Novelty Assessment Report

Paper: SeeDNorm: Self-Rescaled Dynamic Normalization

PDF URL: https://openreview.net/pdf?id=dFL7XExDvv

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-04

Abstract

Normalization layer constitutes an essential component in neural networks. In transformers, the predominantly used RMSNorm constrains vectors to a unit hypersphere, followed by dimension-wise rescaling through a learnable scaling coefficient γ to maintain the representational capacity of the model. However, RMSNorm discards the input norm information in forward pass and a static scaling factor γ may be insufficient to accommodate the wide variability of input data and distributional shifts, thereby limiting further performance improvements, particularly in zero-shot scenarios that large language models routinely encounter. To address this limitation, we propose SeeDNorm, which enhances the representational capability of the model by dynamically adjusting the scaling coefficient based on the current input, thereby preserving the input norm information and enabling data-dependent, self-rescaled dynamic normalization. During backpropagation, SeeDNorm retains the ability of RMSNorm to dynamically adjust gradient according to the input norm. We provide a detailed analysis of the training optimization for SeedNorm and proposed corresponding solutions to address potential instability issues that may arise when applying SeeDNorm. We validate the effectiveness of SeeDNorm across models of varying sizes in large language model pre-training as well as supervised and unsupervised computer vision tasks. By introducing a minimal number of parameters and with negligible impact on model efficiency, SeeDNorm achieves consistently superior performance compared to previously commonly used normalization layers such as RMSNorm and LayerNorm, as well as element-wise activation alternatives to normalization layers like DyT.

Core Task Landscape

This paper addresses: **Dynamic Normalization for Neural Networks**

A total of **50 papers** were analyzed and organized into a taxonomy with **26 categories**.

Taxonomy Overview

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

- **Core Normalization Mechanisms and Architectures**
- **Domain Adaptation and Transfer Learning**
- **Time Series and Temporal Data Processing**
- **Input Preprocessing and Data Enhancement**
- **Specialized Architecture Applications**
- **Multi-Task and Multi-Objective Learning**
- **Domain-Specific Applications**

Complete Taxonomy Tree

- Dynamic Normalization for Neural Networks Survey Taxonomy
- Core Normalization Mechanisms and Architectures
  - Batch-Free and Online Normalization (2 papers)
  - [1] Online normalization for training neural networks (Vitaliy Chiley, 2019) View paper
  - [4] Is normalization indispensable for training deep neural network? (Jie Shao, 2020) View paper
  - Dynamic and Adaptive Scaling Mechanisms ★ (4 papers)
  - [0] SeeDNorm: Self-Rescaled Dynamic Normalization (Anon et al., 2026) View paper
  - [15] Dynamic normalization (Liu Chuan, 2020) View paper
  - [35] Differentiable dynamic normalization for learning deep representation (Ping Luo, 2019) View paper
  - [41] Enhancing deep neural network training through learnable adaptive normalization (Jan Benedikt Ruhland, 2025) View paper
  - Switchable and Multi-Scope Normalization (2 papers)
  - [30] Switchable normalization for learning-to-normalize deep representation (Ping Luo, 2019) View paper
  - [36] Exploring the Efficacy of Group-Normalization in Deep Learning Models for Alzheimer's Disease Classification (Habib, 2024) View paper
  - Normalization-Activation Integration (3 papers)
  - [12] Transformers without normalization (Zhu Jiachen, 2025) View paper
  - [25] The Mathematical Relationship Between Layer Normalization and Dynamic Activation Functions (Stollenwerk, 2025) View paper
  - [29] Evolving normalization-activation layers (Hanxiao Liu, 2020) View paper
  - Theoretical Analysis and Mathematical Foundations (2 papers)
  - [9] On Centralization and Unitization of Batch Normalization for Deep ReLU Neural Networks (Wen Fei, 2024) View paper
  - [40] A note on factor normalization for deep neural network models. (Haobo Qi, 2022) View paper
  - Alternative Normalization Formulations (3 papers)
  - [26] Cluster-Based Activation Normalization for Neural Networks (Bilal Faye, 2024) View paper

