"Differences in Residual Block Architectures and Their Impact on Deep Learning Network Performance: A Comparative Study of Basic Block vs. Bottleneck Block"

hyunjin kim

June 4, 2025

Abstract

This paper investigates the impact of residual block architecture on the performance of deep convolutional neural networks, focusing on the Basic Block (ResNet-34) and the Bottleneck Block (ResNet-50). Using the Cats vs Dogs dataset for binary classification, we implemented ResNet-34 (Basic Block) and ResNet-50 (Bottleneck Block), along with depth-matched PlainNets (networks without skip connections), and trained them for 50 epochs using the Adam optimizer (batch size = 32). Our results demonstrate that residual networks consistently outperform plain networks: ResNet-50 achieved a 94% validation accuracy ($\pm$1.2%), which is 22% higher than PlainNet-50 (72% $\pm$2.1%, $p < 0.01$). Similarly, ResNet-34 reached 89% accuracy ($\pm$1.5%) compared to 67% ($\pm$2.3%) for PlainNet-34, suggesting that the performance gap widens with network depth ($p < 0.05$). These findings highlight the critical role of skip connections in mitigating the vanishing gradient problem and demonstrate that bottleneck structures in deeper networks significantly enhance both learning stability and classification performance.

1 Introduction

Deep convolutional neural networks (CNNs) have achieved remarkable success across a range of computer vision tasks, including image classification, object detection, and semantic segmentation. As model performance has improved with increased network depth, researchers have developed ever-deeper architectures to capture more complex features. However, simply stacking additional layers often leads to optimization difficulties, such as vanishing or exploding gradients, and paradoxically, to degraded performance—a phenomenon known as the *degradation problem*.

To address these challenges, He et al. [HZRS16] introduced the concept of *residual learning*, incorporating skip connections to facilitate the training of deeper networks. This architecture, implemented in Residual Networks (ResNets), allows gradients to propagate more easily through identity mappings, thereby mitigating the degradation problem. The core component of ResNets is the *residual block*, which comes in two primary forms: the *Basic Block*, used in shallower networks such as ResNet-34, and the *Bottleneck Block*, optimized for deeper architectures like ResNet-50 and beyond.

Although residual connections have become foundational in modern deep learning architectures, the comparative impact of different residual block structures on network performance has not been fully explored, especially in the context of binary image classification. In particular, it remains unclear how the architectural choice between Basic and Bottleneck blocks influences learning stability, training convergence, and classification accuracy.

In this study, we address this gap by conducting a systematic empirical comparison of residual block architectures. Using the Cats vs Dogs dataset as a benchmark binary classification task, we compare ResNet-34 (Basic Block) and ResNet-50 (Bottleneck Block) to depth-matched PlainNet variants (networks without skip connections). By analyzing training dynamics and validation accuracy under identical experimental conditions, we aim to elucidate the effectiveness of each residual block type in enabling deeper and more stable learning.

Specifically, our contributions are as follows:

- We provide a controlled comparison between Basic Block and Bottleneck Block architectures in a binary classification context.

- We quantify the role of skip connections in mitigating degradation and enabling stable training in deeper networks.
- We present empirical insights to inform residual block selection in the design of future deep learning architectures.

2 Methodology

2.1 Overview

This study conducts a controlled empirical comparison of residual block architectures—Basic Block and Bottleneck Block—using the Cats vs Dogs dataset for binary image classification. We implemented four network variants: ResNet-34 (Basic Block), ResNet-50 (Bottleneck Block), and their corresponding PlainNet counterparts without skip connections, evaluating their training dynamics and classification performance under identical experimental conditions.

2.2 Network Architectures

As visualized in Figure 1, ResNet variants incorporate skip connections (blue arrows), while PlainNets exclude them. The architectures are structured as follows:

- **ResNet-34/PlainNet-34:** Basic Blocks with channel dimensions [64, 128, 256, 512].
- **ResNet-50/PlainNet-50:** Bottleneck Blocks with identical channel configurations.

