

Research on Image Recognition Problems Based on Deep Learning Models

Ao Liu

Computer Science and Technology, Tianjin University of Technology, Tianjin, China

liuao@stud.tjut.edu.cn

Abstract. Deep learning has achieved impressive results in computer vision but typically requires extensive datasets, which are often costly and difficult to obtain, particularly in specialized domains like medical imaging and autonomous driving. To address this challenge, various image augmentation algorithms have been introduced, aimed at enhancing the size and diversity of training datasets. While model based techniques make use of picture production models, model-free techniques employ conventional image processing techniques. The goal of optimizing policy-based techniques is to enhance model performance by determining the optimal set of augmentation procedures. The seminal AlexNet model, introduced in 2012, marked a significant breakthrough in image classification, demonstrating the power of deep learning. Through advanced techniques such as multi-scale convolution, cross-connections, and global average pooling, these approaches have substantially improved image classification accuracy. Despite ongoing challenges, these advancements underscore the critical role of data augmentation and optimized neural network architectures in enhancing performance when data is limited.

Keywords: Alexnet, Flower, Image Recognition, Image Augmentation.

1. Introduction

Despite the satisfactory performance of deep learning in computer vision, its effectiveness typically demands an abundance of pictures for training. However gathering a sufficiently large and diverse set of images can often be expensive and challenging, particularly in specialized fields such as medical imaging or autonomous driving. Numerous picture augmentation algorithms have been developed to address this problem. Comprehending these current algorithms is crucial to determining appropriate techniques and creating innovative strategies for particular assignments. Using a novel informative taxonomy, this work provides an extensive assessment of picture augmentation for deep learning. The work presents problems in the vicinity distribution of computer vision to investigate the fundamental goal of picture augmentation. The model-free category employs image processing techniques, whereas the model-based approach uses image creation models to generate images. On the other hand, the policy-based optimization method seeks to identify the best possible combination of procedures. Based on this analysis, the study broadens the knowledge required to choose appropriate techniques and create original algorithms.

Alex Krizhevshy created the deep learning model known as AlexNet in 2012. When compared with the conventional classification method, AlexNet exhibits significant advantages. Over 80% of millions of Image Net data had higher recognition rates than 70%. With enough training data, DCNNs can facilitate the learning of layered hierarchical image characteristics that are highly representative and data-driven [1]. The authors of Anwar took advantage of three significant, but little-known, aspects of using DCNNs for computer-aided detection issues [2]. Lofe and Szegedy considered deep convolution networks for large-scale photo recognition. Hinton used the AlexNet algorithm for image identification and took part in the Picture Net picture classification competition in the area of deep learning [3].

On numerous Computer Vision tasks, deep convolution neural networks have demonstrated exceptionally good performance. But to prevent this, these networks rely a lot on big data. describes the phenomena that occurs when a network learns a highly variable function, such as modeling the



training set of data accurately. Sadly, many application disciplines, including medical image analysis, lack access to huge data. The survey's main focus is on data augmentation, a data-space approach to the limited data issue series of techniques referred to as "data augmentation" aim to increase the number and quality of training datasets. So that more advanced models can be constructed with them [4].

The research object for the study is the AlexNet network. The data-space resolution for the restricted data problem that is the focus of this research is data augmentation.

2. Methods:

2.1. Data set (Source, Description, Reasons for Selection)

The data set downloaded from CSDN contains images of flowers, which has been widely used in training image classification tasks, achieving excellent results. This data consists of images of five types of flowers, all in jpg format. Sunflowers, tulips, daisies, dandelion, and roses are the five kinds of flowers. The total quantity of samples is 3670, with 633 daisy images, 898 dandelion images, 641 rose images, 699 sunflower images, and 799 tulip images. There are two sets of data: a training set consisting of 3306 samples and a validation set consisting of 364 samples.

