Analysis of Residual Block in the Resnet for Image Classification

Xiayuan Jin
School of Engineering, The University of Manchester, Manchester, U.K.

Keywords: ResNet, Image Classification, Convolutional Neural Network, CIFAR-10 Dataset.

Abstract: Image recognition is of paramount importance in our contemporary world, with diverse applications across domains such as traffic control, medical diagnosis, educational tools, and workplace automation. Its impact is profound and multifaceted. This study highlights Resnet's effectiveness in building robust deep-learning models for image classification. Through the integration of residual blocks, residual network (ResNet) overcomes challenges like vanishing gradients, enabling the training of very deep networks. Experiments on the CIFAR-10 dataset showcase ResNet's impressive accuracy in image recognition, with loss fluctuations mitigated via hyperparameter tuning. ResNet excels in feature extraction and precise image classification. The important topic of this research is trying to figure out the efficiency comparing convolutional neural network (CNN) and Resnet using Resnet's residual block to find out the difference of parameter changes the accuracy of the model, by inducting Resnet, the performance of the model behaves much better, and solve the problem of gradient vanishment, Resnet plays a pivotal role in image classification by enabling the training of very deep neural networks, enhancing feature extraction, and achieving state-of-the-art accuracy in various visual recognition tasks.

1 INTRODUCTION

In today's digitally driven world, the surge in modern technology has led to an era of unprecedented data generation and spread over the internet. This rapid growth in information necessitates robust tools for managing diverse data forms, particularly images. Various industries now require proficient image classification techniques to make sense of this information influx.

The previous machine learning material provided insights into fundamental algorithms and concepts, spanning across supervised learning, unsupervised learning, and deep learning domains. You acquired skills in data preprocessing, feature engineering, and model evaluation, establishing a robust foundation for tackling real-world problem-solving (Schaetti 2018). Data scientists used a basic Convolutional Neural Network (CNN) network to do machine learning work. For example, In the realm of Deep Learning (DL), CNNs have gained immense prominence (Krizhevsky et al 2017). One of their distinguishing strengths compared to their predecessors is their ability to autonomously extract important features without human supervision (Gu et al 2018). CNNs provide extensive applications across diverse fields, including computer vision (Fang et al 2020), gesture processing (Palaz et al 2019), and even Face Recognition (Lu et al 2018). The basic concept of CNNs takes inspiration from the neural organization in both human and animal health bodies, notably in the visual cortex. For instance, the complex network of cells constituting a feline's visual cortex finds an analogous representation in CNNs (Li et al 2020). Goodfellow underscores three pivotal features of CNNs: CNNs stand out due to their capacity for creating equivalent representations, promoting sparse interactions, and enabling parameter sharing. In contrast to conventional fully connected (FC) networks, CNNs leverage shared weights and localized connections to exploit the inherent 2D structure within input data, particularly in the case of image signals. This architectural choice not only reduces unnecessary parameters but also mirrors the selective information processing observed in our brain's visual system, similar to how our visual cells focus on specific areas (Hubel and Wiesel 1962). This method not only reduces unnecessary parameters but also makes training more efficient, much like how our brain's visual cells selectively process information. Just as these cells concentrate on specific areas, CNNs use local filters to extract important details from the input (Goodfellow et al 2016). CNNs can face challenges like overfitting, high computational load,

Jin, X.
Analysis of Residual Block in the ResNet for Image Classification.
DOI: 10.5220/0012800400003885

In *Proceedings of the 1st International Conference on Data Analysis and Machine Learning (DAML 2023)*, pages 246-250
ISBN: 978-989-758-705-4

limited contextual understanding, and issues with translation invariance. Residual Networks, addresses these by introducing residual blocks, which enable direct gradient flow during training. ResNet, a form of deep neural network, incorporates skip connections or shortcuts in its architecture (Schaetti 2018). Applying in the domain of autonomous driving, ResNet emerges as a key for accurately identifying and categorizing road objects. This capability significantly fortifies the safety and reliability of self-driving vehicles. Similarly, the medical sector benefits from ResNet's prowess, which accelerates the detection of anomalies in medical images, thereby expediting disease diagnosis and prognostication. ResNet's adaptability and effectiveness across these complicated sectors underscore its compelling ability to address real-world complexities in image analysis. Consequently, ResNet stands out as a multifunctional solution to address the contemporary challenges of image processing.

