel reductions as in the original paper's FlashRNN framework.

---

3. Adjoint recurrent neural network technique for nonlinear electronic component modeling

URL: View paper

Brief Assessment

Adjoint RNN Modeling[48] focuses on electronic component modeling using derivative information for training efficiency, not on sequence-parallelization methods for general nonlinear RNN training in language modeling contexts.

---

4. Hybrid series/parallel all-nonlinear dynamic-static neural networks: development, training, and application to chemical processes

URL: View paper

Brief Assessment

Hybrid Series Parallel[4] focuses on hybrid static-dynamic neural network architectures for chemical process modeling, not on parallel training methods for recurrent neural networks. The paper addresses parameter estimation for hybrid networks combining static and dynamic components, which is fundamentally different from FlashRNN's sequence-parallelization approach using Newton iterations.

---

5. Resurrecting recurrent neural networks for long sequences

URL: View paper

Brief Assessment

Resurrecting Recurrent Networks[45] focuses on linear recurrent units (LRUs) with diagonal recurrence matrices, not on parallel training methods for general nonlinear RNNs. The candidate explicitly states 'we show that careful design of deep RNNs using standard signal propagation arguments can recover the impressive performance of deep SSMs' through linearization, which is fundamentally different from the original paper's approach of parallelizing nonlinear recurrences via Newton iterations.

---

6. Neural Network Approaches for Intelligent Decisionâ□□Making in Automation

URL: View paper

Brief Assessment

Neural Decision Making[47] discusses network topologies and non-linearity in general terms but does not address parallel training methods for recurrent neural networks or sequence-parallelization techniques.

---

7. A recurrent neural network-based identification of complex nonlinear dynamical systems: a novel structure, stability analysis and a comparative study

URL: View paper

Brief Assessment

RNN Identification Nonlinear[8] focuses on system identification and control applications using RNNs for modeling nonlinear dynamical systems, not on parallel training methods or computational efficiency improvements for RNN training at scale.

---

8. Recurrent neural network for the identification of nonlinear dynamical systems: A comparative study

URL: View paper

Brief Assessment

RNN Identification Comparative[50] focuses on comparing different RNN architectures for system identification tasks in dynamical systems, not on developing parallel training frameworks for nonlinear RNNs at scale for language modeling.

---

9. Machineâ□□learningâ□□based predictive control of nonlinear processes. Part II: Computational implementation

URL: View paper

Brief Assessment

Machine Learning MPC[49] focuses on implementing RNN models within model predictive control for chemical processes, not on developing parallel training frameworks for nonlinear RNNs at scale for language modeling tasks.

---

10. Hybrid data-model parallel training for sequence-to-sequence recurrent neural network machine translation

URL: View paper

Brief Assessment

Hybrid Data Model[46] focuses on distributing training across multiple GPUs using data and model parallelism for seq2seq models, not on parallelizing computation along the sequence length dimension for nonlinear RNNs as FlashRNN does.

Contribution 2: Adapted LSTM and GRU architectures for large-scale training

Description: The authors demonstrate that classical nonlinear RNN models (LSTM and GRU) can be trained at unprecedented scales of 7 billion parameters using FlashRNN, achieving competitive performance with Transformers and Mamba2 on language modeling tasks. This shows that nonlinear RNNs remain viable alternatives when computational barriers are removed.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Prediction of super-large diameter shield attitude based on LSTM-Transformer

URL: View paper

Brief Assessment

Shield Attitude LSTM[52] applies LSTM to shield attitude prediction in tunnel construction, not large-scale language model training. The candidate focuses on civil engineering applications with different objectives and scale requirements.

---

2. A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks

URL: View paper

Brief Assessment

Multi Head Attention[58] focuses on traffic flow forecasting using transformers compared to RNNs, not on large-scale training of LSTM/GRU models or architectural adaptations for parallel training at billion-parameter scales.