  - [33] Layer normalization (Ba, 2016) View paper
  - [50] Batch Layer Normalization A new normalization layer for CNNs and RNNs (Amir Ziaee, 2022) View paper
- Domain Adaptation and Transfer Learning
  - Unsupervised and Test-Time Domain Adaptation (3 papers)
  - [8] Adaptive batch normalization for practical domain adaptation (Yanghao Li, 2018) View paper
  - [14] Dyno: Dynamic normalization based test-time adaptation for 2d medical image segmentation (Yihang Fu, 2024) View paper
  - [22] The norm must go on: Dynamic unsupervised domain adaptation by normalization (Mirza, 2022) View paper
  - Single Domain Generalization (1 papers)
  - [6] Adversarially adaptive normalization for single domain generalization (Xinjie Fan, 2021) View paper
  - Transferable Normalization Architectures (1 papers)
  - [23] Transferable normalization: Towards improving transferability of deep neural networks (Ximei Wang, 2019) View paper
- Time Series and Temporal Data Processing
  - Non-Stationary Time Series Forecasting (2 papers)
  - [3] DDN: Dual-domain dynamic normalization for non-stationary time series forecasting (Tao Dai, 2024) View paper
  - [5] Adaptive normalization for non-stationary time series forecasting: A temporal slice perspective (Z Liu, 2023) View paper
  - Temporal Attention and Dynamic Visual Processing (2 papers)
  - [2] A dynamic normalization model of temporal attention (Rachel N. Denison, 2021) View paper
  - [10] A dynamic spatiotemporal normalization model captures perceptual and neural effects of spatial and temporal context (Angus F Chapman, 2025) View paper
  - Spiking and Recurrent Neural Networks (1 papers)
  - [13] Temporal effective batch normalization in spiking neural networks (C Duan, 2022) View paper
- Input Preprocessing and Data Enhancement
  - Adaptive Input Normalization (3 papers)
  - [18] Extended deep adaptive input normalization for preprocessing time series data for neural networks (Passino, 2024) View paper
  - [27] Adaptive Input Normalization for Quantized Neural Networks (Jan Schmidt, 2024) View paper
  - [45] Global adaptive input normalization for short-term electric load forecasting (Nikolaos Passalis, 2020) View paper
  - Generative Data Augmentation (2 papers)
  - [7] Adaptive normalization data enhancement algorithm based on DCGAN (Qingxuan Wang, 2024) View paper
  - [37] Sketch-guided spatial adaptive normalization and high-level feature constraints based GAN image synthesis for steel strip defect detection data augmentation (Guangjun Ran, 2024) View paper
- Specialized Architecture Applications
  - Vision Transformers and Token-Based Models (1 papers)
  - [21] Dynamic token normalization improves vision transformers (Shao, 2021) View paper
  - Graph Neural Networks (1 papers)
  - [24] GRANOLA: Adaptive Normalization for Graph Neural Networks (Beatrice Bevilacqua, 2024) View paper
  - Video and Spatiotemporal Convolutional Networks (1 papers)
  - [19] Dynamic normalization and relay for video action recognition (Dongqi Cai, 2021) View paper
- Multi-Task and Multi-Objective Learning (1 papers)
  - [16] Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks (Chen Zhao, 2018) View paper
- Domain-Specific Applications
  - Medical Image Analysis (2 papers)
  - [39] AMIN-CNN: Enhancing Brain Tumor Segmentation through Modality-Aware Normalization and Deep Learning (Sivakumar Depuru, 2025) View paper
  - [44] Adaptive Normalization Enhances the Generalization of Deep Learning Model in Chest X-Ray Classification (Jatsada Singthongchai, 2025) View paper
  - Hyperspectral and Remote Sensing Imaging (2 papers)
  - [11] Dynamic super-pixel normalization for robust hyperspectral image classification (Wang Cong, 2023) View paper
  - [48] Underwater image enhancement using adaptive standardization and normalization networks (Cheol Woo Park, 2024) View paper
  - Industrial Fault Diagnosis and Predictive Maintenance (5 papers)
  - [17] Dynamic normalization supervised contrastive network with multiscale compound attention mechanism for gearbox imbalanced fault diagnosis (Yutong Dong, 2024) View paper
  - [38] Developing Distance-Aware, and Evident Uncertainty Quantification in Dynamic Physics-Constrained Neural Networks for Robust Bearing Degradation Estimation (Waleed Razzaq, 2025) View paper
  - [42] Time-varying Gaussian encoder-based adaptive sensor-weighted method for turbofan engine remaining useful life prediction (Lei Ren, 2023) View paper
  - [43] Speed-Normalized Expansion-and-Shrinkage network for health indicator generating under time-varying conditions (Wenyi Liu, 2025) View paper
  - [49] Gas Turbine Anomaly Detection under Time-Varying Operation Conditions Based on Spectra Alignment and Self-Adaptive Normalization (Dongyan Miao, 2024) View paper
  - Speech and Audio Processing (1 papers)
  - [34] Dynamic layer normalization for adaptive neural acoustic modeling in speech recognition (Taesup Kim, 2017) View paper
  - Gesture and Sign Language Recognition (2 papers)
  - [28] Real-time spatial normalization for dynamic gesture classification (Sofiane Zeghoud, 2022) View paper
  - [46] Gesture Recognition of Filipino Sign Language Using Convolutional and Long Short-Term Memory Deep Neural Networks (Karl Jensen F. Cayme, 2024) View paper
  - Financial and Stock Market Prediction (1 papers)
  - [31] Dynamic normalization BPN for stock price forecasting (Chiaâ□□Chi Chen, 2015) View paper
  - Physics-Informed and Scientific Computing (2 papers)
  - [20] Dynamic & norm-based weights to normalize imbalance in back-propagated gradients of physics-informed neural networks (Shota Deguchi, 2023) View paper
  - [32] Time-varying trajectory modeling via dynamic governing network for remaining useful life prediction (Zheng Zhou, 2023) View paper