The main objective of this study is to utilize ResNet for constructing an efficient image classification model and using different numbers of residual models to find out the change in efficiency. By introducing residual blocks, the aim is to counter overfitting and enhance the model's ability to retain crucial original feature information, leading to improved feature representation. Specifically, firstly, the incorporation of ResNet's residual blocks addresses issues such as vanishing gradients and degradation, which often hinder the training of deep networks. This helps ensure smoother gradient propagation during training, facilitating the learning process. Secondly, the utilization of skip connections within ResNet aids in maintaining and transmitting essential features across layers, thereby mitigating the loss of valuable information. Thirdly, an in-depth analysis and comparison of predictive performances across various models are conducted. Furthermore, the integration of ResNet addresses the challenge of training deep neural networks effectively. ResNet's architecture with residual blocks helps alleviate the degradation problem, enabling successful training of networks with extensive depth. Simultaneously, the incorporation of skip connections supports gradient flow during backpropagation, effectively tackling the issue of vanishing gradients that frequently arise during the training of deep networks. These strategic enhancements collectively contribute to ResNet's efficacy in overcoming challenges associated with training deep networks and accomplishing image classification tasks.

2 METHODOLOGY

2.1 Dataset Description and Preprocessing

The research used the Canadian Institute for Advanced Research (CIFAR) dataset to explore the problem around ResNet (Abouelnaga et al 2016). The CIFAR-10 dataset, a cornerstone in the realm of computer vision, originates from the CIFAR and serves as a vital benchmark for image classification tasks. Comprising 60,000 color images with dimensions of 32x32 pixels, the dataset encompasses ten diverse classes, each containing 6,000 images. This study delves into the enhancement of the CIFAR-10 dataset for robust model training through a sequence of preprocessing techniques. Segmented into 50,000 training images and 10,000 test images, the dataset facilitates rigorous model evaluation.

The core focus of this study lies in the application of preprocessing methods to amplify the dataset's efficacy. The "Rescaling" technique, normalizing pixel values to the [0, 1] range, is coupled in horizontal and vertical modes, introducing data augmentation by enabling random image flips. Horizontal flips emulate diverse object orientations, while vertical flips inject variability in object positioning. These techniques collectively bolster model robustness, alleviate overfitting concerns, and elevate generalization capabilities. The integration of these preprocessing measures results in an elevated quality of the CIFAR-10 dataset, rendering it particularly suitable for training CNNs and other image recognition models. The optimized dataset plays a pivotal role in advancing classification accuracy and fostering adaptability for object recognition in real-world settings. This paper underscores the significance of preprocessing methodologies in refining image datasets, underscoring their role in advancing the performance of machine learning models in the realm of computer vision.

2.2 Proposed Approach

Introduction to the research technology. ResNet is an exciting and groundbreaking deep learning architecture known for its ability to train extremely deep networks effectively. By leveraging the concept of residual blocks, ResNet has shown remarkable performance in various computer vision tasks. This architecture combines several components, such as convolutional layers, residual pathways, global average pooling, and fully connected layers, all working together harmoniously. The pipeline of the ResNet architecture can be visualized in Fig. 1,

providing a comprehensive overview of its structure and flow. At the heart of this architecture lies the Convolution operation, specifically the 2-dimensional convolution (Conv2D). Conv2D is a fundamental operation in deep learning and serves as a powerful tool for extracting meaningful features from images. It employs a sliding window mechanism that scans the input image, highlighting patterns and capturing relevant information necessary for subsequent processing. To optimize computational efficiency and enhance the network's ability to focus on critical features, ResNet incorporates another essential operation known as 2-dimensional Max Pooling (MaxPooling2D). MaxPooling2D downsamples the data by selecting the maximum values within specific regions, effectively reducing the dimensionality of feature maps while emphasizing the most important features. By discarding non-maximal values, MaxPooling2D helps to reduce noise, improve robustness, and enhance the overall performance of the deep learning model.

By combining Conv2D and MaxPooling2D within the ResNet architecture, researchers and practitioners have unlocked new possibilities and achieved state-of-the-art results in image classification, object detection, semantic segmentation, and various other computer vision tasks. This powerful combination of operations has proven crucial in extracting rich and discriminative features, enabling deep networks to learn intricate patterns and make accurate predictions in complex visual tasks.

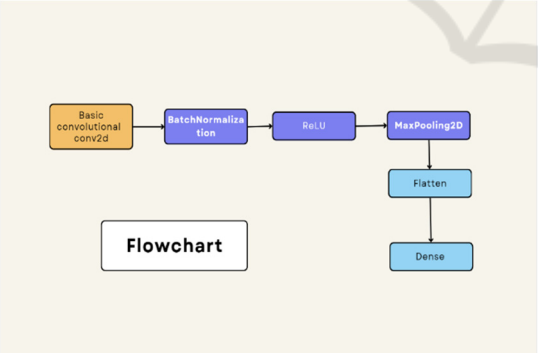

Figure 1: The pipeline of the model (Picture credit: Original).

