Novelty Assessment Report

Paper: MrRoPE: Mixed-radix Rotary Position Embedding
PDF URL: https://openreview.net/pdf?id=1J63FJYJKg
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-29

Abstract

Rotary Position Embedding (RoPE)-extension refers to modifying or generalizing the Rotary Position Embedding scheme to handle longer sequences than those encountered during pre-training. However, current extension strategies are highly diverse and lack a unified theoretical foundation. In this paper, we propose $\textbf{\textit{MrRoPE (Mixed-radix RoPE)}}$, a generalized encoding formulation based on a radix system conversion perspective, which elegantly unifies various RoPE-extension approaches as distinct radix conversion strategies. Based on this theory, we introduce two training-free extensions, $\textbf{\textit{MrRoPE-Uni}}$ and $\textbf{\textit{MrRoPE-Pro}}$, which leverage uniform and progressive radix conversion strategies, respectively, to achieve “train short, test long” generalization. Without fine-tuning, MrRoPE-Pro sustains over 85% recall in the 128K-context Needle-in-a-Haystack test and achieves more than double YaRN’s accuracy on Infinite-Bench retrieval and dialogue subsets. Theoretical analysis confirms that MrRoPE-Pro effectively raises the upper bound of RoPE's attainable encoding length, which further validates the reliability and utility of our theory and methodology.

Core Task Landscape

This paper addresses: **extending context window of rotary position embedding in language models**
A total of **50 papers** were analyzed and organized into a taxonomy with **22 categories**.

Taxonomy Overview

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

- **RoPE Modification and Rescaling Methods**
- **Training-Free and Inference-Time Adaptation**
- **Alternative and Hybrid Position Encoding Architectures**
- **Training Strategies and Data Efficiency**
- **Analysis and Understanding of RoPE Extensions**
- **Domain-Specific and Multimodal Applications**
- **Efficient Inference and Computational Optimization**
- **Extrapolation and Generalization Beyond Training Length**

Complete Taxonomy Tree

- extending context window of rotary position embedding in language models Survey Taxonomy
- RoPE Modification and Rescaling Methods
  - Base and Frequency Adjustment Approaches (4 papers)
  - [16] Base of rope bounds context length (Men Xin, 2024) View paper
  - [18] Scaling laws of rope-based extrapolation (Liu Xiaoran, 2023) View paper
  - [23] Extending context window in large language models with segmented base adjustment for rotary position embeddings (Rongsheng Li, 2024) View paper
  - [29] An efficient recipe for long context extension via middle-focused positional encoding (Tong Wu, 2024) View paper
  - Position Interpolation and Scaling (4 papers)
  - [1] Yarn: Efficient context window extension of large language models (Peng Bowen, 2023) View paper
  - [2] Longrope: Extending llm context window beyond 2 million tokens (Ding Yiran, 2024) View paper
  - [7] Extending context window of large language models via positional interpolation (Chen, 2023) View paper
  - [22] LongRoPE2: Near-Lossless LLM Context Window Scaling (Shang Ning, 2025) View paper
  - Dimension-Wise and Segmented Manipulation (3 papers)
  - [17] 3d-rpe: Enhancing long-context modeling through 3d rotary position encoding (Liu Wenyuan, 2025) View paper
  - [31] On the token distance modeling ability of higher rope attention dimension (Jiang Che, 2024) View paper
  - [41] Effective Length Extrapolation via Dimension-Wise Positional Embeddings Manipulation (Lu Yi, 2025) View paper
  - Unified Theoretical Frameworks for RoPE Extension ★ (2 papers)
  - [0] MrRoPE: Mixed-radix Rotary Position Embedding (Anon et al., 2026) View paper
  - [35] Extending Context Window of Large Language Models from a Distributional Perspective (WU Yingsheng, 2024) View paper
- Training-Free and Inference-Time Adaptation
  - Dynamic and Adaptive Position Encoding (3 papers)
  - [37] LaMPE: Length-aware Multi-grained Positional Encoding for Adaptive Long-context Scaling Without Training (Zhang Si-kui, 2025) View paper
  - [46] Extending LLM Context Window with Adaptive Grouped Positional Encoding: A Training-Free Method (Xinhao Xu, 2025) View paper