2.2. Model (Concept, Function, Advantages, and Disadvantages)

AlexNet Model:

AlexNet is a neural network (CNN) architecture proposed by Alex Krizhevskiy in 2012 for image classification tasks. AlexNet performed exceptionally well in the ILSVRC 2012 competition, significantly improving image classification accuracy and marking a breakthrough for deep learning in computer vision.

Important AlexNet layers:

AlexNet's architecture comprises three pooling layers and five convolution layers.

While pooling layers shrink feature maps while preserving crucial information, convolution layers extract local features from images.

Activation Function: AlexNet uses the ReLU (Rectified Linear Unit) activation function, which accelerates the training process.

Fully Connected Layers: Three fully connected layers make up AlexNet, with the last fully connected layer being utilized to produce categorization results.

Dropout Layer: Dropout technique is used in the fully connected layers to reduce

Functions of AlexNet:

Image Classification: AlexNet excels in large-scale image classification tasks, such as those involving the Image Net data.

Feature Extraction: It extracts high-level features from images through deep layers.

Technological Advancements: AlexNet introduced ReLU activation function, dropout techniques, and GPU-accelerated training, laying the foundation for subsequent deep learning models.

Advantages of AlexNet:

High Performance: It performs exceptionally well in image classification tasks, significantly improving classification accuracy.

ReLU Activation Function: Quicker than conventional Sigmoid or Tanh activation functions in terms of training speed.

Dropout Technique: Effectively reduce, enhancing the model's generalization capability.

GPU Acceleration: Utilizes GPU for parallel computation, greatly speeding up the training process.

Disadvantages of AlexNet:

High Computational Resource Demand: AlexNet requires a large amount of computational resources, especially for large-scale datasets, resulting in longer training times.

3. Result

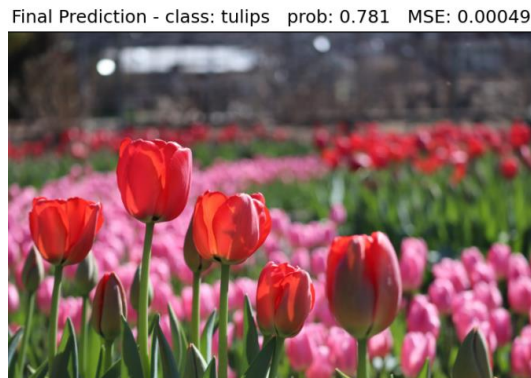


Fig. 1 The enhanced image, and its MSE is 0.00049 (Photo/Picture credit: Original).



Fig. 2 The enhanced image, and its MSE is 0.000734 (Photo/Picture credit: Original).



Fig. 3 The improved image, and its MSE is 9.16e-05 (Photo/Picture credit: Original).

Final Prediction - class: roses prob: 0.443 MSE: 0.000867



Fig. 4 The improved image, and its MSE is 0.000867 (Photo/Picture credit: Original).

From Figs 1-4, it can be seen that the model performance is moderate, it still has room for improvement. This may be because the images in the test set do not clearly express the characteristics of these flowers, leading to inaccurate predictions by the model. Through the Mean Squared Error (MSE), it can see that the MSE after image augmentation is within the standard range, indicating that image augmentation has played a certain role in making the predictions more accurate.

4. Discussion

Convolution neural networks (CNNs) learn by directly extracting image features, enabling the acquisition of higher-level, more abstract feature representations. This approach allows the extraction of more essential features from images, resulting in more accurate and generalized learned features. Zhou et al. proposed a multi-scale convolution neural network (CNN), which involves using two or more different convolution scales as inputs to the CNN model [5]. Sub-CNNs in the multi-scale CNN benefit from input images of various scales to learn optimized parameters. However, this method has numerous parameters and presents higher training difficulty [6]. Using manually derived features, the AlexNet network lowered the mistake rate in the 2012 Image Net competition by about 10% when compared to traditional methods. This proves that when it comes to image recognition and classification, CNNs are far superior to manual feature extraction techniques [7]. It was created to classify pictures for the Image Net LSFRC-2010 competition, which it triumphed in [8]. For the parameter explosion and overfitting phenomenon generated across connections, optimizing the network structure by pooling can increase the robustness and overfitting resistance of the network model, thus improving the speed and accuracy of network training [9]. Inspired by GoogLeNet, this paper proposes using multi-scale convolution to extract features from flower images. This approach allows for capturing feature information at different scales, obtaining richer image features, and thereby more accurately expressing the characteristics of the input data [10].

