Comparative Analysis of Image Classification Algorithms Based on Traditional and Advanced Convolutional Neural Networks

Taimingwang Liu
School of AI and Advanced Computing, Xi'an Jiaotong-Liverpool University, Taicang, Jiangsu, China

Keywords: Image Classification, Convolutional Neural Networks, Architectural Depth, Dataset Characteristics, Comparative Analysis.

Abstract: This study presents a comprehensive comparative analysis of image classification algorithms across diverse datasets and distinct convolutional neural network (CNN) architectures. The datasets considered—CIFAR-10, CALTECH-101, and STL-10—embody varying complexities characteristic of real-world scenarios. They span scenarios of limited categories and low-resolution images to challenges involving diverse instances with fewer categories and high-resolution demands. The selected CNN architectures—LeNet5, VGG16, and ResNet50—exhibit varying depths and design philosophies, offering a diverse landscape for evaluation. Systematic experimentation and evaluation unveil the intricate interplay between architectural complexity and dataset characteristics. The findings underscore the pivotal role of architectural depth in addressing diverse dataset challenges. Notably, VGG16 and ResNet50 consistently outperform LeNet5 across all datasets, emphasizing the importance of deeper architectures in image classification tasks. These insights provide valuable guidance for architectural choices in image classification, ensuring alignment with specific dataset characteristics. Additionally, the study lays the foundation for future research endeavors aimed at refining architectural designs and enhancing image classification algorithm performance across various real-world scenarios.

1 INTRODUCTION

Image classification, a foundational task in the realm of computer vision, holds a central role in a multitude of real-world applications, ranging from object recognition and medical imaging to enabling autonomous vehicles. Throughout the years, CNN has emerged as the predominant approach for image classification, achieving remarkable success due to its innate capability to autonomously extract pertinent features from raw image data. Moreover, it is worth noting that Deep Learning (DL) stands as an effective solution for addressing various image processing challenges, such as facial recognition (Yu et al, 2019). The ever-evolving landscape of CNN architectures, encompassing both traditional and advanced models, presents a compelling opportunity to assess their comparative performance across diverse datasets (Bhatt et al, 2021).

This paper aims to conduct a comparative analysis of image classification algorithms, focusing on both traditional and advanced CNN architectures. Specifically, the performance of LeNet-5, ResNet50, and VGG16 was investigated across three diverse benchmark datasets: CALTECH-101, CIFAR-10, and STL-10. This study sheds light on the architectures' suitability for handling various image classification challenges, while also offering insights into the significance of employing advanced networks.

2 RELATED WORK

The landscape of image classification has been significantly shaped by the emergence of CNNs, which have revolutionized the field and demonstrated exceptional performance in a variety of applications (Pei et al, 2019). This section provides an overview of the pertinent literature surrounding image classification algorithms and their comparative evaluations.

2.1 Image Classification and CNNs

Image classification involves assigning a label to an image from a predefined set of categories. CNNs,

Liu, T.
Comparative Analysis of Image Classification Algorithms Based on Traditional and Advanced Convolutional Neural Networks.
DOI: 10.5220/0012798600003885

In *Proceedings of the 1st International Conference on Data Analysis and Machine Learning (DAML 2023)*, pages 218-223
ISBN: 978-989-758-705-4

inspired by the organization of the visual cortex in animals, have shown remarkable performance in image classification tasks. LeNet-5, introduced by LeCun et al., marked a significant step in the evolution of CNNs, demonstrating the potential of deep learning for feature extraction (Lecun et al, 1998 & Liu et al, 2022). Subsequent architectures, such as VGG16 and ResNet50, introduced deeper architectures with skip connections and improved accuracy (Simonyan and Zisserman, 2015 & He et al, 2016).

2.2 Comparative Analysis of CNN Architectures

Comparative analyses of CNN architectures have become essential to understand the strengths and weaknesses of different models. Previous studies have focused on benchmark datasets to assess architecture performance. For instance, Krizhevsky et al. evaluated different CNN architectures on the ImageNet dataset, demonstrating the effectiveness of deep networks in large-scale image classification (Krizhevsky et al, 2012). More recent work by Tan & Le introduced EfficientNet, showcasing superior performance while being computationally efficient (Tan and Le, 2019).

2.3 Challenges and Dataset Selection

Dataset selection plays a pivotal role in shaping the outcomes of comparative studies. It is crucial to choose diverse datasets with distinct characteristics to provide a comprehensive evaluation of the capabilities of different architectural models. This study selected three widely used benchmark datasets, each with its unique features and challenges.