- [48] CoWPE: Adaptive Context Window Adjustment in LLMs for Complex Input Queries (Venkata Mohit Tamanampudi, 2024) View paper
  - Optimization-Based Extension Strategies (1 papers)
  - [14] Optimal RoPE extension via Bayesian Optimization for training-free length generalization (Xinrong Zhang, 2025) View paper
  - Denoising and Outlier Management (3 papers)
  - [26] DoPE: Denoising Rotary Position Embedding (Jing Xiong, 2025) View paper
  - [43] Q-ROAR: Outlier-Aware Rescaling for RoPE Position Interpolation in Quantized Long-Context LLMs (Qiao Ye, 2025) View paper
  - [47] Rethinking RoPE Scaling in Quantized LLM: Theory, Outlier, and Channel-Band Analysis with Weight Rescaling (Qiao Ye, 2025) View paper
- Alternative and Hybrid Position Encoding Architectures
  - Geometric and Complex-Valued Extensions (2 papers)
  - [8] HoPE: Hyperbolic Rotary Positional Encoding for Stable Long-Range Dependency Modeling in Large Language Models (Dai Chang, 2025) View paper
  - [24] Beyond Real: Imaginary Extension of Rotary Position Embeddings for Long-Context LLMs (Xiaoran Liu, 2025) View paper
  - Learnable and Context-Aware Position Encoding (3 papers)
  - [12] Context-aware Rotary Position Embedding (Veisi Ali, 2025) View paper
  - [21] SeqPE: Transformer with Sequential Position Encoding (Li, 2025) View paper
  - [34] Positional Encoding via Token-Aware Phase Attention (Wang Yu, 2025) View paper
  - Hybrid Attention and Memory Mechanisms (3 papers)
  - [15] SWAN-GPT: An Efficient and Scalable Approach for Long-Context Language Modeling (Puvvada, 2025) View paper
  - [25] Rope to nope and back again: A new hybrid attention strategy (Yang, 2025) View paper
  - [39] Recurrent Memory-Augmented Transformers with Chunked Attention for Long-Context Language Modeling (Kashyap, 2025) View paper
- Training Strategies and Data Efficiency
  - Efficient Fine-Tuning and Continual Pretraining (4 papers)
  - [5] E^ 2-LLM: Efficient and Extreme Length Extension of Large Language Models (Liu Jiaheng, 2024) View paper
  - [6] Breaking the stage barrier: A novel single-stage approach to long context extension for large language models (Chen Junmin, 2025) View paper
  - [27] Extending llms' context window with 100 samples (Zhang, 2024) View paper
  - [30] Extending the context window of a Generative Pre-trained Transformer Large Language Model with positional embeddings.: A comparative study in methods for â□¦ (Hartler, 2024) View paper
  - Synthetic Data Generation for Long Context (2 papers)
  - [32] Resonance RoPE: Improving Context Length Generalization of Large Language Models (Kobyzev, 2024) View paper
  - [49] Scaling Instruction-Tuned LLMs to Million-Token Contexts via Hierarchical Synthetic Data Generation (He Linda, 2025) View paper
  - Mixed Training and Attention Strategies (1 papers)
  - [33] Sliding Window Attention Training for Efficient Large Language Models (Song Wen-tao, 2025) View paper
- Analysis and Understanding of RoPE Extensions
  - Attention and Distribution Analysis (1 papers)
  - [10] Understanding the rope extensions of long-context llms: An attention perspective (Zhang, 2025) View paper
  - Numerical Precision and Implementation Issues (1 papers)
  - [28] When precision meets position: Bfloat16 breaks down rope in long-context training (Wang Haonan, 2024) View paper
- Domain-Specific and Multimodal Applications
  - Vision-Language Model Context Extension (3 papers)
  - [9] InternLM-XComposer-2.5: A Versatile Large Vision Language Model Supporting Long-Contextual Input and Output (Zhang Pan, 2024) View paper
  - [38] Design Choices for Extending the Context Length of Visual Language Models (Mukai Li, 2025) View paper
  - [45] GIRAFFE: Design Choices for Extending the Context Length of Visual Language Models (Li, 2024) View paper
  - Specialized Domain Applications (3 papers)
  - [3] UltraLLaDA: Scaling the Context Length to 128K for Diffusion Large Language Models (HE Guangxin, 2025) View paper
  - [13] Clinical ModernBERT: An efficient and long context encoder for biomedical text (Wu, 2025) View paper
  - [50] On Pretraining for Project-Level Code Completion (Maksim Sapronov, 2025) View paper
- Efficient Inference and Computational Optimization
  - Token Pruning and Hierarchical Processing (2 papers)
  - [19] InfiniteHiP: Extending Language Model Context Up to 3 Million Tokens on a Single GPU (Lee HeeJun, 2025) View paper
  - [36] Adaptive Computation Pruning for the Forgetting Transformer (Obando-Ceron, 2025) View paper
  - Embedding Model Context Extension (1 papers)
  - [20] Longembed: Extending embedding models for long context retrieval (Zhu Da-wei, 2024) View paper
- Extrapolation and Generalization Beyond Training Length
  - Continuous and Periodic Extrapolation (2 papers)
  - [40] CLEX: Continuous Length Extrapolation for Large Language Models (Chen, 2023) View paper
  - [42] PEPE: Long-context Extension for Large Language Models via Periodic Extrapolation Positional Encodings (Jikun Hu, 2025) View paper
  - Random-Access and Infinite Context Mechanisms (1 papers)
  - [4] Random-access infinite context length for transformers (A Mohtashami, 2023) View paper
  - Resonance and Collinear Constraint Methods (2 papers)
  - [11] Gpt rotational position embedding for length extrapolation (Zhijie Qu, 2023) View paper
  - [44] CoCA: Fusing Position Embedding with Collinear Constrained Attention in Transformers for Long Context Window Extending (Shiyi Zhu, 2024) View paper

