Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 11, No. 3, 2024

Image Feature Selection based on Attention Mechanism

Zuyong Lu *

College of Artificial Intelligence, Automation, Nanjing Agricultural University, Nanjing, Jiangsu, 210031, China
*** Corresponding author Email:** 9213020510@stu.njau.edu.cn

Abstract: In the field of deep learning, the selection and extraction of image features are the key factors affecting model performance. Traditional image feature selection methods often rely on artificially designed features, which is not only time-consuming but also difficult to capture complex patterns in the image. In recent years, attention mechanism, as a technique that enables models to automatically focus on key parts of input data, has shown significant advantages in many fields such as natural language processing and image recognition. In this paper, an attention-mechanism-based image feature selection method is proposed to improve the accuracy and efficiency of image classification and object detection tasks. First, we introduce the basic principles of the attention mechanism, and then we design a convolutional neural network (CNN) framework with integrated attention modules that can adaptively adjust the weights during training to highlight important areas in the image and ignore irrelevant backgrounds. By introducing the attention mechanism, our model can learn the key features in the image more effectively, reduce the waste of computing resources, and improve the generalization ability of the model. Finally, we verify the reliability of the model on several data sets.

Keywords: Attention; Neural Network; Feature Extraction.

1. Introduction

In the wide field of deep learning, the selection and extraction of image features is one of the core factors affecting model performance. With the development of technology, the field has gradually shifted from traditional manual feature design to automated feature learning. However, while automated feature learning methods, particularly convolutional neural networks, have made significant progress in tasks such as image recognition, classification, and detection, these methods often rely on large amounts of data and computational resources, and in some cases still struggle to capture complex patterns and subtle differences in images. Therefore, how to effectively select and extract the key image features helpful to the task has become the key to improve the performance of the model.

In this research context, attention mechanism as a new technology, the core idea is derived from the working principle of the human visual system, that is, when processing a large amount of information, it can automatically filter and focus on the key information. This mechanism allows the model to dynamically adjust the attention of different regions or features when processing data, thus achieving remarkable results in many fields such as natural language processing, speech recognition, and image recognition. The attention mechanism not only improves the efficiency of the model in processing information, but also enhances the ability of the model to capture important features in the data, enabling the model to make more accurate and robust predictions in the complex data environment [1].

In view of the great potential of attention mechanism in improving model performance, a novel image feature selection method based on attention mechanism is proposed in this paper. This method aims to optimize the feature learning ability of convolutional neural networks by introducing attention mechanisms, so as to improve the accuracy and efficiency of image classification and object detection tasks [2]. We first introduce the fundamentals of the attention mechanism and detail how it helps the model focus on key areas in the image. Next, we design and implement a CNN framework with an integrated attention module, which adaptively adjusts the weights to highlight important features and suppress irrelevant background, thus making the model more focused on learning meaningful image information.

Through this innovative approach, our model can not only learn the key features in the image more effectively, but also reduce the waste of computing resources and improve the generalization ability of the model. To verify the validity of the proposed method, we conducted extensive experiments on multiple public datasets and compared them with traditional methods [3,4]. The experimental results demonstrate the superiority of our method in improving the performance of image processing tasks.

In addition, we further explore the effect of attention mechanism on different types of image data, including remote sensing image, medical image and natural scene image. These experimental results show that our approach can effectively improve the performance of the model in both high resolution and low light conditions, and in both simple and complex scenes. This is particularly important because it shows that our approach has strong generalization and adaptability, and can meet various challenges in practical application scenarios.

To sum up, the research in this paper not only provides a new solution for image feature selection and extraction, but also opens up new possibilities for the research and application of deep learning. We believe that as technology continues to advance and improve, attention mechanisms will play an even more important role in image processing and computer vision tasks in the future.

2. Convolutional Neural Network

In the fully connected neural network (Fully Connected Neural Network, FC), it is assumed that the number of neurons at the n layer is n^n and the number of neurons at the $n-1$ layer is $nn-1$. Then $n^{n-1}\times n^n$ connection edges are required to connect the $n-1$ layer with the n layer, that is, the number of parameters contained in the weight matrix is

$n^{n-1}\times n^n$. When the number of neurons is large, the number of parameters in the weight matrix will be large, resulting in a decrease in the training efficiency of the neural network [5,6]. In the convolutional layer, the input z^n of layer n is the convolution of the activity value a^{n-1} of layer $n-1$ and the filter, as follows:

$$z^n = w^n \otimes a^{n-1} + b^n \tag{1}$$

Where the filterw^n is the weight vector and b^n is the bias quantity.