2.2.1 Conv2D

Introduction to the Conv2D technique. The Conv2D layer plays a pivotal role in CNN, serving as a fundamental element for feature extraction from images. Operating on the principle of two-dimensional convolution, this layer convolves filters across input images or feature maps to uncover intricate spatial patterns, edges, and textures. This iterative process progressively constructs a hierarchical representation of features essential for image analysis. Practically, Conv2D requires a 4D tensor input encompassing parameters like batch size, image dimensions, and channel count. Through convolution operations, it computes dot products between filters and localized input regions, generating feature maps. Its primary strength lies in automated feature extraction, enhancing the network's capability to capture significant visual cues and enabling robust image comprehension. Integrated into the ResNet architecture, the Conv2D layer initiates the process of feature extraction, producing intermediary feature maps that undergo activation functions for further processing. Its significance resonates across deep learning models, solidifying its status as a cornerstone in image recognition and computer vision.

2.2.2 ResNet

Another interesting module is the Residual block. Residual Blocks, a groundbreaking innovation introduced by ResNet, offer a transformative solution to challenges posed by training deep neural networks. These blocks effectively address the vanishing gradient issue through residual connections, enabling direct information flow across layers. The core functionality of Residual Blocks revolves around their capacity to learn incremental changes in feature representations. Instead of focusing on complex transformations, these blocks allow networks to concentrate on residual or incremental updates, streamlining learning within deeper architectures.

Residual Blocks find optimal utility within deep networks, particularly those susceptible to gradient degradation. By retaining earlier features via shortcut connections, these blocks ensure seamless gradient propagation—a pivotal factor in optimizing networks with extensive layers. In our implementation, Residual Blocks materialize using Conv2D layers and shortcut connections, forming essential components of the ResNet architecture. Their unique capability to mitigate gradient vanishing while facilitating the training of exceptionally deep networks has revolutionized the realm of deep learning. This innovation not only rekindles the pursuit of deeper architectures but also establishes new benchmarks for image recognition, object detection, and other computer vision tasks.

2.2.3 Loss Function

Loss function is one of the important elements in the

research. In the realm of multi-class classification tasks, an important concept of the loss function is the variance, which is the difference between the expectation value and the true value, and it quantifies the gap between real value and predicted value. The Sparse Categorical Cross-Entropy (SparseCE) loss function emerges as a pivotal construct within this context, meticulously tailored to measure the alignment between model predictions and ground truth labels. The core equation encapsulating its essence involves the summation of the negative logarithm of predicted probabilities across all samples: Loss (L) equals the negative sum of the logarithm of the predicted probabilities (pi) for each sample:

$$L = \sum_{i=1}^{N} \log(pi) \quad (1)$$

where L signifies the computed loss value, representing the degree of divergence between predicted probabilities and true labels. Loss function indicates summation across all N samples in the dataset, underscoring its comprehensive nature. $\log(pi)$denotes the natural logarithm of the predicted probability pi assigned to the true class of the ith sample, quantifying the model's confidence in its prediction. pi represents the predicted probability assigned to the true class of the ith sample, encapsulating the model's estimation of the likelihood that the sample belongs to its correct class.

2.3 Implementation Details

The research provides the construction and training of CNN models for image classification tasks using TensorFlow and Keras. The foundation of the system is established by importing the necessary libraries and modules. It introduces classes and functions that enable the creation of diverse CNN architectures, data preprocessing, training, and performance evaluation. Notably, data augmentation techniques are harnessed through a sequential layer, enhancing the dataset's variety by applying rescaling and random flips. This augmentation strategy bolsters the model's robustness and capacity to generalize to different image variations. The code also acknowledges the significance of hyperparameters by allowing customization of crucial factors such as learning rate, batch size, and data storage path. These parameters wield considerable influence over the model's training trajectory and final performance. In essence, the code amalgamates these elements into a cohesive framework that amalgamates system background awareness, data augmentation practices, and flexible hyperparameter configurations to cultivate effective image classification models.

3 RESULTS AND DISCUSSION

The ResNet loss function curve exhibits some rebounds between 20 epochs, possibly due to overfitting, suboptimal hyperparameters, and complex optimization. Strategies to mitigate rebounds include adjusting learning rates and regularization. The accuracy curve indicates ResNet's convergence around 20 epochs, with steeper changes attributed to its depth. ResNet's training "rebound" relates to its depth, gradient challenges, epoch times, learning rate, and number of layers.

The Fig. 2 below illustrates the loss function of the Resnet method after 20 times epochs, the loss function rebounds and turns better after this. The rebound observed in the loss curve during ResNet training can be attributed to factors such as model overfitting, suboptimal hyperparameters, and the intricate optimization landscape of deep networks. The occasional rise in loss indicates potential challenges in achieving convergence due to overfitting, while oscillations can stem from a high learning rate causing overshooting. Moreover, complex optimization layers may result in temporary fluctuations as the algorithm seeks the global minimum. Strategies to mitigate rebounds include adjusting learning rates, regularizing the model, and applying learning rate schedules.