Narrative

Core task: extending context window of rotary position embedding in language models. The field has organized itself around several complementary directions. A large cluster of work focuses on RoPE Modification and Rescaling Methods, exploring how to adjust frequency bases, interpolation factors, and dimension-specific scaling to enable longer contexts without full retraining. Training-Free and

Inference-Time Adaptation approaches seek lightweight solutions that avoid expensive fine-tuning, while Alternative and Hybrid Position Encoding Architectures investigate whether entirely different encoding schemes or combinations can outperform standard RoPE. Training Strategies and Data Efficiency examine how to minimize the computational cost of extending context, and Analysis and Understanding branches provide theoretical insights into why certain extensions succeed. Domain-Specific and Multimodal Applications adapt these techniques to specialized settings, Efficient Inference and Computational Optimization address runtime costs, and Extrapolation and Generalization Beyond Training Length tackle the challenge of generalizing far beyond the original training window.

Within RoPE Modification and Rescaling Methods, a particularly active line of work has emerged around unified theoretical frameworks that aim to explain and systematize the zoo of ad-hoc rescaling tricks. Early methods like Positional Interpolation[7] and Yarn[1] introduced interpolation and non-uniform scaling, while later efforts such as LongRope[2] and UltraLLaDA[3] refined these ideas with search-based or evolutionary strategies. MrRoPE[0] contributes to this unifying thread by proposing a principled framework that connects multiple rescaling approaches under a common theoretical lens, contrasting with more empirical or heuristic methods like Single Stage Extension[6] or E2-LLM[5]. A neighboring work, Distributional Perspective Extension[35], offers a complementary angle by analyzing RoPE extensions through the lens of attention score distributions. Together, these efforts reflect a maturing field where initial empirical successes are now being consolidated into more systematic and interpretable design principles.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. Extending Context Window of Large Language Models from a Distributional Perspective