- Autonomous Driving and Pedestrian Detection (1 papers)
- [47] Pedestrian intention prediction in Adverse Weather Conditions with Spiking Neural Networks and Dynamic Vision Sensors (Sakhai, 2024) View paper

Narrative

Core task: dynamic normalization for neural networks. The field encompasses a diverse set of strategies for adaptively adjusting feature statistics during training and inference, aiming to improve convergence, generalization, and robustness across varying data distributions. The taxonomy organizes these approaches into several main branches: Core Normalization Mechanisms and Architectures focuses on foundational techniques such as adaptive scaling, learnable parameters, and novel normalization layers (e.g., Layer Normalization[33], Switchable Normalization[30]); Domain Adaptation and Transfer Learning addresses methods that adjust normalization to handle distribution shifts (e.g., Transferable Normalization[23]); Time Series and Temporal Data Processing targets sequential and spatiotemporal contexts (e.g., Dynamic Spatiotemporal[10], Temporal Effective Batch[13]); Input Preprocessing and Data Enhancement explores normalization at the data level; Specialized Architecture Applications tailors normalization to specific network designs; Multi-Task and Multi-Objective Learning examines normalization in settings with multiple objectives (e.g., GradNorm[16]); and Domain-Specific Applications applies these ideas to fields like medical imaging, gesture recognition, and anomaly detection.

Within the Core Normalization Mechanisms branch, a particularly active line of work centers on dynamic and adaptive scaling mechanisms that learn or compute normalization parameters on-the-fly rather than relying on fixed statistics. SeeDNorm[0] exemplifies this direction by introducing a mechanism that dynamically adjusts normalization based on input characteristics, positioning it alongside works like Dynamic Normalization[15] and Differentiable Dynamic[35], which similarly emphasize learnable or context-sensitive scaling. In contrast, Learnable Adaptive[41] explores parameterized normalization strategies that balance flexibility with computational efficiency, while Dual Domain Dynamic[3] extends adaptive normalization to handle multiple feature domains simultaneously. A recurring theme across these studies is the trade-off between expressiveness and stability: highly adaptive schemes can better capture input variability but may introduce training instability or overfitting, whereas more constrained approaches sacrifice some flexibility for robustness. SeeDNorm[0] navigates this landscape by proposing a scaling mechanism that remains computationally tractable while offering richer adaptivity than earlier fixed-parameter methods, situating it as a middle ground between fully static normalization layers and more complex, domain-specific adaptive schemes.

Related Works in Same Category

The following **3 sibling papers** share the same taxonomy leaf node with the original paper:

1. Dynamic normalization

Authors: Liu Chuan, Chuan Liu, Gao Yi, Yi Gao, Lv, et al. (7 authors total) | **Year/Venue**: 2020 | **URL**: View paper

Abstract

Batch Normalization has become one of the essential components in CNN. It allows the network to use a higher learning rate and speed up training. And the network doesn't need to be initialized carefully. However, in our work, we find that a simple extension of BN can increase the performance of the network. First, we extend BN to adaptively generate scale and shift parameters for each mini-batch data, called DN-C (Batch-shared and Channel-wise). We use the statistical characteristics of mini-bat...

Relationship Analysis

Both papers belong to the Dynamic and Adaptive Scaling Mechanisms category, focusing on normalization layers with learnable or input-dependent parameters. The candidate paper (Dynamic Normalization) extends Batch Normalization by adaptively generating scale and shift parameters using mini-batch statistics (DN-C) or per-sample statistics (DN-B), primarily for CNNs in computer vision tasks. In contrast, the original paper (SeeDNorm) extends RMSNorm by dynamically adjusting scaling coefficients based on current input while preserving input norm information, targeting Transformer architectures in both language modeling and vision tasks with a focus on zero-shot generalization.