---

3. A comparative analysis of LSTM, GRU, and Transformer models for construction cost prediction with multidimensional feature integration

URL: View paper

Brief Assessment

Construction Cost Prediction[56] focuses on comparing LSTM, GRU, and Transformer models for construction cost prediction tasks, not on large-scale training of these architectures or removing computational barriers for language modeling.

---

4. LSTMâ□□transformer-based robust hybrid deep learning model for financial time series forecasting

URL: View paper

Brief Assessment

LSTM Transformer Hybrid[60] focuses on financial time series forecasting using LSTM combined with transformers for stock price prediction, not on large-scale language model training or parallel training methods for nonlinear RNNs.

---

5. Recurrent neural networks: A comprehensive review of architectures, variants, and applications

URL: View paper

Brief Assessment

Recurrent Networks Review[51] is a general survey paper covering various RNN architectures and applications. It does not present empirical work on training 7B parameter LSTM/GRU models or demonstrate competitive performance with Transformers at scale, which is the core novelty claim of the original paper.

---

6. Comparative Study of LSTM and Transformer

URL: View paper

Brief Assessment

LSTM Transformer Comparative[54] focuses on comparing standard LSTM and Transformer architectures for stock price prediction tasks, not on large-scale training of nonlinear RNNs or architectural adaptations for parallel training at billion-parameter scales.

---

7. Time series forecasting using deep learning: a comparative study of LSTM, GRU, and transformer models

URL: View paper

Brief Assessment

Time Series Forecasting[57] focuses on comparative analysis of LSTM, GRU, and Transformer models for time series prediction tasks in finance, energy, and healthcare domains. It does not address large-scale training (7B parameters) or language modeling tasks, which are the core novelty claims of the original paper.

---

8. Low-Resource Neural Machine Translation Using Recurrent Neural Networks and Transfer Learning: A Case Study on English-to-Igbo

URL: View paper

Brief Assessment

Low Resource Translation[53] focuses on English-to-Igbo translation using standard LSTM/GRU architectures for a low-resource language task, not on large-scale training (7B parameters) or parallel training methods like the original paper's FlashRNN framework.

---

9. Comparative Analysis of Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) and Transformer Models in Predicting Stock Prices

URL: View paper

Brief Assessment

Stock Price Prediction[59] focuses on comparing LSTM, GRU, and Transformer models for stock price forecasting tasks with small-scale models, not on large-scale training (7B parameters) or language modeling as in the original paper.

---

10. Advanced hybrid LSTM-transformer architecture for real-time multi-task prediction in engineering systems

URL: View paper

Brief Assessment

Hybrid LSTM Transformer[55] focuses on engineering system predictions (underground drilling, stormwater management) using a hybrid architecture for multi-task learning, not on large-scale language model training or demonstrating that classical RNNs can scale to 7B parameters for language modeling tasks.

Contribution 3: Open-source PyTorch+CUDA library for automatic RNN parallelization

Description: The authors provide a high-performance PyTorch and CUDA library that automates sequence-parallel training for any nonlinear RNN cell from only the specification of its recurrence step. This enables researchers to explore new nonlinear RNN architectures at scale without manually implementing the underlying parallelization complexity.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Parallelizing non-linear sequential models over the sequence length

URL: View paper

Prior Art Analysis

Parallelizing Sequence Length[37] demonstrates that prior work exists on parallel algorithms for sequential models. Both papers address the same fundamental problem: parallelizing non-linear sequential models (specifically RNNs) over the sequence length to overcome training bottlenecks. The candidate paper explicitly states it provides a 'parallel algorithm that accelerates gpu evaluation of sequential models' and makes sequential models applicable 'to a wide range of architectures' without needing 'any special structure', which directly overlaps with the original paper's claim of providing an automated library for any nonlinear RNN cell.

