Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/41/20230743

Performance analysis and comparison of cat and dog image classification based on different models

Dong Ma[1,2], Haoyang Song[1]

[1]School of AI and Advanced Computing/School of Advanced Technology, Xi'an Jiaotong-Liverpool University, TaiCang, China

[2]Dong.Ma22@student.xjtlu.edu.cn

Abstract. Image classification has widespread applications in computer vision, with significant advancements in performance due to deep learning models. Cat and dog image classification, as a classic problem, has attracted considerable research interest. This study aims to conduct a comprehensive analysis and comparison of deep learning models, including LeNet, ResNet, and VGG, in the context of cat and dog image classification. This paper employed two datasets: traditional cat and dog images and non-traditional, diverse images. Data preprocessing and augmentation were applied, and various model architectures were constructed. Through training and testing, this paper assessed the performance of these models under different conditions. The research findings indicate that ResNet excels in handling various datasets and different dataset sizes, demonstrating outstanding image classification performance. LeNet performs well on traditional datasets but experiences performance degradation when dealing with non-traditional datasets and smaller dataset sizes. VGG performs reasonably well on the original dataset but needs help processing non-traditional datasets. These results provide valuable insights for guiding model selection and optimization in image classification tasks.

Keywords: VGG, ResNet, LeNet, Image Classification.

1. Introduction

In today's digital world, image classification is crucial in computer vision. With the rapid advancement of deep learning techniques, the performance of deep learning models in image classification tasks has significantly improved. Cat and dog image classification, as a typical image classification problem, has attracted widespread research interest. Understanding the performance of different deep learning models in cat and dog image classification and how the diversity of datasets affects model performance is essential for advancing the field of computer vision. The objectives of this study are as follows: 1) Compare the performance of different deep learning models in cat and dog image classification tasks. 2) Explore the impact of dataset diversity on model performance. 3)Provide practical recommendations on how to select

This study analyzes and compares different deep learning models' performance in cat and dog image classification tasks. This paper uses three classical deep learning models (LeNet, ResNet, and VGG) to investigate their performance in handling different datasets and dataset sizes [1-3]. To achieve this goal, this paper Collects two different datasets, including traditional cat and dog images and non-traditional images. Besides, utilize data augmentation techniques to enhance model generalization and construct

DOI: 10.54254/2755-2721/41/20230743

different model architectures for the three deep learning models. Finally, this paper evaluates model performance through training and testing, comparing their performance under different conditions.

2. Methods

This paper utilized three prominent deep-learning architectures:

2.1. LeNet

LeNet, short for LeNet-5, is a pioneering convolutional neural network (CNN) architecture crucial in developing deep learning for image recognition [4]. It was initially proposed by Yann LeCun and his colleagues in the 1990s. LeNet is one of the foundational models that paved the way for modern CNNs and has been employed in various computer vision tasks. The LeNet architecture typically consists of multiple layers: convolutional, pooling, and fully connected. Here is a more detailed breakdown of its components: 1) Convolutional Layers: LeNet starts with one or more convolutional layers responsible for detecting various features in the input images. These layers use learnable filters (kernels) to perform convolution operations on the input data. 2) Activation Functions: Suitable activation functions, such as the hyperbolic tangent (tanh) or rectified linear unit (ReLU), are applied after each convolutional layer to introduce non-linearity into the network and capture complex patterns in the data. 3) Pooling Layers: Pooling layers often use max-pooling to downsample the feature maps produced by the convolutional layers. This helps reduce the spatial dimensions of the data while retaining important information [5]. 4) Fully Connected Layers: LeNet typically has one or more fully connected layers after several convolutional and pooling layers. These layers are used for high-level feature extraction and classification. LeNet has proven effective in tasks like handwritten digit recognition, making it a classic choice for image classification problems.

2.2. ResNet

ResNet, short for Residual Network, is a breakthrough CNN architecture designed to address the vanishing gradient problem in intense neural networks. It was introduced by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun in 2015. It has since become a go-to choose for various computer vision tasks due to its exceptional performance. Key features of the ResNet architecture include: 1) Residual Blocks: ResNet employs residual blocks, which are building blocks that include skip connections, also known as shortcut connections or identity mappings. These connections allow the network to learn residual functions, making it easier to train intense networks. 2) Deep Stacking: ResNet architectures can be extremely deep, with hundreds of layers. The skip connections in residual blocks help propagate gradients effectively during training, enabling the training of intense networks. 3) Convolutional Layers: Like LeNet and other CNNs, ResNet consists of convolutional layers with suitable activation functions. These layers extract hierarchical features from the input data. 4) Batch Normalization: Batch normalization is often used in ResNet to stabilize and accelerate the training process. It helps normalize the activations within each layer. ResNet has achieved state-of-the-art results in various image classification challenges, including the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [6].