- *CIFAR-10:* CIFAR-10 is a classic dataset that contains samples from ten major image classes shown in Fig. 1. This dataset focuses on the basic class classification of objects and is suitable for evaluating the performance of image classification algorithms in identifying common objects.

Figure 1: Images from CIFAR-10 (Picture credit: Original).

- *STL-10:* The composition of STL-10 is similar to CIFAR-10. It also includes ten categories with natural scenes and objects common in life, which are almost the same as the previous one (Coates et al, 2021) . Fig. 2 shows examples of each class. STL-10 images are often complex, which can be recognized as an upgraded version of CIFAR-10 with higher resolution and more instances. It can better reflect real-world image diversity.

Figure 2: Images from STL-10 (Picture credit: Original).

- *CALTECH-101:* The CALTECH-101 dataset is more challenging rather the others, which includes 101 different object categories. These categories cover a wide variety of objects, including animals, food, tools, and more (Li et al, 2004). Fig. 3 shows some examples of part of the classes in this dataset, where the images have not been preprocessed. CALTECH-101 provides a rigorous test of object recognition algorithms because it requires the algorithms to be able to handle a variety of different classes of objects.

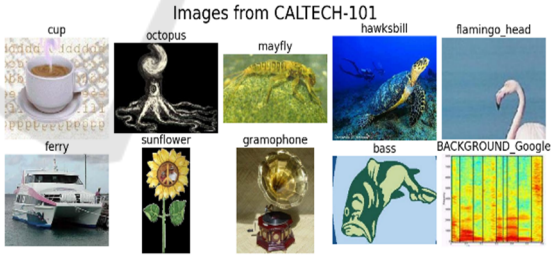

Figure 3: Images from CALTECH-101 (without preprocessing) (Picture credit: Original).

These dataset selections ensure that the comparative analysis encompasses a spectrum of image classification challenges, ranging from small, simple images to high-resolution, complex scenes, and diverse object categories.

2.4 Significance of Comparative Studies

Comparative analyses guide researchers and practitioners in selecting appropriate architectures for specific tasks. Such studies facilitate a deeper

understanding of architectural behavior across different scenarios, aiding informed decision-making. The insights gained from these analyses contribute to the design of more effective models and the advancement of the field.

3 METHODOLOGY

The methodology applied in this research underscores a meticulous and strategic approach to conducting a comprehensive and rigorous comparative analysis of image classification algorithms. The primary objective of this section is to delve into the key aspects of the methodology, highlighting the preprocessing steps and the rationale behind the selection of convolutional neural network (CNN) architectures. This strategic approach ensures holistic evaluation of algorithmic performance, taking into consideration both data preparation and architectural considerations.

3.1 Preprocessing

Not only the quality of data but also the quality of preprocessing can affect the performance of the model (Gulati and Raheja, 2021). To ensure rigorous evaluation, each dataset was divided into three subsets: training, validation, and test set with a ratio of 7:2:1. The training set was used for model parameter optimization, the validation set aided in hyperparameter tuning, and the test set, consisting of unseen data, served as the final performance evaluation metric. Table 1 shows how the datasets are split.

Table 1: Split of the datasets.

Dataset	Train set	Validation set	Test set
CIFAR-10	49,000	14,000	7,000
STL-10	9,100	2,600	1,300
CALTECH-101	6,400	18,28	916

Minimal resizing was performed on the images to retain their original dimensions. However, since the images in the CALTECH-101 dataset have varying sizes, they were uniformly resized to 224x224 pixels. For data augmentation, horizontal flipping was applied. Additionally, the images were normalized with 0.485, 0.456, 0.406, and 0.229, 0.224, 0.225 respectively for the mean and standard deviation values for the RGB channels. These values are from the famous ImageNet dataset.

3.2 CNN Architectures

Choosing the right CNN architectures is crucial for this study as it ensures a comprehensive evaluation of image classification algorithms at different developmental stages. Three significant architectures were selected that have played a pivotal role in computer vision. By comparing these architectures, the results on various aspects of image classification can be obtained clearer and more meaningful.

- *LeNet-5:* Introduced by LeCun et al. in 1998, LeNet-5 holds a paramount place in the history of convolutional neural networks (Lecun et al, 1998). Fig. 4. illustrates the classic architecture of LeNet-5. As one of the earliest CNNs, its "Convolutional + Sampling(Pooling) + Fully Connection" structure laid the foundation for modern image classification techniques (Lecun et al, 1998). LeNet-5's significance lies in its utilization of convolutional and subsampling layers, showcasing the effectiveness of hierarchical feature learning. Despite its relatively simple structure, its inclusion in this study allows for a benchmark evaluation of fundamental model performance.