Authors: WU Yingsheng, Gu Yuxuan, Yingsheng Wu, Feng, Xiaocheng, et al. (18 authors total) | **Year/Venue**: 2024 | **URL**: View paper

Abstract

Scaling the rotary position embedding (RoPE) has become a common method for extending the context window of RoPE-based large language models (LLMs). However, existing scaling methods often rely on empirical approaches and lack a profound understanding of the internal distribution within RoPE, resulting in suboptimal performance in extending the context window length. In this paper, we propose to optimize the context window extending task from the view of rotary angle distribution. Specifically, ...

Relationship Analysis

Both papers belong to the 'Unified Theoretical Frameworks for RoPE Extension' category, providing generalized formulations for extending RoPE-based context windows. While the original paper (MrRoPE) unifies existing methods through a radix system conversion perspective and proposes progressive/uniform radix conversion strategies, the candidate paper approaches the problem from a distributional perspective by analyzing and minimizing disturbances in rotary angle distributions. The key difference lies in their theoretical foundations: MrRoPE uses radix theory to design new extension strategies (MrRoPE-Pro/Uni), whereas the candidate paper optimizes extensions by preserving the statistical distribution of rotary angles from pre-training.

Contributions Analysis

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

Contribution 1: MrRoPE unified theoretical framework for RoPE-extension methods

Description: The authors introduce MrRoPE, a theoretical framework that unifies existing RoPE-extension methods (such as Position Interpolation, NTK-aware Interpolation, and YaRN) by interpreting them as different radix conversion strategies. This framework provides a systematic way to understand and compare various context extension approaches through the lens of mixed-radix positional encoding.

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. Found in the Middle: How Language Models Use Long Contexts Better via Plug-and-Play Positional Encoding

URL: View paper

Brief Assessment

Found in Middle[58] focuses on addressing the 'lost-in-the-middle' challenge through multi-scale positional encoding (MS-POE), not on unifying existing RoPE-extension methods through a theoretical framework.

---

2. Rethinking RoPE: A Mathematical Blueprint for N-dimensional Positional Encoding

URL: View paper

Brief Assessment

N-dimensional Blueprint[54] is not available for comparison as the candidate paper's full text context is marked as 'n/a'. Without access to the actual content, no assessment of novelty refutation can be made.

---

3. Rotary position embedding for vision transformer

URL: View paper

Brief Assessment

Rotary Vision Transformer[53] focuses on applying RoPE to vision transformers for 2D image data, not on unifying RoPE-extension methods through theoretical frameworks for context length scaling in language models.

---

4. Learning the RoPEs: Better 2D and 3D Position Encodings with STRING

URL: View paper

Brief Assessment

STRING[56] focuses on generalizing RoPE to multi-dimensional position encodings (2D/3D) for robotics applications, not on unifying RoPE-extension methods for context window scaling. The candidate addresses different dimensions of spatial coordinates, while the original addresses temporal sequence extension strategies.

---

5. Liere: Generalizing rotary position encodings

URL: View paper

Brief Assessment

Liere[55] focuses on generalizing RoPE to higher-dimensional inputs (2D/3D) across modalities, not on unifying existing RoPE-extension methods through a radix conversion framework.

---

6. Rethinking RoPE: A Mathematical Blueprint for N-dimensional Positional Embedding

URL: View paper

Brief Assessment

N-dimensional Blueprint[57] focuses on extending RoPE to higher-dimensional input domains (e.g., 2D images) using Lie group theory, not on unifying context extension methods like Position Interpolation or YaRN through radix conversion strategies.