2.3. VGG

The VGG (Visual Geometry Group) architecture is a well-known and influential CNN architecture for image classification and object recognition. The Visual Geometry Group developed it at the University of Oxford and was a runner-up in the ILSVRC 2014 competition. Key characteristics of the VGG architecture include: 1) Uniform Architecture: VGG follows a uniform architecture consisting of multiple convolutional layers with small 3x3 filters and max-pooling layers. The depth of the network is controlled by stacking these blocks. 2) Convolutional Layers: VGG uses convolutional layers with small receptive fields to capture fine-grained features in the input data. The 3x3 filters are applied with a stride of 1, which helps preserve spatial information. 3) Pooling Layers: Max-pooling layers reduce the spatial dimensions of feature maps while retaining important information. VGG uses max-pooling

DOI: 10.54254/2755-2721/41/20230743

with 2x2 windows and a stride of 2. 4) Fully Connected Layers: Like LeNet, VGG ends with one or more fully connected layers for high-level feature extraction and classification. 5) Activation Functions: Common activation functions like ReLU are used after each convolutional layer to introduce non-linearity into the network [7]. VGG is known for its simplicity and effectiveness. Although it may be more profound than ResNet, it achieved impressive results in image classification tasks and has served as a benchmark for other CNN architectures.

3. Experimental results and analysis

3.1. Data collection

This paper meticulously gathered data from two distinct sources, enriching the study with diverse images. The primary datasets used for this research are as follows: 1) Original Cat-Dog Image Dataset. The initial dataset comprises a total of 100,000 traditional cat and dog images. These images were carefully curated from well-established online repositories, ensuring they were numerous and relevant to the study's objectives. Each image was selected with precision to maintain high data quality. 2) Non-Traditional Image Dataset. The second dataset is a unique collection of 40,000 images encompassing various content, including artworks, abstract visuals, and unconventional depictions of cats and dogs [8]. This dataset was meticulously curated to introduce diversity and challenge deep learning models with non-standard image content, going beyond the conventional representations of these animals.

3.2. Data pre-processing

Critical pre-processing steps were meticulously applied before the data could be fed into deep learning models to ensure data consistency and reliability: 1) Resizing and Normalization. All images underwent a resizing process to conform to a uniform dimension, typically set at 224x224 pixels. This standardization was imperative to ensure compatibility with the architectural requirements of deep learning models. Additionally, pixel values within each image were meticulously normalized, scaling them to fall from 0 to 1. This normalization process aimed to mitigate potential issues related to variations in pixel intensity. 2) Data Quality Checks. Rigorous data quality checks were conducted to identify and rectify any potential issues within the datasets. This included the detection of corrupted images and the verification of accurate labeling. Any problematic images discovered during this process were removed from the dataset or corrected to maintain data integrity. This step was pivotal in ensuring that the data used for training and evaluation were of the highest quality.

3.3. Data augmentation

This paper implemented various data augmentation techniques. These techniques introduced controlled variations into the training data, thus making models more robust to real-world variations: 1) Image Rotation and Flipping. During training, images were subjected to random rotations and horizontal flips. This augmentation strategy exposed the models to a broader range of image orientations and improved their ability to handle images with varying perspectives. It was precious in scenarios where the orientation of cats and dogs in images could vary significantly.[9] 2) Brightness and Contrast Adjustments. To simulate different lighting conditions, this paper introduced controlled variations in brightness and contrast across the dataset. This augmentation technique enhanced the adaptability of our models, ensuring that they could effectively handle images taken under diverse lighting circumstances. By applying these adjustments, models were better equipped to generalize from the training data to real-world scenarios [10]. By incorporating these comprehensive data collection, pre-processing, and augmentation procedures, this paper aimed to establish a robust foundation to thoroughly evaluate the performance of different deep learning models in cat and dog image classification tasks under various conditions and scenarios.

DOI: 10.54254/2755-2721/41/20230743

3.4. Results and analysis

When trained and tested on the original cat and dog image dataset, LeNet exhibited good classification accuracy. Training accuracy increased gradually with the number of epochs, while validation accuracy showed an upward trend. Training loss decreased gradually, and validation loss showed an overall downward trend. Performance decreased when using the unconventional image dataset, but it still maintained a certain level of accuracy. LeNet's classification accuracy gradually decreased with a decrease in dataset size, but it still demonstrated acceptable performance at 1/5 of the dataset size. ResNet showed the best classification performance on the original cat and dog image dataset. Training accuracy was positively correlated with the number of training epochs, and validation accuracy and loss showed more fluctuations than other models. ResNet's performance remained relatively good when using the unconventional image dataset but slightly lower than the original dataset. Reducing dataset size had a minimal impact on ResNet's performance. Even at 1/5 of the dataset size, it maintained excellent performance (Table 1).