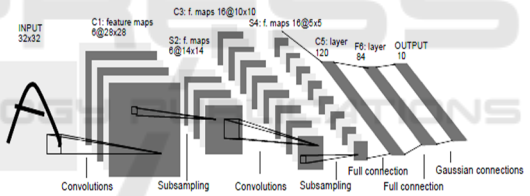

Figure 4: Architecture of LeNet-5.

- *VGG16:* The year 2014 witnessed the introduction of VGG16, a seminal architecture crafted by Simonyan and Zisserman. Renowned for its uniformity and depth, VGG16's distinctive trait is the consistent use of 3x3 convolutional filters across its layers (shown in fig. 5) (Simonyan and Zisserman, 2015) This architectural simplicity contributes to its ease of implementation and interpretation. Although resource-intensive due to its 16-layer depth, VGG16's deep structure empowers it to capture intricate features within images. Its selection in this study enables the exploration of the impact of architectural complexity on classification outcomes.

input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 5: Architecture of VGG16.

- *ResNet50:* Introduced by He et al. in 2015, ResNet50 represents a breakthrough in addressing the challenges posed by vanishing gradients in deep neural networks (shown in fig. 6) (He et al, 2016). This architecture's innovative use of skip connections allows for the training of remarkably deep networks without encountering diminishing gradient magnitudes. ResNet50's 50-layer depth showcases its ability to capture intricate image features while maintaining gradient flow. Its inclusion in this study offers insights into the advantages conferred by skip connections and depth in the field of image classification.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	[3×3, 64; 3×3, 64] ×2	[3×3, 64; 3×3, 64] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3
conv3_x	28×28	[3×3, 128; 3×3, 128] ×2	[3×3, 128; 3×3, 128] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×8
conv4_x	14×14	[3×3, 256; 3×3, 256] ×2	[3×3, 256; 3×3, 256] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×23	[1×1, 256; 3×3, 256; 1×1, 1024] ×36
conv5_x	7×7	[3×3, 512; 3×3, 512] ×2	[3×3, 512; 3×3, 512] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Figure 6: Architecture of ResNet50.

In conclusion, the methodology of this study strategically selects and explores the historical significance and unique characteristics of three distinct CNN architectures. These selections span pivotal moments in the timeline of deep learning, allowing for a comprehensive evaluation of image classification algorithms. The careful consideration of these architectures ensures a robust foundation for the ensuing comparative analysis.

4 EXPERIMENT

4.1 Dataset Selection Rationale and Training Settings

- *Dataset Selection:* The selection of datasets in this study is based on a deliberate strategy to encompass a range of challenges that progressively increase algorithmic demands. The details are shown in Table II. In detail, four factors were considered, number of classes, number of images, image size, and color channel. What all data sets have in common is the RGB color channel. Specifically, CIFAR-10 and STL-10 have limited categories, while the former has more instances with lower resolution. Moreover, the CALTECH-101 dataset is much more challenging where all the attributes are larger or higher, except the number of instances. This selection ensures a systematic evaluation of CNN architectures across varying complexities.

Table 2: Description of the datasets.

Dataset	# of classes	# of images	Image size(pixel)	Color channel
CIFAR-10	10	70,000	32x32	RGB
STL-10	10	130,000	96x96	RGB
CALTECH-101	101	9,144	variable (higher than 224x224 in average)	RGB

- *Training Settings:* The models were trained for 120 epochs using the stochastic gradient descent (SGD) optimizer with a momentum of 0.9. The criterion used for training was the cross-entropy loss function.

4.2 Experiment Results

The results section presents the performance of LeNet-5, VGG16, and ResNet50 across the three chosen datasets: CIFAR-10, CALTECH-101, and STL-10. The performance metrics are summarized in Fig. 7 providing a concise overview of each architecture's accuracy on each dataset.

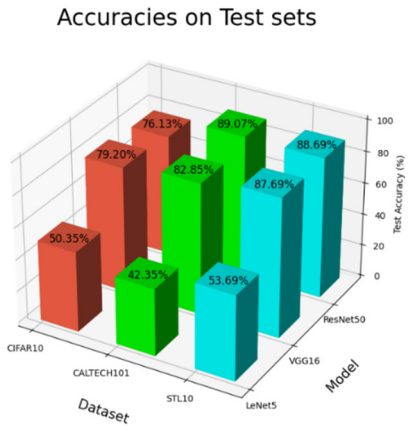

Figure 7: Accuracies on the test sets of the datasets (Picture credit: Original).