---

7. 3d-rpe: Enhancing long-context modeling through 3d rotary position encoding

URL: View paper

Brief Assessment

3D-RPE[17] proposes a novel 3D rotary position encoding method inspired by the Bloch sphere, focusing on controllable long-term decay and improved position resolution. It does not present a unified theoretical framework for existing RoPE-extension methods like PI, NTK, and YaRN through radix conversion strategies.

---

8. Context-aware Rotary Position Embedding

URL: View paper

Brief Assessment

Context Aware RoPE[12] proposes a context-aware dynamic frequency generation mechanism for RoPE, not a theoretical framework for unifying existing RoPE-extension methods through radix conversion strategies.

---

9. Round and round we go! what makes rotary positional encodings useful?

URL: View paper

Brief Assessment

Round and Round[52] focuses on understanding how RoPE frequencies are used mechanistically within trained models (Gemma 7B), analyzing high vs. low frequency usage patterns and proposing p-RoPE. It does not present a unified theoretical framework for RoPE-extension methods or interpret them as radix conversion strategies.

---

10. Pose: Efficient context window extension of llms via positional skip-wise training

URL: View paper

Brief Assessment

Pose[51] focuses on efficient context window extension through positional skip-wise training that manipulates position indices during fine-tuning, rather than providing a theoretical framework that unifies different RoPE-extension methods through radix conversion strategies.

Contribution 2: MrRoPE-Pro training-free extension method with progressive radix conversion

Description: The authors propose two novel training-free RoPE extension methods: MrRoPE-Uni (using uniform radix conversion) and MrRoPE-Pro (using progressive radix conversion). These methods enable models to generalize to longer contexts than seen during pre-training without requiring additional fine-tuning, with MrRoPE-Pro demonstrating superior performance by progressively scaling the radix base across dimensions.

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. Resonance RoPE: Improving Context Length Generalization of Large Language Models

URL: View paper

Brief Assessment

Resonance RoPE[32] focuses on reducing interpolation gaps in pre-critical dimensions by rounding wavelengths to integers, while MrRoPE-Pro addresses the OOD problem in post-critical dimensions through progressive radix conversion. These are complementary approaches targeting different aspects of RoPE extension.

---

2. Effective Length Extrapolation via Dimension-Wise Positional Embeddings Manipulation

URL: View paper

Brief Assessment

Dimension Wise Manipulation[41] focuses on dimension-wise manipulation of position embeddings through detecting effective lengths and identifying key dimensions, rather than radix conversion strategies. The candidate's approach of selective dimension manipulation differs fundamentally from MrRoPE-Pro's progressive radix scaling methodology.

---

3. Q-ROAR: Outlier-Aware Rescaling for RoPE Position Interpolation in Quantized Long-Context LLMs

URL: View paper

Brief Assessment

Q-ROAR[43] addresses quantization-aware RoPE scaling for long contexts, not training-free RoPE extension methods. It focuses on stabilizing position interpolation under post-training quantization, which is orthogonal to MrRoPE-Pro's progressive radix conversion approach.

---

4. Optimal RoPE extension via Bayesian Optimization for training-free length generalization

URL: View paper

Brief Assessment

Optimal RoPE Bayesian[14] focuses on Bayesian optimization for RoPE extension parameters, not on radix conversion-based methods like MrRoPE-Pro's progressive strategy.

---

5. Extending the Context of Pretrained LLMs by Dropping Their Positional Embeddings

URL: View paper

Brief Assessment

Dropping Positional Embeddings[59] proposes removing positional embeddings after pretraining rather than modifying RoPE frequencies. This is a fundamentally different approach from MrRoPE-Pro's progressive radix conversion strategy for extending context windows.

---

6. Probing Rotary Position Embeddings through Frequency Entropy

URL: View paper

Brief Assessment

Frequency Entropy Probing[61] focuses on analyzing and measuring the utilization of RoPE frequency dimensions through entropy metrics, not on extending context windows. The candidate does not propose training-free extension methods or radix conversion strategies.

