An Image Classification Algorithm Based on Multidomain Convolution Neural Network

Tao-wei JIANG, Meng-yu ZHU*, Yong-hao HAI and De-zhuang KONG
Schools of Life Science, Beijing Institute of Technology, Beijing China
*Corresponding author

Keywords: Convolutional Neural Network, Deep learning, Image classification.

Abstract. Deep Convolutional Neural Networks (CNNs) have outperformed humans in many computer vision tasks, such as object recognition and image classification, but it is almost impossible to run a large-scale CNN network structure in the platform and application scenarios with limited calculate ability. For the limited computing platform and application scenarios, we propose a novel CNN architecture: Multi-Domain CNN (MD-CNN). Multi-domain images are obtained by different pre-processing of the images input to the multi-domain convolutional neural network, and each image domain independently obtains the output features through a feedforward network. Then, the output features of multiple parallel multi-domain images are spliced together as the output characteristic of MD-CNN. This structure can effectively convert the network of CNN from depth to width, extract more efficient features quickly, which improves the training speed of the network. In the test of MNIST and CIFAR-10 datasets, the accuracy of our method reached 96.3% and 91.4% respectively, moreover, compared with the traditional CNN training speed increased about 1.5 times.

Introduction

Image classification is one of the basic and challenging tasks in computer vision. In recent years, the deep learning[1][2] algorithm based on convolutional neural network has made outstanding achievements in the traditional image classification tasks. Deep convolutional neural network is a model of deep learning, and the improvement of algorithm’s effect depends on the depth of neural network and a large amount of training data. Embedded vision applications such as: Security surveillance, UAV and so on, facing the limited calculate ability and real-time challenges. In order to improve the timeliness of convolutional neural networks, it is necessary to design a new network architecture.

In 1988, LeCun[3] et al. proposed CNN-based artificial neural network structure, which effectively solves the weakness of extracting image features by hand and it’s automatically extracting the advanced abstract features of the images. With the development of deep artificial neural network technology, deep neural network model based on CNN has made great progress in the image classification task. In 2012, AlexNet[4] proposed a large multi-layer CNN-based image recognition algorithm, AlexNet, which achieved a high recognition rate on the ImageNet dataset. Then there are many deep neural network models based on CNN proposed and achieved a breakthrough in the application of computer vision such as image classification and image identification. VGG-Net[5] appeared in 2014, another GoogleLeNet[6] in 2015, ResNet[7] and other network architectures were proposed. Although the above architecture has its own unique, but they all have a common characteristic, that is, are deep large neural networks.

The above algorithms have achieved good results in the image classification tasks, but they are all driven by big data, which leads to the slow training of the network model, numerous network model parameters, and difficulty in adjusting the algorithm performance and other issues. Rely on deepening CNN neural network layer to enhance the effectiveness of the algorithm has encountered a bottleneck. In view of the above problems, this paper proposes a multi-domain CNN-based image classification algorithm. First of all, different preprocessing of the original image to get multi-domain image.
Secondly, a multi-CNN architecture is proposed to deal with multi-domain images. Finally, by comparison experiment, the algorithm is compared with the traditional CNN to verify the effectiveness of the algorithm.

**Algorithm Architecture**

This paper proposes a CNN network structure that connects multiple image domains at the same time, which can quickly learn the effective features of input images. The feedforward network of each image domain is similar to the traditional CNN network, and both convolutional layers and pooling are alternately connected. However, the networks of each image domain are parallel structures. By converting the depth network to a certain extent to the width, the parameters of the model to be trained are greatly reduced, thus speeding up the training. As shown in Figure 1, a convolutional neural network with 3 convolution layers, which have 5 convolution kernels of 5×5 size for each output feature map, then there are $5^9$ parameters need to be trained. If we change to two multi-CNN architectures with two convolution layers, the parameters to be trained are $2 \times 5^6$. The parameters to be trained are greatly reduced.

![Figure 1. Parameters compare between single CNN and MD-CNN.](image)

The deeper the neural network, the stronger its ability to learn image features. Although we reduce the number of layers in the MD-CNN, it is still a multi-layer structure and at the same time we solve the performance loss caused by reducing the network depth by introducing multi-domain images. The performance of image classification algorithms is highly dependent on the image feature representation, so efficient image classification requires multiple image fields, which can be color, image gradient, and different pixel filter responses. There is almost no correlation between these image regions, but they contain complementary image features [8]. Therefore, we need to consider these image fields at the same time, and then we introduce these image fields separately.

**Multi-domain Image**

**Image Enhancement Based on Mean and Standard Deviation.** Let the mean value of image is $\mu$, the standard deviation of the image gray is $\sigma$, then the adaptive enhancement algorithm[9][10] is:

$$f(i, j) = k \cdot \frac{\mu}{\sigma} (x(i, j) - m_s(i, j)) + m_s(i, j)$$  \hspace{1cm} (1)

Where $k$ is a positive constant and $m_s(i, j)$ is the average grayscale value of the local area centered on the pixel $f(i, j)$. In this paper, the selected data sets are MNIST (28×28) and CIFAR-10 (32×32). Considering that the image is not large, our image normalization algorithm regards the entire image as a partial image:

$$f(i, j) = k \cdot \frac{\mu}{\sigma} (x(i, j) - \mu) + \mu$$  \hspace{1cm} (2)

**Local Normalization Based on Mean and Standard Deviation.** Let the mean value of image is $\mu$, the standard deviation of the image gray is $\sigma$, then the image local normalization algorithm[11][12] is:
\[ f(i, j) = \frac{x(i, j) - \mu(i, j)}{\sigma(i, j) + C} \]  
(3)

Where \( C \) is a constant, here 1/255, \( x(i, j) \) denotes a pixel, \( \mu(i, j) \) and \( \sigma(i, j) \) the gray mean and standard deviation of local blocks.