---

2. Differentiable dynamic normalization for learning deep representation

Authors: Ping Luo, Zhanglin Peng, Wenqi Shao, Ruimao Zhang, Jiamin Ren, et al. (6 authors total) | **Year/Venue**: 2019 | **URL**: View paper

Abstract

â¦ This work presents Dynamic Normalization (DN), which is able to learn arbitrary normalization â¦ approaches are indispensable components in recent deep neural networks (DNNs), such â¦

Relationship Analysis

Both papers belong to the Dynamic and Adaptive Scaling Mechanisms category, focusing on normalization layers with learnable parameters that adapt during forward pass. They overlap in addressing dynamic adjustment of normalization statistics, but differ fundamentally in their approach: SeeDNorm dynamically adjusts scaling coefficients based on input norm information through a self-rescaling mechanism ($\sigma(x \cdot \beta^T) \cdot \alpha + \gamma$), while Dynamic Normalization learns arbitrary normalization operations by transforming statistics through learnable binary matrices (U and V) that partition batch and channel dimensions into groups. SeeDNorm preserves input norm information explicitly, whereas DN focuses on learning diverse grouping strategies for computing statistics.

---

3. Enhancing deep neural network training through learnable adaptive normalization

Authors: Jan Benedikt Ruhland, Iraj Masoudian, Benedikt Ruhland, Dominik Heider | **Year/Venue**: 2025 | **URL**: View paper

Abstract

â¦ adaptive normalization â¦ deep learning architectures, including feed-forward, convolutional, and transformer-based neural networks. The results demonstrate that adaptive normalization â¦

Relationship Analysis

Both papers belong to the Dynamic and Adaptive Scaling Mechanisms category, focusing on normalization layers with learnable or input-dependent parameters that adapt during forward pass. The candidate paper proposes a learnable adaptive normalization approach applicable across feed-forward, convolutional, and transformer architectures, overlapping with SeeDNorm's goal of dynamic normalization. However, SeeDNorm specifically addresses input norm preservation through self-rescaled dynamic adjustment based on RMSNorm, incorporating input-dependent scaling via tanh-activated projections ($\sigma(x \cdot \beta^T) \cdot \alpha + \gamma$), while the candidate paper appears to focus on a broader learnable adaptive framework without the specific self-rescaling mechanism or detailed gradient analysis that characterizes SeeDNorm.

Contributions Analysis

Overall novelty summary. The paper proposes SeeDNorm, a dynamic normalization layer that adjusts scaling coefficients based on current input rather than using static parameters. It resides in the 'Dynamic and Adaptive Scaling Mechanisms' leaf of the taxonomy, which contains four papers including the original work. This leaf sits within the broader 'Core Normalization Mechanisms and Architectures' branch, indicating a moderately populated research direction focused on learnable or input-dependent normalization parameters. The taxonomy shows this is an active but not overcrowded area, with sibling papers exploring related adaptive scaling strategies.

The taxonomy reveals several neighboring research directions that contextualize SeeDNorm's contribution. Adjacent leaves include 'Batch-Free and Online Normalization' (2 papers), 'Switchable and Multi-Scope Normalization' (2 papers), and 'Normalization-Activation Integration' (3 papers), suggesting the field explores diverse approaches to adaptive normalization beyond dynamic scaling. The 'Domain Adaptation and Transfer Learning' branch (5 papers across 3 leaves) addresses distribution shifts through different mechanisms, while 'Time Series and Temporal Data Processing' (5 papers) tackles temporal dynamics. SeeDNorm's focus on input-dependent scaling distinguishes it from these parallel directions, which emphasize scope selection, domain transfer, or temporal adaptation rather than dynamic coefficient adjustment.

Among 30 candidates examined through semantic search and citation expansion, none clearly refute any of the three contributions: the SeeDNorm mechanism itself (10 candidates examined, 0 refutable), the theoretical stability analysis (10 candidates, 0 refutable), and empirical validation across language and vision tasks (10 candidates, 0 refutable). This limited search scope suggests that within the examined literature, the specific combination of input-dependent scaling with norm preservation appears relatively unexplored. However, the analysis explicitly notes this is not an exhaustive search, and the sibling papers in the same taxonomy leaf indicate related adaptive scaling work exists in the broader field.

Based on the limited search of 30 semantically similar papers, SeeDNorm appears to occupy a distinct position within dynamic normalization research. The taxonomy structure shows it contributes to an active but not saturated research direction, with clear boundaries separating it from domain adaptation, temporal processing, and architecture-specific methods. The absence of refuting candidates among examined papers suggests novelty within the search scope, though the analysis acknowledges this does not constitute comprehensive coverage of all prior work in adaptive normalization mechanisms.