---

7. Breaking the stage barrier: A novel single-stage approach to long context extension for large language models

URL: View paper

Brief Assessment

Single Stage Extension[6] focuses on continual pretraining with head-adaptive RoPE bases during training, not training-free extension methods. The candidate requires training with 6B tokens, while the original proposes training-free inference-time methods.

---

8. Extending Audio Context for Long-Form Understanding in Large Audio-Language Models

URL: View paper

Brief Assessment

Audio Context Extension[60] focuses on audio-specific context extension in multimodal audio-language models, not general RoPE extension for text-based LLMs. The candidate's 'partial yarn' method modifies only audio token positions while preserving text positions, which is fundamentally different from MrRoPE-Pro's progressive radix conversion strategy applied uniformly across dimensions for text context extension.

---

9. LaMPE: Length-aware Multi-grained Positional Encoding for Adaptive Long-context Scaling Without Training

URL: View paper

Brief Assessment

LaMPE[37] focuses on dynamic mapping strategies using sigmoid functions and multi-grained attention mechanisms, rather than radix conversion approaches. The technical mechanisms differ fundamentally from MrRoPE-Pro's progressive radix conversion methodology.

---

10. Extending LLM Context Window with Adaptive Grouped Positional Encoding: A Training-Free Method

URL: View paper

Brief Assessment

Adaptive Grouped Encoding[46] focuses on progressively increasing reuse count of relative positions with dynamic adaptation to sequence length, which differs from MrRoPE-Pro's progressive radix conversion strategy across dimensions. The candidate does not provide sufficient technical detail to demonstrate prior work on progressive radix conversion.

Contribution 3: Theoretical analysis showing MrRoPE-Pro raises RoPE encoding length upper bound

Description: The authors provide theoretical evidence demonstrating that MrRoPE-Pro significantly extends the theoretical context window upper bound of RoPE-based models. Their analysis shows that MrRoPE-Pro stabilizes attention score distributions in intermediate dimensions and maximally restores high-frequency information, thereby increasing the effective context window limit.

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. Decoupling the "What" and "Where" With Polar Coordinate Positional Embeddings

URL: View paper

Brief Assessment

Polar Coordinate Embeddings[63] focuses on decoupling content and position information in RoPE to improve performance, not on theoretical upper bounds of encoding length. The candidate does not address or challenge the original paper's theoretical analysis of context window limits.

---

2. Beyond Position: the emergence of wavelet-like properties in Transformers

URL: View paper

Brief Assessment

Wavelet Properties[64] focuses on emergent wavelet-like properties in attention heads with RoPE, not on theoretical upper bounds of encoding length or context window extension methods.

---

3. ComRoPE: Scalable and Robust Rotary Position Embedding Parameterized by Trainable Commuting Angle Matrices

URL: View paper

Brief Assessment

ComRoPE[67] focuses on trainable commuting angle matrices for RoPE rather than context window extension methods. It does not address theoretical upper bounds of encoding length or mixed-radix conversion strategies.

---

4. Fast Gradient Computation for RoPE Attention in Almost Linear Time

URL: View paper

Brief Assessment

Fast Gradient RoPE[65] focuses on computational efficiency (almost linear time algorithms for forward/backward computations), not on extending the theoretical context window upper bound of RoPE encodings.

---

5. The impact of positional encoding on length generalization in transformers

URL: View paper

Brief Assessment

Positional Encoding Impact[62] focuses on comparing different positional encoding schemes (APE, T5's relative PE, ALiBi, Rotary, NoPE) for length generalization in transformers, not on extending RoPE's theoretical upper bound through mixed-radix conversion strategies like MrRoPE-Pro.