Because of local connections, each neuron in the convolutional layer (we assume layer $n-1$) is connected only to certain neurons in the next layer (layer n), which are identified by the local window. Compared with the fully connected layer, the number of connections between the convolutional layer and the next layer is greatly reduced. The number of connections between the original$n^{n-1}\times n^n$ is reduced to $n^{n-1}\times m$. The size of filters in the convolutional layer is expressed as m. Because of the weight sharing, filters can act on all neurons together. As shown in Figure 1, connections of the same color all have the same weight. However, in weight sharing, one filter can only extract one local feature, so in order to extract multiple features, it is often necessary to use multiple filters in the convolution layer.

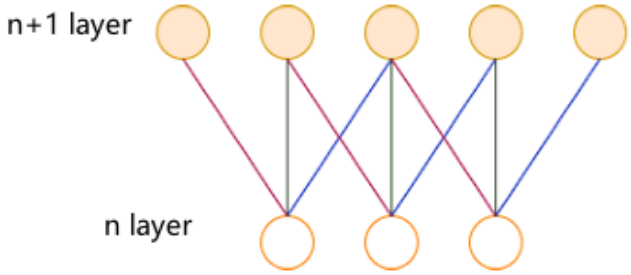

Figure 1. Convolution layer scheme

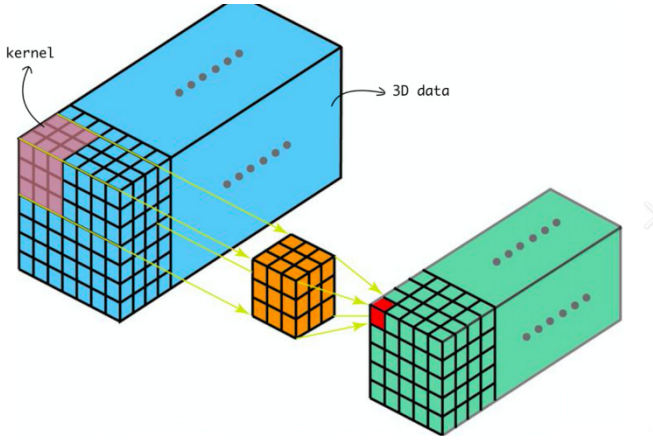

Figure 2. The three-dimensional structure of the convolution layer

Convolutional layer (convolutional layer) is the core infrastructure of convolutional neural networks [7]. The function of convolution is to extract a local feature of the input data. Every convolution kernel in the convolution layer is a feature extractor, so features can be extracted fully. At present, convolutional neural networks are mainly applied in image processing, while images belong to two-dimensional data in data processing. Therefore, in order to fully extract image data features, the convolutional layer needs to be reconstructed into a three-dimensional structure with the dimensions $M\times N\times D$, where M is the width, N is the width and D is the depth. Figure 2 shows the three-dimensional structure of the convolutional layer. In FIG. 2, D M×N feature maps together constitute the basic structure of the convolutional layer. Feature mapping (Feature Map) is the feature extracted from the image after product operation, and each individual feature mapping represents a feature of the image. Therefore, it is necessary to use multiple feature maps in each convolutional layer to improve the feature representation capability of the network.

One-dimensional convolution was originally used in signal processing, often to calculate the delay accumulation of various signals. We assume that a one-dimensional convolution of the input signal sequence x $x1$, $x2$, ···, xt, convolution layer of convolution kernels or filter $\omega\omega 1,\omega 2,\cdots,\omega m$, then filter and convolution sequence of the input signal as shown in type 2:

$$y_t = \sum_{k=1}^{m} \omega_k \cdot x_{t-k+1} \tag{2}$$

The one-dimensional convolution of the signal sequence x and the ω of the filter is defined as follows:

$$y = \omega * x \tag{3}$$

Where $*$ is the convolution operator.

In image processing, two-dimensional convolution is more commonly used than one-dimensional convolution because the image data itself is a two-dimensional structure. Two-dimensional convolution is the extension for one-dimensional convolution, assume that the input image data for $X \in \mathbb{R}$ M x N, convolution layer filter for $W \in \mathbb{R}$ m x n, m and n is greater than M N and its convolution is:

$$y_{i,j} = \sum_{u=1}^{m}\sum_{v=1}^{n} \omega_{uv} \cdot x_{i-u+1,j-v+1} \tag{4}$$

An example of two-dimensional convolution is shown in Figure 3. The input is a 4×4 two-dimensional image, and a 2×2 feature map is obtained by convolution operation with a convolution kernel of 3×3. The specific operation is as follows: on the feature plane of the input image data, the convolutional kernel uses step size 1 to slide on the feature plane in sequence from left to right and from top to bottom, and then performs weighted sum to obtain the final feature mapping.

