Deep Residual Learning for Image Recognition: Implementation and Analysis

Vinayak Raju Shinde

September 13, 2025

1 Introduction

The degradation problem in deep neural networks represents a fundamental challenge in computer vision, where increasing network depth paradoxically leads to higher training error rather than improved performance. This phenomenon occurs not due to overfitting, but because of optimization difficulties inherent in training very deep networks. He et al. [1] introduced Deep Residual Learning to address this optimization challenge through skip connections that enable the training of networks with substantially increased depth.

This paper presents an empirical comparison between plain convolutional networks and their residual counterparts on the CIFAR-10 dataset. We implement both 20-layer and 56-layer networks with identical architectures, differing only in the presence of residual connections, to isolate and analyze the impact of skip connections on training dynamics and final performance across different depths.

The key contributions of this work are: (1) faithful reproduction of ResNet architecture principles on CIFAR-10, (2) systematic analysis of training dynamics between plain and residual networks at multiple depths, (3) quantitative evaluation of how degradation effects amplify with network depth, and (4) empirical validation of the identity mapping hypothesis.

2 Methodology

2.1 Network Architecture

We implemented networks following the architectural principles described in the original ResNet paper with two depth configurations:

20-Layer Networks: Using the formula $6n + 2$ with $n = 3$, containing three stages with [3,3,3] blocks of 3×3 convolutions and filter counts [16,32,64]. Each block consists of two convolutional layers with batch normalization and ReLU activation.

56-Layer Networks: Using $n = 9$, containing three stages with [9,9,9] blocks, maintaining the same filter progression [16,32,64] but with significantly increased depth.

For each depth, we implemented both: - **Plain Networks:** Standard sequential processing where each layer processes the output of the previous layer - **Residual Networks:** Identical architectures augmented with skip connections implementing $\mathcal{H}(x) = \mathcal{F}(x) + x$

All networks process 32×32 RGB images for 10-class CIFAR-10 classification. Parameter counts are 0.27M for 20-layer and 0.85M for 56-layer networks.

2.2 Training Configuration

Identical training procedures were employed across all experiments:

- Optimizer: SGD with momentum (0.9) and weight decay (1e-4)
- Learning rate: 0.1, reduced by factor of 10 at epochs 80 and 120
- Batch size: 128, Training epochs: 200
- Data augmentation: 4-pixel padding with random crops and horizontal flips
- Normalization: CIFAR-10 standard statistics

3 Results

3.1 20-Layer Comparison

The 20-layer experiments demonstrate measurable benefits of residual connections. ResNet-20 achieved 91.95% test accuracy compared to 90.84% for Plain-20, representing a 1.11 percentage point improvement with $2\times$ faster convergence.

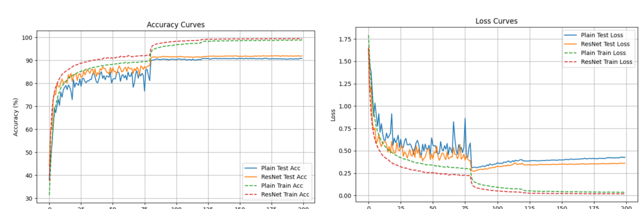

Figure 1: Training dynamics for 20-layer networks showing accuracy (left) and loss (right) curves. ResNet-20 demonstrates superior convergence speed and stability.

3.2 56-Layer Comparison

The 56-layer experiments reveal dramatically amplified benefits, confirming the degradation hypothesis. The performance gap increased to 6.54 percentage points (Plain-56: 86.71%, ResNet-56: 93.25%), representing a 6-fold amplification of residual learning benefits.

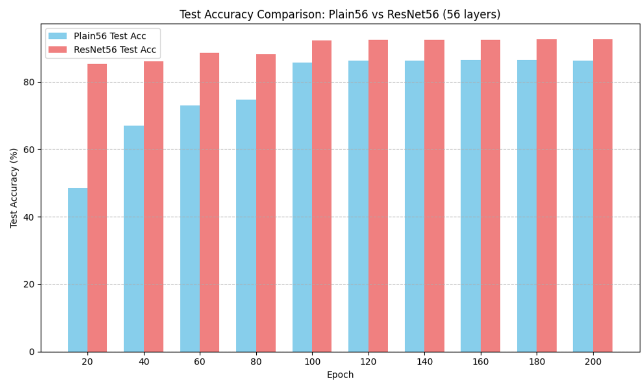

Figure 2: Test accuracy comparison for 56-layer networks showing dramatic performance divergence. Plain-56 exhibits severe degradation while ResNet-56 maintains consistent improvement throughout training.

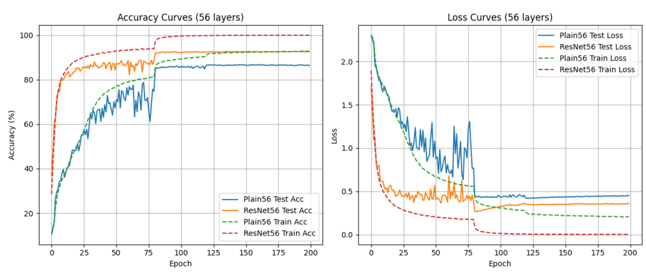

Figure 3: Complete training dynamics for 56-layer networks. The stark differences in convergence behavior illustrate how residual connections enable successful deep network training.