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

Contribution 1: SeeDNorm: Self-Rescaled Dynamic Normalization

Description: The authors introduce SeeDNorm, a novel normalization layer that dynamically adjusts its scaling coefficient conditioned on the input. Unlike RMSNorm, which uses a static scaling factor, SeeDNorm preserves input norm information in the forward pass while maintaining the ability to adaptively adjust gradients during backpropagation.

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. GammaGAN: Gamma-Scaled Class Embeddings for Conditional Video Generation

URL: View paper

Brief Assessment

GammaGAN[70] focuses on conditional video generation using scaled class embeddings and output normalization in a two-stream architecture. This is fundamentally different from SeeDNorm's input-dependent dynamic normalization for general neural networks.

---

2. Enhancing cross-domain generalization in retinal image segmentation via style randomization and style normalization

URL: View paper

Brief Assessment

Style Randomization Retinal[71] focuses on domain adaptation for retinal image segmentation through style randomization and normalization techniques, not on general-purpose dynamic normalization layers with input-dependent scaling for neural networks.

---

3. Gradient-Weighted, Data-Driven Normalization for Approximate Border Bases -- Concept and Computation

URL: View paper

Brief Assessment

Gradient Weighted Border[73] focuses on gradient-weighted normalization for approximate border bases in computational commutative algebra, not neural network normalization layers. The candidate addresses polynomial vanishing ideals and data-driven normalization in algebraic geometry contexts, which is fundamentally different from SeeDNorm's dynamic scaling in neural network architectures.

---

4. SC-GAN: A Style-Conditioned Generative Adversarial Network for High-Quality Artistic Image Generation

URL: View paper

Brief Assessment

SC GAN[76] focuses on style-conditioned image generation using adaptive instance normalization for GANs, not on dynamic normalization layers with input-dependent scaling coefficients for general neural network architectures.

---

5. Global partitioning elevation normalization applied to building footprint prediction

URL: View paper

Brief Assessment

Global Partitioning Elevation[72] addresses elevation data normalization for topographical CNN applications, not general neural network normalization with input-dependent scaling coefficients that preserve norm information during forward and backward passes.

---

6. StatAvg: Mitigating Data Heterogeneity in Federated Learning for Intrusion Detection Systems

URL: View paper

Brief Assessment

StatAvg[69] addresses data normalization in federated learning for intrusion detection systems by sharing statistical information (mean and variance) across clients for universal feature scaling. This is fundamentally different from SeeDNorm, which proposes a novel neural network normalization layer with input-dependent dynamic scaling coefficients that preserve norm information during forward and backward passes in deep learning models.

---

7. NormSoftmax: Normalizing the Input of Softmax to Accelerate and Stabilize Training

URL: View paper

Brief Assessment

NormSoftmax[78] focuses on normalizing inputs to softmax functions (in cross-entropy loss and attention), not on general normalization layers with dynamic, input-dependent scaling coefficients that preserve norm information throughout network layers.

---

8. Medium-scale projection of reference evapotranspiration beyond available data using sequential deep learning models: a case study from Bangladesh

URL: View paper

Brief Assessment

Sequential Deep Learning[74] focuses on evapotranspiration prediction using sequential models with standard normalization techniques. It does not address dynamic normalization layers with input-dependent scaling coefficients.

9. Enhanced Model Robustness to Input Corruptions by Per-corruption Adaptation of Normalization Statistics

URL: View paper

Brief Assessment

Per Corruption Adaptation[75] focuses on adapting normalization statistics based on corruption type identification for robustness to input corruptions in robotic vision, not on dynamic input-dependent scaling coefficients that preserve norm information during forward/backward passes as in SeeDNorm.

10. MINTIN: Maxout-Based and Input-Normalized Transformation Invariant Neural Network

URL: View paper

Brief Assessment

MINTIN[77] focuses on transformation invariance (rotation/scaling) in CNNs through maxout operators and input normalization modules, not on dynamic normalization with input-dependent scaling coefficients that preserve norm information as in SeeDNorm.

Contribution 2: Theoretical analysis and stability solutions for SeeDNorm

Description: The authors conduct a comprehensive theoretical analysis of SeeDNorm's forward and backward propagation properties, including scale invariance and gradient behavior. They propose techniques such as multi-head SeeDNorm and weight decay strategies to enhance training stability.