---

6. DoPE: Denoising Rotary Position Embedding

URL: View paper

Brief Assessment

DoPE[26] focuses on denoising positional encodings via truncated matrix entropy to address attention sinks and extrapolation, not on extending the theoretical upper bound of RoPE's encoding length through radix conversion strategies.

---

7. Head-wise Adaptive Rotary Positional Encoding for Fine-Grained Image Generation

URL: View paper

Brief Assessment

Head Wise Adaptive[68] focuses on head-wise adaptive transformations for fine-grained image generation in vision transformers, not on theoretical upper bounds of RoPE encoding length in language models. The candidate addresses spatial relations, color fidelity, and object counting in images, which is a fundamentally different domain and contribution from extending context windows in text-based transformers.

---

8. Eulerformer: Sequential user behavior modeling with complex vector attention

URL: View paper

Brief Assessment

Eulerformer[66] focuses on sequential user behavior modeling in recommendation systems using complex vector attention, not on extending RoPE's theoretical context window bounds for language models.

---

9. Liere: Generalizing rotary position encodings

URL: View paper

Brief Assessment

Liere[55] does not address theoretical upper bounds of RoPE encoding length or context window extension; it focuses on adapting RoPE to multi-dimensional spatial data.

---

10. Positional Encoding via Token-Aware Phase Attention

URL: View paper

Brief Assessment

Token Aware Phase[34] focuses on proving that RoPE introduces distance-dependent bias in attention scores and proposes a new positional encoding method (TAPA) to address this. It does not challenge the novelty of MrRoPE-Pro's theoretical analysis about raising the upper bound of RoPE's encoding length through progressive radix conversion strategies.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] MrRoPE: Mixed-radix Rotary Position Embedding View paper
- [1] Yarn: Efficient context window extension of large language models View paper
- [2] Longrope: Extending llm context window beyond 2 million tokens View paper
- [3] UltraLLaDA: Scaling the Context Length to 128K for Diffusion Large Language Models View paper
- [4] Random-access infinite context length for transformers View paper
- [5] E^ 2-LLM: Efficient and Extreme Length Extension of Large Language Models View paper
- [6] Breaking the stage barrier: A novel single-stage approach to long context extension for large language models View paper
- [7] Extending context window of large language models via positional interpolation View paper
- [8] HoPE: Hyperbolic Rotary Positional Encoding for Stable Long-Range Dependency Modeling in Large Language Models View paper
- [9] InternLM-XComposer-2.5: A Versatile Large Vision Language Model Supporting Long-Contextual Input and Output View paper
- [10] Understanding the rope extensions of long-context llms: An attention perspective View paper
- [11] Gpt rotational position embedding for length extrapolation View paper
- [12] Context-aware Rotary Position Embedding View paper
- [13] Clinical ModernBERT: An efficient and long context encoder for biomedical text View paper
- [14] Optimal RoPE extension via Bayesian Optimization for training-free length generalization View paper
- [15] SWAN-GPT: An Efficient and Scalable Approach for Long-Context Language Modeling View paper
- [16] Base of rope bounds context length View paper
- [17] 3d-rpe: Enhancing long-context modeling through 3d rotary position encoding View paper
- [18] Scaling laws of rope-based extrapolation View paper
- [19] InfiniteHiP: Extending Language Model Context Up to 3 Million Tokens on a Single GPU View paper
- [20] Longembed: Extending embedding models for long context retrieval View paper
- [21] SeqPE: Transformer with Sequential Position Encoding View paper
- [22] LongRoPE2: Near-Lossless LLM Context Window Scaling View paper
- [23] Extending context window in large language models with segmented base adjustment for rotary position embeddings View paper
- [24] Beyond Real: Imaginary Extension of Rotary Position Embeddings for Long-Context LLMs View paper
- [25] Rope to nope and back again: A new hybrid attention strategy View paper
- [26] DoPE: Denoising Rotary Position Embedding View paper
- [27] Extending llms' context window with 100 samples View paper
- [28] When precision meets position: Bfloat16 breaks down rope in long-context training View paper
- [29] An efficient recipe for long context extension via middle-focused positional encoding View paper
- [30] Extending the context window of a Generative Pre-trained Transformer Large Language Model with positional embeddings.: A comparative study in methods for â View paper
- [31] On the token distance modeling ability of higher rope attention dimension View paper
- [32] Resonance RoPE: Improving Context Length Generalization of Large Language Models View paper
- [33] Sliding Window Attention Training for Efficient Large Language Models View paper
- [34] Positional Encoding via Token-Aware Phase Attention View paper
- [35] Extending Context Window of Large Language Models from a Distributional Perspective View paper
- [36] Adaptive Computation Pruning for the Forgetting Transformer View paper
- [37] LaMPE: Length-aware Multi-grained Positional Encoding for Adaptive Long-context Scaling Without Training View paper
- [38] Design Choices for Extending the Context Length of Visual Language Models View paper
- [39] Recurrent Memory-Augmented Transformers with Chunked Attention for Long-Context Language Modeling View paper
- [40] CLEX: Continuous Length Extrapolation for Large Language Models View paper