Evidence

Evidence 1 - **Rationale**: Both papers claim to provide general-purpose parallelization solutions for sequential models. The candidate explicitly states their algorithm works without special architectural structure, directly paralleling the original's claim of automating parallelization for 'any nonlinear cell'. - **Original**: we enable the exploration of new nonlinear rnn architectures for language modeling at scale, by introducing flashrnn: a high-performance pytorch+cuda library that implements sequence-parallel training for any nonlinear cell from only the specification of its recurrence step, thereby automating the u... - **Candidate**: we challenge this long-held belief with our parallel algorithm that accelerates gpu evaluation of sequential models by up to 3 orders of magnitude faster without compromising output accuracy. the algorithm does not need any special structure in the sequential models' architecture, making it applicab...

Evidence 2 - **Rationale**: Both papers claim to overcome the training bottleneck for non-linear sequential models and enable their use at scale. The candidate's claim to be 'the first step to unlock the potential' suggests prior work in this area, which refutes the original's novelty claim. - **Original**: to accelerate research in efficient sequence modeling, we release the flashrnn codebase as an open-source framework for automatic training-parallelization of nonlinear rnns, enabling researchers and practitioners to explore new nonlinear rnn models at scale. - **Candidate**: by overcoming the training bottleneck, our work serves as the first step to unlock the potential of non-linear sequential models for long sequence problems.

---

2. Parallelizing linear recurrent neural nets over sequence length

URL: View paper

Prior Art Analysis

Parallelizing Linear Recurrent[44] demonstrates prior work on developing CUDA kernels for parallelizing RNN training. The candidate paper explicitly describes developing 'a parallel linear recurrence cuda kernel' that can be applied to 'immediately speed up training and inference of several state of the art rnn architectures by up to 9x.' This shows that CUDA-based parallelization libraries for RNN training existed before the original paper's submission, directly challenging the novelty claim of being the first to provide such a library.

Evidence

Evidence 1 - **Rationale**: Both papers describe developing CUDA kernels for parallelizing RNN training. The candidate demonstrates that such CUDA-based parallelization implementations existed prior to the original work, refuting the claim of being first to provide this type of library. - **Original**: we enable the exploration of new nonlinear rnn architectures for language modeling at scale, by introducing flashrnn: a high-performance pytorch+cuda library that implements sequence-parallel training for any nonlinear cell from only the specification of its recurrence step, thereby automating the u... - **Candidate**: we develop a parallel linear recurrence cuda kernel and show that it can be applied to immediately speed up training and inference of several state of the art rnn architectures by up to 9x.

Evidence 2 - **Rationale**: Both papers address the fundamental challenge of parallelizing RNN training over sequence length and provide solutions. The candidate's work on parallel scan algorithms for RNN training predates the original paper's claims of enabling such parallelization. - **Original**: to accelerate research in efficient sequence modeling, we release the flashrnn codebase as an open-source

framework for automatic training-parallelization of nonlinear rnns, enabling researchers and practitioners to explore new nonlinear rnn models at scale. - **Candidate**: recurrent neural networks (rnns) are widely used to model sequential data but their non-linear dependencies between sequence elements prevent parallelizing training over sequence length. we show the training of rnns with only linear sequential dependencies can be parallelized over the sequence lengt...

---

3. An optimized parallel implementation of non-iteratively trained recurrent neural networks

URL: View paper

Brief Assessment

Optimized Parallel RNN[1] implements GPU-accelerated ELM training for specific RNN architectures but does not provide an automatic parallelization library that works from only the specification of a recurrence step, as claimed in the original paper.

---

4. Graph computing system and application based on large-scale information network

URL: View paper

Brief Assessment

Graph Computing System[43] focuses on graph computing systems for large-scale information networks, not on RNN parallelization libraries or sequence modeling frameworks.

---

5. Parallelizing legendre memory unit training

URL: View paper

Brief Assessment

Parallelizing Legendre Memory[39] focuses on parallelizing a specific linear time-invariant (LTI) system (the Legendre Memory Unit) rather than providing a general framework for arbitrary nonlinear RNN cells. The candidate does not demonstrate prior work on automatic parallelization libraries for general nonlinear RNNs.