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. Continuous-Time Analysis of Adaptive Optimization and Normalization

URL: View paper

Brief Assessment

Continuous Time Adaptive[67] focuses on continuous-time analysis of Adam optimizer and normalization layers' implicit effects, not on dynamic normalization layers like SeeDNorm. The candidate analyzes scale-invariant architectural components (e.g., layer-norm) theoretically, while the original proposes a novel dynamic normalization method with input-dependent scaling coefficients.

2. BNPO: Beta Normalization Policy Optimization

URL: View paper

Brief Assessment

BNPO[61] focuses on policy optimization for reinforcement learning with binary-valued rewards using beta distribution normalization, not on dynamic normalization layers in neural networks like SeeDNorm.

3. An Adaptive Mixed-Step Size Normalized Least Means Fourth Control Approach for Stand-Alone Power Generation System Considering Dynamic Conditions

URL: View paper

Brief Assessment

Mixed Step Normalized[68] addresses stability in power generation systems using adaptive control for voltage converters, not neural network normalization layers. The technical domains are entirely different—power systems engineering versus deep learning optimization.

4. Pangu Light: Weight Re-Initialization for Pruning and Accelerating LLMs

URL: View paper

Brief Assessment

Pangu Light[63] focuses on structured pruning and weight re-initialization for LLM acceleration, not on normalization layer design or training stability analysis. The candidate addresses a completely different technical problem domain.

5. Fixup initialization: Residual learning without normalization

URL: View paper

Brief Assessment

Fixup Initialization[60] addresses training stability through initialization rescaling for residual networks without normalization, not through dynamic normalization layers. The candidate focuses on gradient explosion problems at initialization via proper weight scaling, while the original paper analyzes forward/backward properties of a dynamic normalization method (SeeDNorm) that adjusts scaling coefficients based on input.

6. DyTTP: Trajectory Prediction with Normalization-Free Transformers

URL: View paper

Brief Assessment

DyTTP[64] applies DynamicTanh (DyT) to trajectory prediction tasks but does not provide theoretical analysis of training stability or propose stability solutions for dynamic normalization layers. The candidate focuses on application rather than theoretical contributions.

7. Quality prediction of industrial process based on Kolmogorovâ Arnold graph convolution aggregation temporal convolution network

URL: View paper

Brief Assessment

Kolmogorov Arnold Graph[62] focuses on industrial process quality prediction using graph convolution and temporal networks with KAN for nonlinear feature extraction. It does not address theoretical analysis of dynamic normalization layers or their training stability properties.

8. Enhancing deep neural network training through learnable adaptive normalization

URL: View paper

Brief Assessment

Learnable Adaptive[41] focuses on adaptive normalization techniques in deep neural networks but does not provide sufficient detail in the available context to assess whether it addresses the specific theoretical analysis of forward/backward propagation properties, scale invariance, gradient behavior, or stability solutions (multi-head variants, weight decay strategies) that are central to the ORIGINAL paper's SeeDNorm contribution.

9. StyDiff: a refined style transfer method based on diffusion models

URL: View paper

Brief Assessment

StyDiff[65] focuses on image style transfer using diffusion models and AdaIN, not on normalization layer training stability or dynamic normalization techniques relevant to SeeDNorm.

10. Stream normalization for ctr prediction

URL: View paper

Brief Assessment

Stream CTR[66] focuses on dynamic normalization for streaming CTR prediction tasks with non-stationary data distributions, while the original paper addresses training stability in general neural networks with static datasets. The technical approaches differ fundamentally.

Contribution 3: Empirical validation across language and vision tasks

Description: The authors demonstrate SeeDNorm's effectiveness through extensive experiments on large language models (both dense and MoE architectures) and computer vision tasks including image generation, supervised classification, and self-supervised learning. SeeDNorm achieves superior performance compared to RMSNorm, LayerNorm, and DyT with minimal parameter overhead.

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. Transformers without normalization

URL: View paper

Brief Assessment

[Final Audit Failure] The model insisted on a refutation claim but failed to provide verifiable evidence after multiple retries. Marked as cannot_refute for safety. Please manually verify the candidate text.

2. Influence of feature normalization methods on transfer Learning-A comparison study

URL: View paper

Brief Assessment

Transfer Learning Comparison[56] focuses on comparing feature normalization methods (min-max, Pareto scaling) for transfer learning scenarios, not on proposing or evaluating a novel normalization layer like SeeDNorm across diverse language and vision tasks.

3. On feature normalization and data augmentation

URL: View paper

Brief Assessment

Feature Normalization Augmentation[55] focuses on data augmentation through moment exchange in normalization layers, not on developing a new normalization method like SeeDNorm. The candidate evaluates augmentation techniques across vision and speech tasks, while the original paper proposes a novel dynamic normalization layer with different technical mechanisms and evaluation focus.