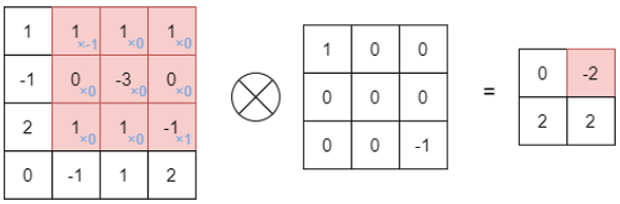

Figure 3. Two-dimensional convolution diagram

3. Attention Mechanisms

The essence of attention mechanism comes from human visual attention, which was first proposed in the field of visual image. The basic idea of the attention mechanism is to choose from input information that is more critical to the objective of the task at hand. According to the importance of each input information, it assigns a corresponding weight, so as to capture more valuable information. With the continuous development and improvement of attention mechanism, attention mechanism has gradually become a research hotspot of scholars around the world, and has been widely used in image recognition, speech recognition, machine translation and other tasks, and has achieved good results [8]. In 2017,

the google team used self-attention mechanisms to learn text representations and achieved good learning results [9]. Self-attention mechanism is an improvement of attention mechanism, which is less dependent on external information and better at capturing internal correlation of data. Its structure is shown in Figure 4. Attention mechanism.

As shown in FIG. 4, $h_t$represents the implicit state vector at moment t, $e_t$represents the similarity weight of h_t, α_t represents the normalized similarity weight of h_t, and R represents the self-attention output matrix obtained by weighted summation. The calculation process can be divided into three steps as follows:

(1) The multi-layer perceptron is used to calculate the similarity weight of the data at each moment, and the calculation formula is as follows:

$$e_t = W_b \tanh(W_a[h_t; H]) \tag{5}$$

Among them, the W a said the MLP connection weights between input and hidden layer, $W_b$said MLP hidden layer and output layer connection weights between H= (h_{t-1}, h_t,...,h_{t+i}) represents the hidden state matrix.

(2) The SoftMax function is used to normalize these similarity weights, and the calculation formula is as follows:

$$a_t = soft\max(e_t) = \frac{exp(e_t)}{\sum_{j=t-1}^{t+i}(e_j)} \tag{6}$$

Where exp() represents an exponential function based on the natural constant e.

(3) The weighted sum of the normalized similarity weight and the corresponding data is carried out to obtain the self-attention output matrix R. The calculation formula is as follows:

$$R = \sum_{j=t-1}^{d} \alpha_j \overline{h}_j \tag{7}$$

Where, $h_j$represents the implicit state vector at time j, and $\alpha_j$represents the similarity weight after normalization of h_j.

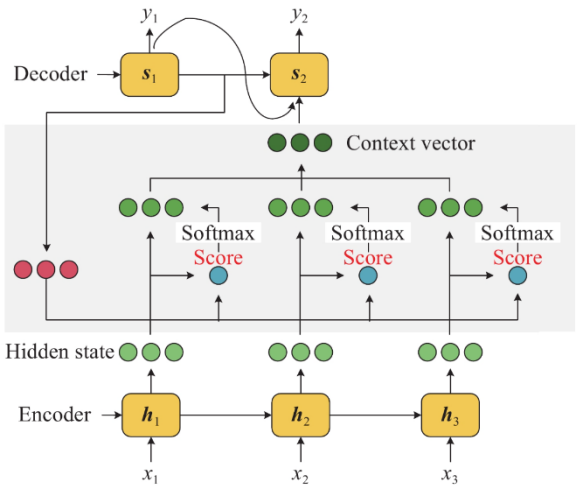

Figure 4. Self-attention mechanism

4. Experiment

In this paper, the performance of the method is verified on the ImageNet dataset, and this experiment is implemented based on PyTorch. All deep convolutional neural network models in the experiment follow the same architecture, and the initial learning rate is set as lr = 6 × 10-1.

The learning rate is gradually adjusted according to the number of iterations.