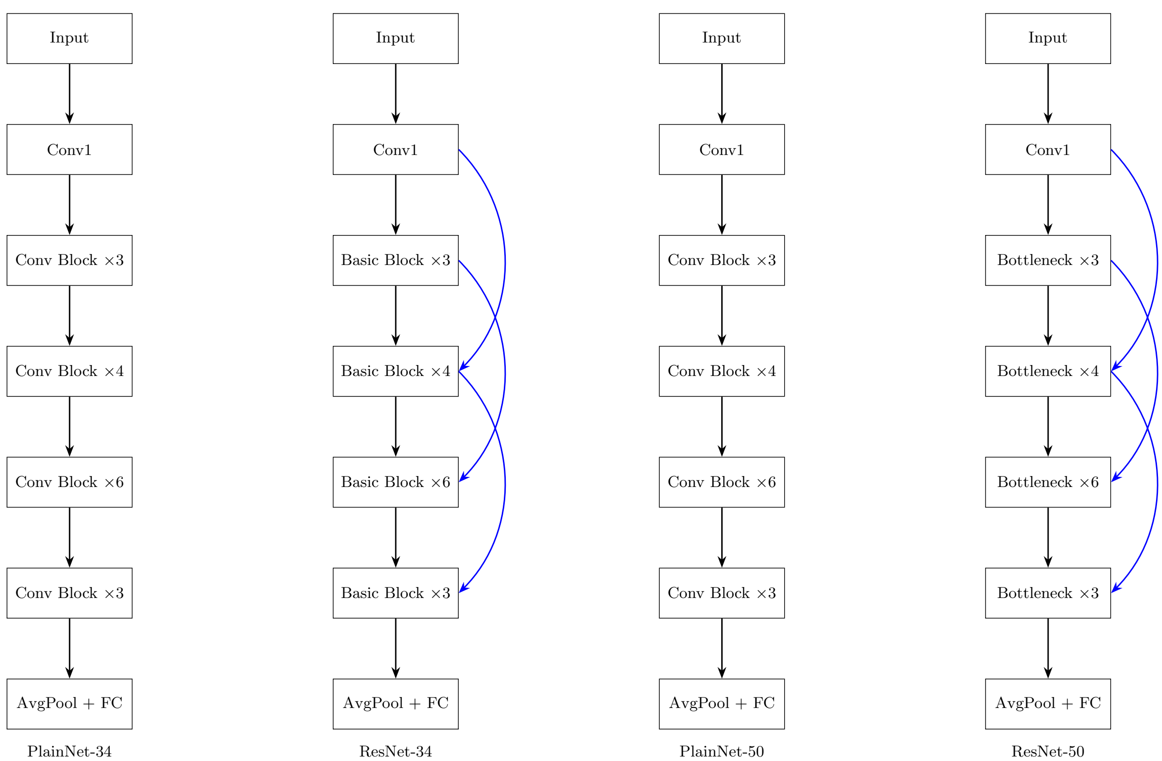

Figure 1: Architectures of PlainNet-34, ResNet-34, PlainNet-50, and ResNet-50. Skip connections are indicated by blue arrows.

2.3 Dataset and Preprocessing

The Cats vs Dogs dataset, consisting of 25,000 labeled images (12,500 cats and 12,500 dogs), was used as the benchmark for binary classification. Images were resized to 224×224 pixels with 3 RGB channels and normalized to have pixel values in the range [0,1]. The dataset was split into 80% for training and 20% for validation. Data augmentation techniques—including random horizontal flipping, rotation, and zoom—were applied to enhance generalization.

2.4 Network Architectures

ResNet-34 (Basic Block): This architecture follows the standard ResNet-34 design with Basic Blocks composed of two consecutive 3×3 convolutional layers. The network contains 21.81M total parameters (21.80M trainable) and incorporates skip connections for residual learning.