**Gradient Image**

Based on the Sobel\[13\][14] operator gradient image, the Sobel operator calculates the horizontal and vertical partial derivatives \( S_x \) and \( S_y \) in the 3x3 domain using the pixel \( f(i, j) \) as neutral. The gradient is:

\[ g(i, j) = \sqrt{S_x^2 + S_y^2} \]  
(4)

Operator template is:

\[
\begin{bmatrix}
-1 & 0 & 1 \\
-2 & 0 & 2 \\
-1 & 0 & 1 \\
\end{bmatrix}
\]

In this paper, we choose three image regions generated by a given grayscale image, which are the enhanced image, the normalized image and the gradient image after local normalization respectively, as shown in Figure 2. In CNN, enhanced image can enhance the image detail feature expression, and the normalized image can not only reduce the saturation problem, but also make the CNN have the robustness to the illumination change. On the other hand, the gradient image contains the shape information of the target, which is a good feature of the normalized gray image.

![Original image, Domain-1, Domain-2, Domain-3](image)

**Figure 2. Multi-domain image.**

By connecting the final responses of these three image domains together, we get a fully connected layer that outputs an inferred vector.

**Multi-domain Neural Network Architecture**

First introduce the CNN structure of a single image domain. Then combine multiple CNNs to construct our multi-domain CNN architecture. Our single image domain CNN structure contains two convolution layers, followed by an average pooling operation for each convolutional layer, and the sigmoid function\[15\] is used as the activation function. Figure 3 shows a single image domain CNN structure, the input is a specific image domain that after the original image pre-processed. The first convolutional layer uses a convolution kernel function of size 5x5 to output eight feature maps, followed by a pooling operation to reduce the resulting feature mapping to a low dimensional level. The second convolutional layer also uses a 5x5 convolution kernel function to output 16 feature maps. After the pooling operation, the output feature vector is obtained.
The multi-domain neural network is formed by combining the neural networks of three single image domains horizontally, as shown in Figure 4. The original image is preprocessed to obtain images of three image domains: Image Domain-1, Image Domain-2, and Image Domain-3. One-dimensional eigenvectors of three single image domains are combined to form a fully connected layer. The fully connected layer is a logical regression operation.

**Experiment**

In order to verify the validity of the multi-domain convolutional neural network proposed in this paper, we compare the two experiments with the traditional CNN in the MNIST and CIFAR-10 databases to verify the effectiveness of the proposed algorithm and the rapidity of training. We verify the ability of our algorithm to learn effective features by setting the dimensionality of the final output of the network in experiments and comparing the accuracy of different algorithms. By setting the same feature output dimension and comparing the training time of the algorithm, we can verify the rapidity of algorithm training in this paper. The experimental environment is CPU 3.5GHz, memory 16GB, 64-bit Windows 10 operating system, and the algorithm implementation environment is Matlab (2016a).

Table 1 records the accuracy of the traditional CNN and MD-CNN when the number of training iterations is set to 5000 and the output feature dimensions take different values. As can be seen from Table 1, MD-CNN has stronger feature abstraction ability than CNN. The correct rate of MD_CNN is always higher than that of traditional CNN in the same output feature dimension, especially when the output feature dimension is low the correct rate of CNN declines rapidly, while the performance of MD-CNN in this paper is better. It shows that MD_CNN can learn more efficient features by adding multi-domain images to ensure the accuracy of classification results. Table 2 is to set the algorithm output characteristic dimension is 240, compare the running time of different iteration times. It can be seen from Table 2 that MD-CNN is easier to train than CNN with less time and speed by about 50% at the same number of iterations.
Table 1. Comparison of feature learning ability.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Output characteristic dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MNIST</td>
<td>CIFAR-10</td>
</tr>
<tr>
<td>CNN</td>
<td>52.1%</td>
<td>50.2% 32</td>
</tr>
<tr>
<td></td>
<td>78.7%</td>
<td>69.1% 128</td>
</tr>
<tr>
<td></td>
<td>92.4%</td>
<td>80.6% 240</td>
</tr>
<tr>
<td>MD-CNN</td>
<td>78.5%</td>
<td>76.3% 32</td>
</tr>
<tr>
<td></td>
<td>86.9%</td>
<td>80.5% 128</td>
</tr>
<tr>
<td></td>
<td>96.3%</td>
<td>91.4% 240</td>
</tr>
</tbody>
</table>

Table 2. Comparison of training speed.

<table>
<thead>
<tr>
<th></th>
<th>Training time (minutes: m)</th>
<th>The number of iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CNN</td>
<td>MD-CNN</td>
</tr>
<tr>
<td>MNIST</td>
<td>32m</td>
<td>20m</td>
</tr>
<tr>
<td></td>
<td>138m</td>
<td>92m</td>
</tr>
<tr>
<td>CIFAR-10</td>
<td>32m</td>
<td>21m</td>
</tr>
<tr>
<td></td>
<td>143m</td>
<td>96m</td>
</tr>
</tbody>
</table>

**Conclusion**

Multi-domain image based multi-domain convolutional neural network algorithm proposed in this paper can enhance the original image's feature information based on the traditional convolutional neural network, so that our algorithm can learn more balanced and comprehensive original image features. Enhance the ability of neural network to deal with the changes of ambient lighting conditions, shooting angle changes and other scenes. Compared with the traditional CNN, experimental results show that this algorithm can effectively learn the characteristics of the image and can quickly train.

**Acknowledgement**

This research was financially supported by the National High Technology Research and Development Program of China (2015AA042308).

**References**