- *LeNet-5:* LeNet-5 exhibits the lowest accuracy on all datasets. The performance on VGG16 and ResNet50 is comparable, while ResNet50 shows better processing of more complex data sets. Furthermore, Since LeNet-5 is inherently designed for simpler tasks, its architecture is relatively shallower. Hence the under-performance of LeNet-5 which struggles to capture the nuances presented in more complex datasets is predictable and acceptable.
- *VGG16 and ResNet50:* It is evident that the more complex CNNs, VGG16 and ResNet50, outperform LeNet-5 across all datasets. Among all of them, ResNet50 achieves the highest accuracy on both CALTECH-101 and STL-10. These results underscore the significance of architectural complexity in enhancing image classification performance across diverse datasets. The advanced performance of VGG16 and ResNet50 on CALTECH-101 and STL-10 can be attributed to their deeper architectures, especially the skip connections of ResNet50, which allow them to capture intricate features in the diverse images present in these datasets. The complexity of CALTECH-101 and STL-10 aligns with ResNet50's strengths, while VGG16's consistent architecture is advantageous in maintaining a relatively good level in the face of a large number of instances. However, it is important to consider the computational demands of VGG16, particularly in resource-constrained scenarios. In this experiment, VGG16 always took the longest training time on all the datasets. In summary, the ability of these architectures to capture both low and high-level features is evident in their superior performance on CALTECH-101 and STL-10.

It is worth noting that the characteristics of a dataset can have a significant impact on the observed performance. In the case of CIFAR-10, CALTECH-101, and STL-10, the selection of datasets presents a set of challenges that reflect real-world scenarios. On the one hand, among all the CNNs, the accuracies on STL-10 are always the highest, which may be attributed to this dataset containing the largest number of instances with limited categories. On the other hand, the accuracies of the easiest dataset, CIFAR-10, on the more complex CNNs, VGG16 and ResNet50, are not the best as expected. This may point out that although more complex networks can adapt to more complex environments, they do not always perform very well in simple environments.

5 CONCLUSION

This study conducted a comprehensive analysis of image classification algorithms using diverse datasets: CIFAR-10, STL-10, and CALTECH-10, and CNN architectures: LeNet-5, VGG16, and ResNet50. The data set and network structure are carefully selected so that the results can cover a large scenario. This paper aims to figure out how architectural complexity impacts performance across varying datasets. Results highlighted the importance of architecture in addressing dataset challenges. Across the datasets, depth and skip connections were key. The depth of VGG16 and the application of skip connections in ResNet50 excelled on complex datasets, capturing intricate features. In conclusion, this study informs architectural decisions for diverse image classification scenarios, bridging CNN design and dataset specifics.

Future research can explore intricate designs, transfer learning, and hybrid models. These efforts will advance image classification, producing enhanced performance and generalization models.

REFERENCES

H. Yu, J. Zhao, and Y. Zhu, “Research on Face Recognition Method Based on Deep Learning,” IEEE Xplore, Oct. 01, 2019.

D. Bhatt et al., “CNN variants for Computer Vision: history, architecture, application, challenges and future scope,” Electronics, vol. 10, no. 20, p. 2470, Oct. 2021.

Y. Pei, Y. Huang, Q. Zou, X. Zhang, and S. Wang, “Effects of Image Degradation and Degradation Removal to CNN-based Image Classification,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2019.

Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

R. Liu, Y. Liu, Z. Wang, and H. Tian, "Research on face recognition technology based on an improved LeNet-5 system," IEEE Xplore, Jan. 01, 2022.

K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv.org, Apr. 10, 2015.

K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Thecvf.com, pp. 770–778, 2016.

A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems, vol. 25, 2012.

M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," proceedings.mlr.press, May 24, 2019.

A. Coates, A. Ng, and H. Lee, "An Analysis of Single-Layer Networks in Unsupervised Feature Learning," proceedings.mlr.press, Jun 2021.

Li Fei-Fei, R. Fergus, and P. Perona, "Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories," 2004 Conference on Computer Vision and Pattern Recognition Workshop.

V. Gulati and N. Raheja, "Efficiency Enhancement of Machine Learning Approaches through the Impact of Preprocessing Techniques," IEEE Xplore, Oct. 01, 2021.