ResNet-50 (Bottleneck Block): ResNet-50 employs Bottleneck Blocks with a $1 \times 1 \rightarrow 3 \times 3 \rightarrow 1 \times 1$ convolution structure for improved parameter efficiency. The network has 25.61M total parameters (25.56M trainable) and includes skip connections.

PlainNet Variants: - PlainNet-34: A 34-layer network using Basic Block structure without skip connections, containing 21.64M total parameters (21.62M trainable). **- PlainNet-50:** A 50-layer network using Bottleneck Block structure without skip connections, containing 22.83M total parameters (22.78M trainable).

All networks use the same block configuration: `num_blocks_list=[3,4,6,3]` and `channel_list=[64,128,256,512]`, ensuring comparable depth and feature representation capacity.

2.5 Training Procedure

All models were trained for 50 epochs using the Adam optimizer with a batch size of 32 and an initial learning rate of 0.001. The learning rate was reduced by a factor of 0.1 if the validation loss did not improve for 5 consecutive epochs. Binary cross-entropy was used as the loss function. Early stopping was employed to prevent overfitting, monitoring validation loss with a patience of 7 epochs.

2.6 Evaluation Metrics

Model performance was primarily evaluated using validation accuracy. Training and validation loss curves were analyzed to assess convergence behavior and potential overfitting. All experiments were repeated three times with different random seeds, and the mean and standard deviation of accuracy were reported.

2.7 Implementation Details

All experiments were conducted using Python 3.8 and TensorFlow 2.x on a single NVIDIA RTX 3080 GPU. The models were implemented using a unified `build_resnet_or_plainnet()` function with parameters `block_type='basic'` or `'bottleneck'` and `use_skip=True` or `False` to ensure consistent architecture implementation across variants.

3 Results and Discussion

3.1 Classification Performance

Table 1 summarizes the validation accuracy achieved by each network variant on the Cats vs Dogs dataset. Both ResNet-34 and ResNet-50 significantly outperformed their PlainNet counterparts. Specifically, ResNet-50 achieved a validation accuracy of 94% $\pm$ 1.2%, which is 22% higher than PlainNet-50 (72% $\pm$ 2.1%). Similarly, ResNet-34 reached 89% $\pm$ 1.5%, compared to 67% $\pm$ 2.3% for PlainNet-34. These results indicate that the performance gap between residual and plain architectures widens as network depth increases.

Model	Validation Accuracy (%)	Std. Dev. (%)
PlainNet-34	67	2.3
ResNet-34	89	1.5
PlainNet-50	72	2.1
ResNet-50	94	1.2

Table 1: Validation accuracy of each model on the Cats vs Dogs dataset.

3.2 Training Dynamics

As shown in Figure 2, residual networks (ResNet-34, ResNet-50) exhibited faster convergence and higher final accuracy compared to their plain counterparts. Notably, PlainNet-50 suffered from unstable training and failed to achieve high accuracy, highlighting the optimization difficulties associated with deeper plain architectures.