- [41] Effective Length Extrapolation via Dimension-Wise Positional Embeddings Manipulation View paper
- [42] PEPE: Long-context Extension for Large Language Models via Periodic Extrapolation Positional Encodings View paper
- [43] Q-ROAR: Outlier-Aware Rescaling for RoPE Position Interpolation in Quantized Long-Context LLMs View paper
- [44] CoCA: Fusing Position Embedding with Collinear Constrained Attention in Transformers for Long Context Window Extending View paper
- [45] GIRAFFE: Design Choices for Extending the Context Length of Visual Language Models View paper
- [46] Extending LLM Context Window with Adaptive Grouped Positional Encoding: A Training-Free Method View paper
- [47] Rethinking RoPE Scaling in Quantized LLM: Theory, Outlier, and Channel-Band Analysis with Weight Rescaling View paper
- [48] CoWPE: Adaptive Context Window Adjustment in LLMs for Complex Input Queries View paper
- [49] Scaling Instruction-Tuned LLMs to Million-Token Contexts via Hierarchical Synthetic Data Generation View paper
- [50] On Pretraining for Project-Level Code Completion View paper
- [51] Pose: Efficient context window extension of llms via positional skip-wise training View paper
- [52] Round and round we go! what makes rotary positional encodings useful? View paper
- [53] Rotary position embedding for vision transformer View paper
- [54] Rethinking RoPE: A Mathematical Blueprint for N-dimensional Positional Encoding View paper
- [55] Liere: Generalizing rotary position encodings View paper
- [56] Learning the RoPEs: Better 2D and 3D Position Encodings with STRING View paper
- [57] Rethinking RoPE: A Mathematical Blueprint for N-dimensional Positional Embedding View paper
- [58] Found in the Middle: How Language Models Use Long Contexts Better via Plug-and-Play Positional Encoding View paper
- [59] Extending the Context of Pretrained LLMs by Dropping Their Positional Embeddings View paper
- [60] Extending Audio Context for Long-Form Understanding in Large Audio-Language Models View paper
- [61] Probing Rotary Position Embeddings through Frequency Entropy View paper
- [62] The impact of positional encoding on length generalization in transformers View paper
- [63] Decoupling the "What" and "Where" With Polar Coordinate Positional Embeddings View paper
- [64] Beyond Position: the emergence of wavelet-like properties in Transformers View paper
- [65] Fast Gradient Computation for RoPE Attention in Almost Linear Time View paper
- [66] Eulerformer: Sequential user behavior modeling with complex vector attention View paper
- [67] ComRoPE: Scalable and Robust Rotary Position Embedding Parameterized by Trainable Commuting Angle Matrices View paper
- [68] Head-wise Adaptive Rotary Positional Encoding for Fine-Grained Image Generation View paper