Table 1. Experimental results of different models

Model	Macro	Micro
CNN	78.5%	77.9%
ResNet	81.3%	81.7%
Our	83.8%	84.2%

According to the experimental results in Table I, the three models -- CNN, ResNet and Our model -- show obvious differences in performance under the two evaluation indexes "Macro" and "Micro". The "Macro" indicator reflects the average performance of the model across various categories, while the "Micro" indicator measures the overall performance of the model across the overall data set. Specifically, the accuracy of the CNN model on the "Macro" and "Micro" indexes is 78.5% and 77.9%, respectively, showing its basic level in the multi-classification task. The accuracy of ResNet model in the two indexes increased to 81.3% and 81.7%, indicating that the generalization ability and prediction accuracy of the model were effectively improved through innovative design such as residual connection. On the other hand, Our model has reached a high level of 83.8% and 84.2% respectively in the "Macro" and "Micro" indexes, which not only means that Our model has excellent average performance in various categories, but also the accuracy of the overall sample prediction is extremely outstanding, surpassing the CNN and ResNet models, showing a strong performance advantage.

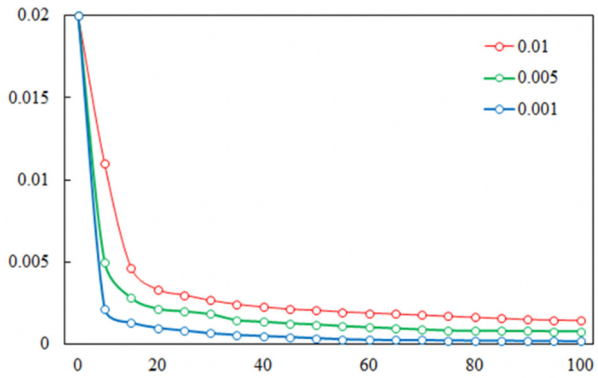

Figure 5. The influence of different learning rate on model loss

Figure 5 shows the different learning rates. The chart shows the changes of the loss function in the training process for three different learning rates (0.01, 0.005 and 0.001 respectively). As the number of iterations increases, each curve has its own manifestation: The learning rate is 0.01 (red curve): It declines rapidly at the beginning, but due to the high learning rate, there are large fluctuations later, which may mean that the model is difficult to converge. The learning rate is 0.005 (green curve): compared to the red curve, it declines slightly more slowly, but behaves more smoothly in the later period, with no significant oscillations. The learning rate is 0.001 (blue curve): Although the initial decline was slow, it eventually approached the minimum and maintained good stability. To sum up, it is very important to select a suitable learning rate for model training. Too high or too low learning rate may affect the model performance.

5. Conclusion

In this paper, attention mechanism is innovatively applied to image feature selection, which solves the limitation of

manual feature design in traditional methods. The proposed CNN framework based on attention mechanism effectively highlights the key information in the image by adaptive weight adjustment, and ignores the redundant background, thus improving the accuracy and efficiency of image classification and object detection tasks. The experimental results show that the model shows excellent performance on multiple datasets, which proves the remarkable advantages of attention mechanism in the field of deep learning image processing, and provides new directions and ideas for subsequent research.

References

[1] Muruganantham, Priyanga, et al. "A systematic literature review on crop yield prediction with deep learning and remote sensing." Remote Sensing 14.9 (2022): 1990.

[2] Xu, Yonghao, and Pedram Ghamisi. "Universal adversarial examples in remote sensing: Methodology and benchmark." IEEE Transactions on Geoscience and Remote Sensing 60 (2022): 1-15.

[3] KATTENBORN, Teja, et al. Spatially autocorrelated training and validation samples inflate performance assessment of convolutional neural networks. ISPRS Open Journal of Photogrammetry and Remote Sensing, 2022, 5: 100018.

[4] Graves, Alex, and Jürgen Schmidhuber. "Framewise phoneme classification with bidirectional LSTM and other neural network architectures." Neural networks 18.5-6 (2005): 602-610.

[5] Zhu, Jiawei, et al. "AST-GCN: Attribute-augmented spatiotemporal graph convolutional network for traffic forecasting." IEEE Access 9 (2021): 35973-35983.

[6] Wen, Guangqi, et al. "MVS-GCN: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis." Computers in Biology and Medicine 142 (2022): 105239.

[7] Hou, Jialu, Hang Wei, and Bin Liu. "iPiDA-GCN: Identification of piRNA-disease associations based on Graph Convolutional Network." PLOS Computational Biology 18.10 (2022): e1010671.

[8] Eliasof, Moshe, Eldad Haber, and Eran Treister. "Pde-gcn: Novel architectures for graph neural networks motivated by partial differential equations." Advances in neural information processing systems 34 (2021): 3836-3849.

[9] Peng, Shaowen, Kazunari Sugiyama, and Tsunenori Mine. "SVD-GCN: A simplified graph convolution paradigm for recommendation." Proceedings of the 31st ACM International Conference on Information & Knowledge Management. 2022.