---

6. Supporting very large models using automatic dataflow graph partitioning

URL: View paper

Brief Assessment

Automatic Dataflow Partitioning[36] focuses on partitioning dataflow graphs across multiple GPUs to reduce memory footprint, not on sequence-parallel training of RNNs. The systems address fundamentally different parallelization problems (model partitioning vs. sequence parallelization).

---

7. Accelerating rnn controllers with parallel computing and weight dropout techniques

URL: View paper

Brief Assessment

Parallel Weight Dropout[38] focuses on accelerating RNN controllers using parallel computing and weight dropout techniques for specific applications, not on providing a general-purpose automatic parallelization library for arbitrary nonlinear RNN architectures.

---

8. AutoML with parallel genetic algorithm for fast hyperparameters optimization in efficient IoT time series prediction

URL: View paper

Brief Assessment

AutoML Parallel Genetic[40] focuses on hyperparameter optimization for LSTM models using parallel genetic algorithms, not on providing a library for automatic parallelization of RNN training. The candidate addresses a completely different problem domain (hyperparameter tuning) rather than training parallelization infrastructure.

---

9. FINN-L: Library extensions and design trade-off analysis for variable precision LSTM networks on FPGAs

URL: View paper

Brief Assessment

FINN Library Extensions[41] focuses on FPGA hardware implementations of LSTM networks with variable precision, not on automatic parallelization libraries for RNN training. The candidate provides HLS library extensions for hardware architectures, which is fundamentally different from the original paper's software-based training parallelization framework.

---

10. An Adaptive Dropout and Parallel Computing Approaches for Accelerating RNN Controller

URL: View paper

Brief Assessment

Adaptive Dropout Parallel[42] focuses on accelerating RNN controller training for closed-loop control systems (e.g., solar inverters) using adaptive dropout and parallel computing with Levenberg-Marquardt algorithm. It does not address automatic parallelization libraries for general RNN architectures or sequence-parallel training frameworks.

Appendix: Text Similarity Detection

Textual similarity detection checked 31 papers and found 1 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. Pararnn: Unlocking parallel training of nonlinear rnns for large language models

Detected in: Core Task (sibling)

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] FlashRNN: Unlocking Parallel Training of Nonlinear RNNs for Large Language Models View paper
- [1] An optimized parallel implementation of non-iteratively trained recurrent neural networks View paper
- [2] Pararnn: Unlocking parallel training of nonlinear rnns for large language models View paper
- [3] Comba: Improving Nonlinear RNNs with Closed-loop Control View paper
- [4] Hybrid series/parallel all-nonlinear dynamic-static neural networks: development, training, and application to chemical processes View paper