4. Different data cleaning techniques and normalization techniques with focus on current normalization techniques: A study

URL: View paper

Brief Assessment

Data Cleaning Techniques[53] focuses on data preprocessing and normalization techniques for data cleaning, not on evaluating normalization layers within neural network architectures across language and vision tasks.

5. High-performance large-scale image recognition without normalization

URL: View paper

Brief Assessment

High Performance Without[58] focuses exclusively on image recognition tasks (ImageNet classification, COCO detection) without normalization layers. It does not address language modeling tasks or the specific normalization methods (RMSNorm, LayerNorm, DyT) that SeeDNorm compares against.

6. Comparing image normalization techniques in an end-to-end model for automated modic changes classification from MRI images

URL: View paper

Brief Assessment

Image Normalization Modic[54] focuses on comparing image normalization techniques for medical imaging (MRI-based Modic changes classification), not on normalization methods for language models or general vision tasks like image generation and self-supervised learning.

7. Gated channel transformation for visual recognition

URL: View paper

Brief Assessment

Gated Channel Transformation[52] focuses on visual recognition tasks (image classification, object detection, video classification) and does not address language models or the specific normalization methods (RMSNorm, LayerNorm, SeeDNorm) discussed in the original paper.

8. A review of convolutional neural networks in computer vision

URL: View paper

Brief Assessment

CNN Review Vision[51] is a review paper surveying convolutional neural networks in computer vision tasks (image classification, object detection, video prediction). It does not propose a new normalization method like SeeDNorm, nor does it conduct empirical validation of a novel technique across language and vision domains.

9. Test-time distribution normalization for contrastively learned visual-language models

URL: View paper

Brief Assessment

Test Time Distribution[59] focuses on test-time distribution normalization for vision-language contrastive models (CLIP), not on normalization layers for language model pre-training and supervised/self-supervised vision tasks as in the original paper.

10. Optimal affine image normalization approach for optical character recognition

URL: View paper

Brief Assessment

Affine Image OCR[57] focuses exclusively on geometric image normalization for optical character recognition tasks, not on normalization layers for neural networks across language and vision domains.

Appendix: Text Similarity Detection

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

References

- [0] SeeDNorm: Self-Rescaled Dynamic Normalization View paper
- [1] Online normalization for training neural networks View paper
- [2] A dynamic normalization model of temporal attention View paper
- [3] DDN: Dual-domain dynamic normalization for non-stationary time series forecasting View paper
- [4] Is normalization indispensable for training deep neural network? View paper
- [5] Adaptive normalization for non-stationary time series forecasting: A temporal slice perspective View paper
- [6] Adversarially adaptive normalization for single domain generalization View paper
- [7] Adaptive normalization data enhancement algorithm based on DCGAN View paper
- [8] Adaptive batch normalization for practical domain adaptation View paper
- [9] On Centralization and Unitization of Batch Normalization for Deep ReLU Neural Networks View paper
- [10] A dynamic spatiotemporal normalization model captures perceptual and neural effects of spatial and temporal context View paper
- [11] Dynamic super-pixel normalization for robust hyperspectral image classification View paper
- [12] Transformers without normalization View paper
- [13] Temporal effective batch normalization in spiking neural networks View paper
- [14] Dyno: Dynamic normalization based test-time adaptation for 2d medical image segmentation View paper
- [15] Dynamic normalization View paper
- [16] Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks View paper
- [17] Dynamic normalization supervised contrastive network with multiscale compound attention mechanism for gearbox imbalanced fault diagnosis View paper
- [18] Extended deep adaptive input normalization for preprocessing time series data for neural networks View paper
- [19] Dynamic normalization and relay for video action recognition View paper
- [20] Dynamic & norm-based weights to normalize imbalance in back-propagated gradients of physics-informed neural networks View paper
- [21] Dynamic token normalization improves vision transformers View paper
- [22] The norm must go on: Dynamic unsupervised domain adaptation by normalization View paper
- [23] Transferable normalization: Towards improving transferability of deep neural networks View paper
- [24] GRANOLA: Adaptive Normalization for Graph Neural Networks View paper
- [25] The Mathematical Relationship Between Layer Normalization and Dynamic Activation Functions View paper
- [26] Cluster-Based Activation Normalization for Neural Networks View paper
- [27] Adaptive Input Normalization for Quantized Neural Networks View paper
- [28] Real-time spatial normalization for dynamic gesture classification View paper
- [29] Evolving normalization-activation layers View paper
- [30] Switchable normalization for learning-to-normalize deep representation View paper
- [31] Dynamic normalization BPN for stock price forecasting View paper
- [32] Time-varying trajectory modeling via dynamic governing network for remaining useful life prediction View paper
- [33] Layer normalization View paper
- [34] Dynamic layer normalization for adaptive neural acoustic modeling in speech recognition View paper
- [35] Differentiable dynamic normalization for learning deep representation View paper
- [36] Exploring the Efficacy of Group-Normalization in Deep Learning Models for Alzheimer's Disease Classification View paper
- [37] Sketch-guided spatial adaptive normalization and high-level feature constraints based GAN image synthesis for steel strip defect detection data augmentation View paper
- [38] Developing Distance-Aware, and Evident Uncertainty Quantification in Dynamic Physics-Constrained Neural Networks for Robust Bearing Degradation Estimation View paper
- [39] AMIN-CNN: Enhancing Brain Tumor Segmentation through Modality-Aware Normalization and Deep Learning View paper
- [40] A note on factor normalization for deep neural network models. View paper
- [41] Enhancing deep neural network training through learnable adaptive normalization View paper
- [42] Time-varying Gaussian encoder-based adaptive sensor-weighted method for turbofan engine remaining useful life prediction View paper
- [43] Speed-Normalized Expansion-and-Shrinkage network for health indicator generating under time-varying conditions View paper
- [44] Adaptive Normalization Enhances the Generalization of Deep Learning Model in Chest X-Ray Classification View paper
- [45] Global adaptive input normalization for short-term electric load forecasting View paper
- [46] Gesture Recognition of Filipino Sign Language Using Convolutional and Long Short-Term Memory Deep Neural Networks View paper