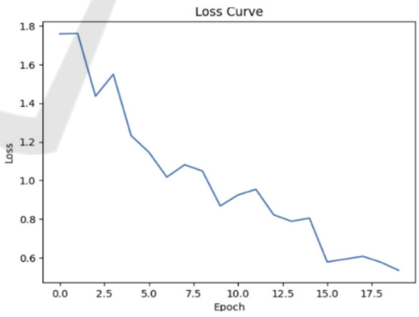

Figure 2: Loss curve (Picture credit: Original).

Based on Fig. 3 (the accuracy function), the diagram can observed that the Resnet model's convergence point converges around 20 times epochs. However, the diagram is still steep since Resnet has more layers than normal CNN, and some of the gradients change faster than normal networks. During ResNet training, the phenomenon of "rebound" may occur due to the network's depth and complexity. Challenges in gradient propagation and learning rate adjustment can lead to oscillations in loss and

accuracy curves. However, with proper optimization strategies, ResNet can recover from these fluctuations and achieve favorable outcomes.

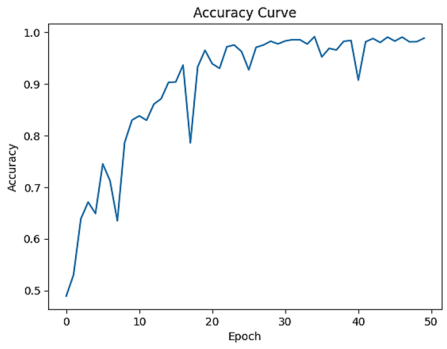

Figure 3: Accuracy of model (Picture credit: Original).

4 CONCLUSION

This study underscores the effectiveness of ResNet in building robust deep-learning models for image classification tasks. By integrating residual blocks into the CNN architecture, ResNet successfully tackles challenges like vanishing gradients and degradation. These residual connections ensure smooth gradient propagation during backpropagation, enabling the training of exceptionally deep networks. Through extensive experiments conducted on the CIFAR-10 dataset, ResNet demonstrates its ability to achieve remarkable accuracy in image recognition. While the loss function exhibits occasional fluctuations during training, meticulous hyperparameter tuning, including learning rate adjustments, effectively mitigates these fluctuations. The results validate ResNet's ability to extract distinctive visual features and accurately classify images. Future research avenues may explore enhancements to the residual blocks, including the implementation of bottleneck architectures to improve computational efficiency. Additionally, evaluating ResNet on datasets with more complex images can assess its generalization potential. In a rapidly evolving landscape of deep learning and image classification, ResNet stands as an exemplar of innovation and advancement. As we continue our journey in this field, we eagerly anticipate the innovations and breakthroughs that will inevitably shape the future of image recognition and deep learning. The potential is vast, and the promise is bright.

REFERENCES

N. Schaetti, “Character-based Convolutional Neural Network and ResNet 18 for Twitter Author Profiling. Notebook for PAN at CLEF 2018,” 2018.

A. Krizhevsky, I. Sutskever, and G.E Hinton, “Imagenet classification with deep convolutional neural networks,” Communications of the ACM, vol. 60, 2017, pp. 84–90.

J. Gu, Z. Wang, J. Kuen, et al, “Recent advances in convolutional neural networks,” Pattern Recognition, vol. 77, 2018, pp. 354–77.

W. Fang, P. E. Love, H. Luo, and L. Ding, “Computer vision for behavior-based safety in construction: a review and future directions,” Advanced Engineering Informatics, vol. 43, 2020, p. 100980.

D. Palaz, M. Magimai-Doss, and R. Collobert, “End-to-end acoustic modeling using convolutional neural networks for hmm-based automatic speech recognition,” Speech Communication, vol. 108, 2019, p. 15–32.

Z. Lu, X. Jiang, and A. Kot, “Deep Coupled ResNet for Low-Resolution Face Recognition,” IEEE Signal Processing Letters, vol. 25, 2018, p. 526-530.

H. C. Li, Z. Y. Deng, and H. H. Chiang, “Lightweight and resource-constrained learning network for face recognition with performance optimization,” Sensors, vol. 20, 2020, p. 6114.

D. H. Hubel, and T. N. Wiesel, “Receptive fields, binocular interaction, and functional architecture in the cat’s visual cortex,” Journal of Physiology, vol. 160, 1962, p. 106.

I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, “Deep learning,” vol. 1, 2016.

Y. Abouelnaga, O. S. Ali, H. Rady, and M. Moustafa, “CIFAR-10: KNN-Based Ensemble of Classifiers,” 2016 International Conference on Computational Science and Computational Intelligence (CSCI), 2016, pp.