5. Conclusion

By incorporating multi-scale convolution, cross-connections, and global average pooling into the AlexNet model, the study used multiple scales of convolution kernels to extract distinct attributes and adjusted to small-sized fruit photos using small-scale convolution kernels. These methods have led to significant improvements and optimizations:

Multi-Scale Convolution: Images of fruit can be used to extract a range of features using different convolution kernel sizes. Larger kernels are great for capturing the overall image, while smaller kernels are perfect for picking up finer details.

Cross-Connections: Combining low-level and high-level feature information across layers allows for a more accurate and complete transfer of features. This method enhances the efficacious utilization and dissemination of feature information by the network.

Global Average Pooling: This technique reduces the number of parameters, avoiding the issue of parameter explosion and making the network easier to train. By averaging each feature map globally, it creates more representative feature vectors, keeping the network's expressive power while cutting down on parameters.

The experiments show that the improved AlexNet algorithm significantly boosts the accuracy of flower image classification. However, some specific flowers still pose challenges in recognition accuracy. Future work will involve traditional feature extraction combined with image features to achieve parameter sharing during training, further enhancing classification accuracy.

Flower image classification using the AlexNet model has made remarkable strides. Since its debut in 2012, AlexNet has dramatically improved image classification accuracy with its deep convolution neural network structure. For flower classification, AlexNet achieved efficient and precise results through extensive data training. The optimized model and expanded datasets have enhanced its performance in complex scenarios. In real-time applications, integration with mobile devices allows users to instantly identify flowers, improving user experience. Looking ahead, AlexNet will play a greater role in smart agriculture, education, and environmental protection, driving technological progress.

The AlexNet model is widely used in flower classification, significantly boosting efficiency in various fields. In botany research, it speeds up the identification of flower species; in agriculture and horticulture, it helps farmers and gardeners manage plants more precisely; in education and science popularization, it makes learning about flowers easier; in the flower market and e-commerce, it improves product classification and customer service efficiency.

The future looks bright for the AlexNet model in flower classification. With ongoing optimization, it will recognize flowers more accurately and quickly. Smart agriculture, mobile applications, and the education sector will increasingly benefit from this technology, enhancing agricultural efficiency, spreading plant knowledge, and supporting ecological and environmental management.

References

- [1] Zixian W, Shuping H. Improvement of the AlexNet Networks for Large-Scale Recognition Applications Received: 6 May 2019/Accepted: 26 October 2020/Published online:7 November 2020 Shiraz University 2020.
- [2] Anwar S, Hwang K, Sung W. Fixed point optimization of deep convolutional neural networks for object recognition.In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2022, 1131-1135.
- [3] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Proceedings of the International Conference on Machine Learning, 2015, 448-456.
- [4] Connor S, Taghi M. A survey on Image data augmentation for deep learning. Journal of Big Data, 2019, 6: 60.
- [5] Yang X, Shang ZH. Expression recognition based on improved. AlexNet Laser&Optoelectronics Progress, 2020, 57(14).
- [6] Zhou S, Liang YY, Wan J, et al. Facial expression recognition based on multi-scale CNN. Biometric Recognition. 2016, 503-510.
- [7] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 2012, 25(2): 1097-1105.
- [8] Gao R, Wang R, Feng L, et al. Dual-branch, efficient, channel attention-based crop disease identification. Comput. Electron. Agric. 2021, 190, 106410.
- [9] Zhou S, Liang YY, Wan J, et al. Facial expression recognition based on multi-scale CNNs. Biometric Recognition. Cham: Springer International Publishing, 2016, 503-510.
- [10] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, 1-9.