- [5] Parallel sequence classification using recurrent neural networks and alignment View paper
- [6] Safety-Certified Multi-Target Circumnavigation With Autonomous Surface Vehicles via Neurodynamics-Driven Distributed Optimization View paper
- [7] Parallel implementations of recurrent neural network learning View paper
- [8] A recurrent neural network-based identification of complex nonlinear dynamical systems: a novel structure, stability analysis and a comparative study View paper
- [9] Multiple Mittagâ□□Leffler Stability of Almost Periodic Solutions for Fractional-Order Delayed Neural Networks: Distributed Optimization Approach View paper
- [10] Machine learningâ□□based distributed model predictive control of nonlinear processes View paper
- [11] Optimal adaptive output regulation of uncertain nonlinear discrete-time systems using lifelong concurrent learning View paper
- [12] Graph Neural Network-Based Distributed Optimal Control for Linear Networked Systems: An Online Distributed Training Approach View paper
- [13] A novel spatiotemporal process feature learning method based on the pseudo-siamese network for complex chemical process concurrent condition monitoring View paper
- [14] A Novel Parallel Recurrent Fusion Network for Stock Market Forecasting View paper
- [15] Nonlinear model predictive control based on a self-organizing recurrent neural network View paper
- [16] Facial Landmark Detection View paper
- [17] Accelerating recurrent neural network training using sequence bucketing and multi-GPU data parallelization View paper
- [18] A parallel-fusion RNN-LSTM architecture for image caption generation View paper
- [19] Stabilization of nonlinear nonminimum phase systems: adaptive parallel approach using recurrent fuzzy neural network View paper
- [20] Nonlinear model-predictive control for industrial processes: An application to wastewater treatment process View paper
- [21] Empirical investigation of stale value tolerance on parallel RNN training View paper
- [22] Minimal Convolutional RNNs Accelerate Spatiotemporal Learning View paper
- [23] Intelligent computational algorithms based on neural networks: a survey View paper
- [24] A modified particle swarm optimization-based dynamic recurrent neural network for identifying and controlling nonlinear systems View paper
- [25] Implicit Language Models are RNNs: Balancing Parallelization and Expressivity View paper
- [26] Concurrent asynchronous learning algorithms for massively parallel recurrent neural networks View paper
- [27] Score Prediction of Sports Events Based on Parallel Self-Organizing Nonlinear Neural Network. View paper
- [28] Design of an SCRFNN-based nonlinear channel equaliser View paper
- [29] A Nonlinear Concurrent Butterfly Equalizer View paper
- [30] LMUFormer: Low Complexity Yet Powerful Spiking Model With Legendre Memory Units View paper
- [31] Towards Scalable and Stable Parallelization of Nonlinear RNNs View paper
- [32] Single stream parallelization of generalized LSTM-like RNNs on a GPU View paper
- [33] Exponential Moving Average Model in Parallel Speech Recognition Training View paper
- [34] Training recurrent neural networks for dynamic system identification using parallel tabu search algorithm View paper
- [35] Distributed Supervised Learning using Neural Networks View paper
- [36] Supporting very large models using automatic dataflow graph partitioning View paper
- [37] Parallelizing non-linear sequential models over the sequence length View paper
- [38] Accelerating rnn controllers with parallel computing and weight dropout techniques View paper
- [39] Parallelizing legendre memory unit training View paper
- [40] AutoML with parallel genetic algorithm for fast hyperparameters optimization in efficient IoT time series prediction View paper
- [41] FINN-L: Library extensions and design trade-off analysis for variable precision LSTM networks on FPGAs View paper
- [42] An Adaptive Dropout and Parallel Computing Approaches for Accelerating RNN Controller View paper
- [43] Graph computing system and application based on large-scale information network View paper
- [44] Parallelizing linear recurrent neural nets over sequence length View paper
- [45] Resurrecting recurrent neural networks for long sequences View paper
- [46] Hybrid data-model parallel training for sequence-to-sequence recurrent neural network machine translation View paper
- [47] Neural Network Approaches for Intelligent Decisionâ□□Making in Automation View paper
- [48] Adjoint recurrent neural network technique for nonlinear electronic component modeling View paper
- [49] Machineâ□□learningâ□□based predictive control of nonlinear processes. Part II: Computational implementation View paper
- [50] Recurrent neural network for the identification of nonlinear dynamical systems: A comparative study View paper
- [51] Recurrent neural networks: A comprehensive review of architectures, variants, and applications View paper
- [52] Prediction of super-large diameter shield attitude based on LSTM-Transformer View paper
- [53] Low-Resource Neural Machine Translation Using Recurrent Neural Networks and Transfer Learning: A Case Study on English-to-Igbo View paper
- [54] Comparative Study of LSTM and Transformer View paper
- [55] Advanced hybrid LSTM-transformer architecture for real-time multi-task prediction in engineering systems View paper
- [56] A comparative analysis of LSTM, GRU, and Transformer models for construction cost prediction with multidimensional feature integration View paper
- [57] Time series forecasting using deep learning: a comparative study of LSTM, GRU, and transformer models View paper
- [58] A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks View paper
- [59] Comparative Analysis of Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) and Transformer Models in Predicting Stock Prices View paper
- [60] LSTMâ□□transformer-based robust hybrid deep learning model for financial time series forecasting View paper