- [47] Pedestrian intention prediction in Adverse Weather Conditions with Spiking Neural Networks and Dynamic Vision Sensors View paper
- [48] Underwater image enhancement using adaptive standardization and normalization networks View paper
- [49] Gas Turbine Anomaly Detection under Time-Varying Operation Conditions Based on Spectra Alignment and Self-Adaptive Normalization View paper
- [50] Batch Layer Normalization A new normalization layer for CNNs and RNNs View paper
- [51] A review of convolutional neural networks in computer vision View paper
- [52] Gated channel transformation for visual recognition View paper
- [53] Different data cleaning techniques and normalization techniques with focus on current normalization techniques: A study View paper
- [54] Comparing image normalization techniques in an end-to-end model for automated modic changes classification from MRI images View paper
- [55] On feature normalization and data augmentation View paper
- [56] Influence of feature normalization methods on transfer Learning-A comparison study View paper
- [57] Optimal affine image normalization approach for optical character recognition View paper
- [58] High-performance large-scale image recognition without normalization View paper
- [59] Test-time distribution normalization for contrastively learned visual-language models View paper
- [60] Fixup initialization: Residual learning without normalization View paper
- [61] BNPO: Beta Normalization Policy Optimization View paper
- [62] Quality prediction of industrial process based on Kolmogorovâ□□Arnold graph convolution aggregation temporal convolution network View paper
- [63] Pangu Light: Weight Re-Initialization for Pruning and Accelerating LLMs View paper
- [64] DyTTP: Trajectory Prediction with Normalization-Free Transformers View paper
- [65] StyDiff: a refined style transfer method based on diffusion models View paper
- [66] Stream normalization for ctr prediction View paper
- [67] Continuous-Time Analysis of Adaptive Optimization and Normalization View paper
- [68] An Adaptive Mixed-Step Size Normalized Least Means Fourth Control Approach for Stand-Alone Power Generation System Considering Dynamic Conditions View paper
- [69] StatAvg: Mitigating Data Heterogeneity in Federated Learning for Intrusion Detection Systems View paper
- [70] GammaGAN: Gamma-Scaled Class Embeddings for Conditional Video Generation View paper
- [71] Enhancing cross-domain generalization in retinal image segmentation via style randomization and style normalization View paper
- [72] Global partitioning elevation normalization applied to building footprint prediction View paper
- [73] Gradient-Weighted, Data-Driven Normalization for Approximate Border Bases -- Concept and Computation View paper
- [74] Medium-scale projection of reference evapotranspiration beyond available data using sequential deep learning models: a case study from Bangladesh View paper
- [75] Enhanced Model Robustness to Input Corruptions by Per-corruption Adaptation of Normalization Statistics View paper
- [76] SC-GAN: A Style-Conditioned Generative Adversarial Network for High-Quality Artistic Image Generation View paper
- [77] MINTIN: Maxout-Based and Input-Normalized Transformation Invariant Neural Network View paper
- [78] NormSoftmax: Normalizing the Input of Softmax to Accelerate and Stabilize Training View paper