3.3 Comprehensive Performance Analysis

Architecture	Layers	Test Accuracy (%)
Plain	20	90.84
ResNet	20	91.95
Plain	56	86.71
ResNet	56	93.25

Table 1: Comprehensive performance comparison demonstrating degradation in plain networks and successful scaling in ResNets

The results reveal three critical findings:

Degradation Problem: Plain-56 performs worse than Plain-20 (86.71% vs 90.84%), clearly demonstrating that optimization difficulties, not model capacity, limit deep network performance.

Successful Scaling: ResNet-56 outperforms ResNet-20 (93.25% vs 91.95%), showing residual connections enable effective depth scaling.

Amplified Benefits: The advantage of residual connections grows substantially with depth, from 1.11% to 6.54% improvement.

3.4 Classification Quality Analysis

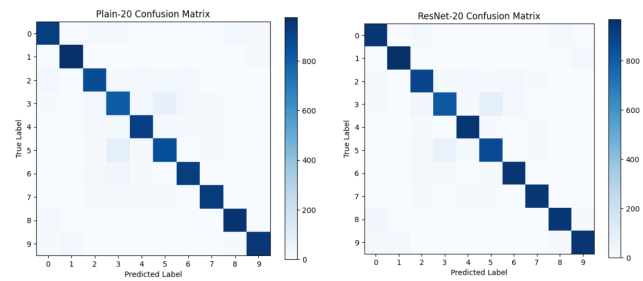

Figure 4: Confusion matrices for 20-layer networks: Plain-20 (left) and ResNet-20 (right)

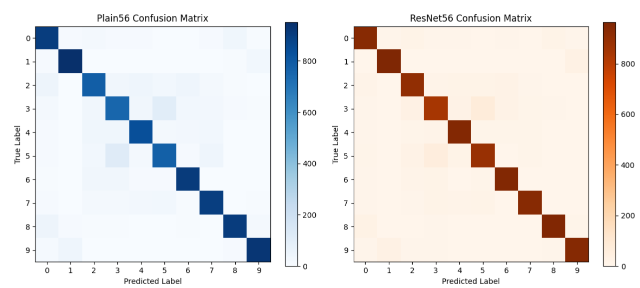

Figure 5: Confusion matrices for 56-layer networks showing ResNet-56's superior classification performance across all CIFAR-10 classes with sharper diagonal patterns indicating more confident predictions.

The confusion matrices reveal that residual networks achieve better classification performance across all classes, with the improvement being more pronounced in deeper architectures. ResNet-56 shows particularly strong diagonal patterns, indicating robust learned representations.

4 Discussion

4.1 Theoretical Validation

Our experiments provide strong empirical validation of core ResNet principles:

Identity Mapping Hypothesis: The superior performance of residual networks suggests that skip connections enable layers to learn identity mappings when beneficial, while plain networks struggle to approximate identity functions through stacked nonlinearities.

Gradient Highway Effect: The smooth training curves of ResNets, particularly evident in 56-layer networks, demonstrate that skip connections provide effective gradient pathways, enabling stable optimization in very deep architectures.

Optimization Landscape: Residual connections appear to create more favorable optimization landscapes with reduced vanishing gradient effects and improved training stability.

4.2 Scaling Behavior Analysis

The depth comparison reveals fundamental differences in scaling behavior:

Plain Network Degradation: As depth increases, plain networks suffer from increasingly severe optimization challenges, resulting in worse performance despite increased capacity.

ResNet Scalability: Residual networks maintain stable training characteristics across depths, with performance improvements that justify increased computational cost.

Practical Implications: For applications requiring high accuracy, deeper ResNets provide substantial benefits (93.25% vs 91.95%) that may justify the increased parameter count and computational requirements.

4.3 Architectural Insights

The success of residual learning can be understood through multiple perspectives:

Feature Learning: Skip connections enable more effective feature learning by allowing layers to build incrementally upon previous representations rather than learning complete transformations.

Ensemble Interpretation: ResNets can be viewed as implicit ensembles of networks with different effective depths, providing robustness and improved generalization.

Information Flow: Identity mappings preserve information from earlier layers, enabling the network to selectively refine features while maintaining access to lower-level representations.

5 Conclusion

This comprehensive study validates the fundamental principles of Deep Residual Learning through systematic comparison of plain and residual networks at 20 and 56-layer depths on CIFAR-10. Our results demonstrate that residual connections provide modest but consistent benefits in shallow networks (1.11% improvement), with advantages becoming dramatic in deeper architectures (6.54% improvement).

The clear manifestation of the degradation problem in Plain-56, combined with ResNet-56's superior performance, confirms that optimization challenges rather than model capacity were the primary barrier to training deep networks. Residual connections address this challenge by providing direct gradient pathways and enabling layers to learn identity mappings when beneficial.

Key findings include:

- ResNet enables successful scaling where plain networks fail
- Training stability improves dramatically with skip connections
- Performance gains amplify substantially as network depth increases
- Convergence characteristics remain favorable even in very deep networks

These results have profound implications for deep learning architecture design. The success of residual learning has transformed computer vision and established skip connections as foundational techniques across numerous domains. Our empirical validation demonstrates that practitioners can confidently employ deeper residual architectures to achieve substantial performance improvements, with benefits becoming more pronounced as architectural complexity increases.

The practical significance extends beyond academic validation, confirming that residual connections should be standard components in deep network design, particularly when depth exceeds 20-30 layers. This supports the continued development of increasingly sophisticated deep learning models built upon residual learning principles.

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.