Table 1. Experimental results.

	Accuracy
LeNet	0.7456
ResNet	0.9292
VGG	0.7272

VGG performed reasonably well on the original cat and dog image dataset but could have been better than ResNet. Performance significantly declined when using the unconventional image dataset, especially at smaller dataset sizes, such as 1/5, where performance dropped noticeably. Based on experimental results, this paper finds that: 1) ResNet demonstrates strong performance in cat and dog image classification across different types of datasets and dataset sizes. 2) LeNet performs well when dataset sizes are more significant but suffers performance degradation with unconventional and smaller datasets. 3) VGG exhibits reasonable performance on the original cat and dog image dataset but experiences a significant drop in performance with unconventional data sets and smaller dataset sizes.

4. Discussion

This paper has observed variations in the performance of different models across distinct datasets and dataset sizes. These disparities can be attributed to the design and complexity of the model architectures. For instance, ResNet's deep architecture and residual connections empower it to handle complex and diverse data effectively. In contrast, LeNet's shallower structure and relative simplicity diminish performance, mainly when dealing with non-traditional datasets and smaller dataset sizes. The introduction of non-traditional image data posed a challenge to model performance. This dataset comprised artworks, abstract images, and unconventional cat-dog depictions, which exhibited more significant visual variations than traditional datasets. Consequently, model performance suffered when faced with these images due to their increased visual diversity. This paper recommends choosing models based on the specific task requirements and dataset characteristics. For conventional cat-dog image classification tasks, ResNet remains a reliable choice. However, more complex models or additional data pre-processing and augmentation may be necessary to achieve optimal performance when dealing with non-traditional datasets. Data Augmentation. Data augmentation played a pivotal role in enhancing model performance. This paper encourages further exploration of advanced data augmentation techniques to bolster model robustness, especially when confronted with non-traditional datasets.

5. Conclusion

This paper embarked on a comprehensive journey to analyze and compare the performance of three distinct deep learning models, LeNet, ResNet, and VGG, within the cat and dog image classification tasks. This paper conducted rigorous experiments across varying datasets and dataset sizes to gain profound insights into the capabilities of these models. this paper finds that: 1) ResNet stands out as a

DOI: 10.54254/2755-2721/41/20230743

star performer in experiments, showcasing remarkable image classification capabilities across a spectrum of dataset types and sizes. Its deep architecture and ingenious residual connections enable it to gracefully navigate the complexities of traditional and unconventional image datasets. 2) LeNet excels when dataset sizes are substantial, consistently demonstrating strong classification performance. 3) VGG, while not reaching the heights of ResNet, serves as a dependable baseline model for cat and dog image classification on traditional datasets. Its straightforward architecture provides a solid foundation for comparison with more complex models.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[2] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[3] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[4] Szytula A and Leciejewicz J 1989 Handbook on the Physics and Chemistry of Rare Earths vol 12, ed K A Gschneidner Jr and L Erwin (Amsterdam: Elsevier) p 133. O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458.

[5] Agrawal, S. C., Tripathi, R. K., Bhardwaj, N., & Parashar, P. (2023, July). Virtual Drawing: An Air Paint Application. In 2023 2nd International Conference on Edge Computing and Applications (ICECAA) (pp. 971-975). IEEE.

[6] Yasin, E. T., Ozkan, I. A., & Koklu, M. (2023). Detection of fish freshness using artificial intelligence methods. European Food Research and Technology, 1-12.

[7] Cai, D. (2017). Fine-grained classification of low-resolution image (Master's thesis).

[8] Howard, J., & Gugger, S. (2020). Deep Learning for Coders with fastai and PyTorch. O'Reilly Media.

[9] Alomar, K., Aysel, H. I., & Cai, X. (2023). Data augmentation in classification and segmentation: A survey and new strategies. Journal of Imaging, 9(2), 46.

[10] Maurya, L., Lohchab, V., Mahapatra, P. K., & Abonyi, J. (2022). Contrast and brightness balance in image enhancement using Cuckoo Search-optimized image fusion. Journal of King Saud University-Computer and Information Sciences, 34(9), 7247-7258.