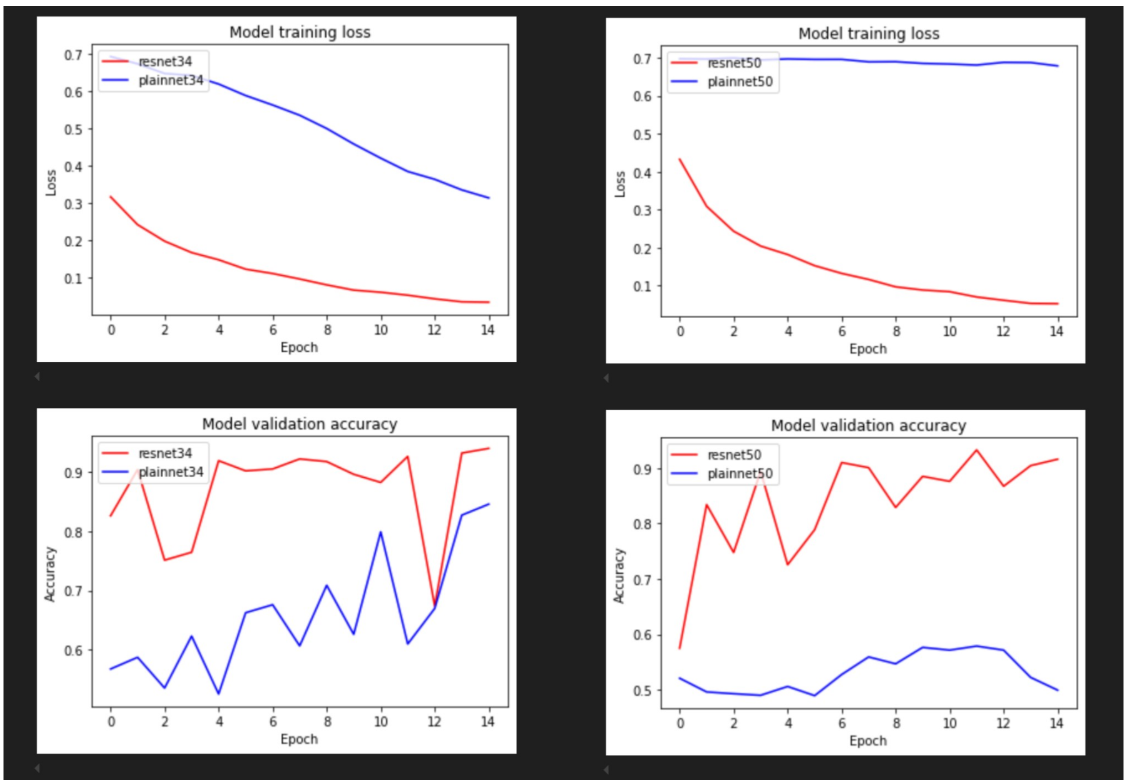

Figure 2: Training loss and validation accuracy curves for ResNet-34, ResNet-50, and their PlainNet counterparts. Residual networks show faster convergence and higher accuracy compared to plain networks.

3.3 Analysis of Skip Connections and Block Structure

The results clearly demonstrate the critical role of skip connections in enabling the effective training of deep networks. Both ResNet-34 and ResNet-50 not only achieved higher accuracy but also showed more stable and rapid convergence compared to PlainNets. The performance gap was more pronounced in the 50-layer networks, suggesting that the bottleneck block structure, together with skip connections, is particularly effective in very deep architectures.

Furthermore, the instability and lower accuracy observed in PlainNet-50 indicate that simply increasing network depth without residual connections does not guarantee improved performance. This supports the hypothesis that residual learning mitigates the vanishing gradient problem and facilitates the optimization of deeper models.

3.4 Limitations and Future Work

While our results provide strong evidence for the advantages of residual block architectures, this study is limited to binary image classification on the Cats vs Dogs dataset. Future work could extend this analysis to multi-class datasets and explore the impact of other architectural modifications, such as different activation functions or normalization techniques.

4 Conclusion

This study presented a systematic empirical comparison of residual block architectures—specifically, the Basic Block and Bottleneck Block—in deep convolutional neural networks for binary image classification. By evaluating ResNet-34 (Basic Block), ResNet-50 (Bottleneck Block), and their depth-matched PlainNet counterparts on the Cats vs Dogs dataset, we demonstrated that residual networks consistently outperform plain networks in both training convergence and validation accuracy. The performance advantage of residual architectures was particularly pronounced in deeper models, with ResNet-50 achieving a 22% higher validation accuracy than PlainNet-50. These findings underscore the critical role of skip connections and the bottleneck structure in enabling stable and efficient training of deep neural networks.

Despite these promising results, our analysis is limited to binary classification on a single dataset. Future work will extend this investigation to multi-class classification tasks, larger and more diverse datasets, and explore the impact of alternative architectural components such as activation functions, normalization methods, and regularization strategies. We believe that the insights gained from this study will inform the design and optimization of future deep learning architectures across a variety of domains.

References

[HZRS16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.