

基于异形卷积核的卷积神经网络图像分类方法

马 双, 董安国, 王长鹏

长安大学理学院, 陕西 西安

Email: 535675651@qq.com

收稿日期: 2020年10月28日; 录用日期: 2020年11月12日; 发布日期: 2020年11月19日

摘 要

卷积神经网络具有强大的图像特征学习能力, 在机器学习问题中取得了突破性的进展。针对目前卷积神经网络方法中的不足之处, 创新性提出了异形卷积核的概念, 在不改变卷积核参数数量的情况下通过改变卷积核的形状, 扩大其感受野, 提高网络提取图像特征的能力。通过图像移位的方法, 解决了异形卷积核可行性问题。构建了一种融合异形卷积核和矩形卷积核的卷积神经网络, 使得异形卷积核能够与传统卷积神经网络有效结合。实验结果表明, 相较于传统的卷积神经网络, 结合异形卷积核的卷积神经网络具有更高的分类精度。

关键词

卷积神经网络, 卷积核, 异形卷积核, 图像分类, 感受野

Based on the Heterogeneous Convolution Kernel Image Classification of Convolutional Neural Network

Shaung Ma, Anguo Dong, Changpeng Wang

College of Faculty of Science, Chang'an University, Xi'an Shaanxi

Email: 535675651@qq.com

Received: Oct. 28th, 2020; accepted: Nov. 12th, 2020; published: Nov. 19th, 2020

Abstract

With a strong ability to learn image features, convolution neural network has made a breakthrough in machine learning. Directing at the shortcomings of the current convolution neural

network methods, the concept of heterogeneous convolution kernel is creatively proposed by changing the shape of the convolution kernel, expanding its receptive field and improving the ability of the network to extract image features, without changing the number of convolution kernel parameters. Through the method of image shift, the feasibility of heterogeneous convolution kernel is solved. The convolution neural network which combines the heterogeneous convolution kernel and the rectangular convolution kernel is constructed in order that the heterogeneous convolution kernel can be effectively combined with the traditional convolution neural network. The experimental results show that, compared with the traditional convolution neural network, the convolution neural network combined with special-shaped convolution kernel has higher classification accuracy.

Keywords

Convolutional Neural Network, Convolution Kernel, Heterogeneous Convolution Kernel, Image Classification, Receptive Field

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

如今图像的特征提取与分类已经成为计算机视觉中最重要的研究领域之一，卷积神经网络(Convolutional Neural Network, CNN)更是依赖其良好的特征提取能力和泛化能力，在其领域获得了巨大成功。卷积神经网络的快速发展一方面得益于计算机性能的大幅提升，使得构建并训练更大规模的网络不再受到硬件水平的限制；另一方面得益于大规模标注数据的增长，增强了网络的泛化能力 Lecun 等[1]提出 Le Net-5 模型后更加积淀了卷积神经网络在图像处理领域的重要性。继 Lecun 之后，人们又对卷积神经网络的结构和内部方法做了许多改动和优化，比如 Krizhevsky 等[2]提出的 Alex Net 中使用了 Relu 激活函数来代替 Sidmorg 函数，并首次引入 Doupout 来增强网络的泛化能力。在这之后又提出了 VGGNet-16 [3]、GoogLeNet [4]、ResNet-50 [5]等优秀的网络架构，不仅增加了网络深度和准确率，还解决了网络过拟合等问题。2016 年 Huang [6]等提出了一种 DenseNet 模型缓解了消失梯度问题，加强了特征传播，鼓励特征重用，并大大减少了参数的数量。此外 Hinton [7]等人在 2017 年提出胶囊网络(Capsule Network, CapsNet)更是改变了人们对卷积神经网络的认识，解决了 CNN 对物体之间的空间辨识度差及物体大幅度旋转之后识别能力低下的两个缺陷。

虽然以上网络对卷积神经网络做了诸多创新，但对卷积这一过程几乎没有改变，基本只是停留在对卷积核大小的改变以及卷积核与前一层的链接方式的改变上。卷积层作为卷积神经网络的主要部分，完成了对特征的提取从而达到更好的分类效果。通过研究发现卷积层的浅层用来提取边缘和背景等轮廓特征可以看作是边缘检测器，深层网络用来提取花纹等细节和更抽象特征[8]，但是现在几乎所有的卷积核都为矩形，若想使其更好地提取图像特征，从而提高网络精确度，只能通过改变卷积核的大小来实现，而卷积核大小的改变会使其参数迅速增长，不利于神经网络的计算与求解[9] [10]。这使得卷积核不能更好地发挥其作用，那么如何使卷积核能够更好地提取目标特征便成为一个重要的问题。本文以 $3 * 3$ 的卷积核为例，通过对卷积核形状的改变使得卷积核的感受野发生变化，从而使得其对于某种特定的特征具有更好的提取效果，使得卷积核更加高效地提取图片特征以至于增加卷积神经网络的特征提取能力以及

分类正确率[11] [12]。

2. 卷积神经网络(CNN)

首先给出卷积神经网络的结构以及卷积核的概念。

卷积神经网络通常包括卷积层、池化层、全连接层和输出层，其组合方式一般为卷积层和池化层的交替，最后是全连接层和输出层。如图 1 所示：

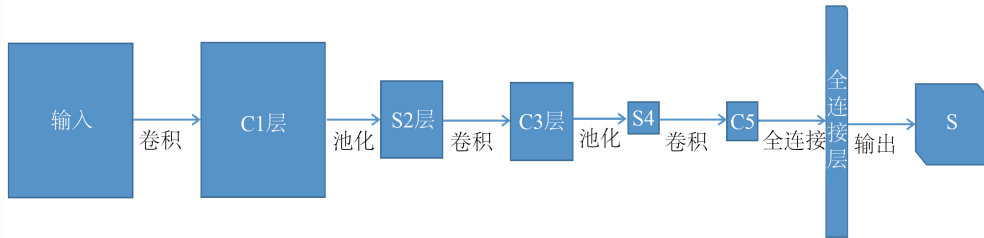


Figure 1. CNN network structure diagram
图 1. CNN 网络结构图

卷积神经网络的参数是通过最小化代价函数 $C(\omega, b)$ 来得到，即

$$\min_{w, b} C(\omega, b) = \frac{1}{2n} \sum_{x \in X} \|y(x) - S(w, b, x)\|^2, \quad (1)$$

其中 X 为所有样本的集合， n 为样本容量， $y(x)$ 为样本 $x \in X$ 的监督值， K 为卷积层个数，第 k 个卷积层的输入和输出单元数分别为 I_k 和 J_k ， w_j^k 为第 k 个卷积层中第 j 个输出单元所对应的卷积核， b_j^k 为第 k 个卷积层中第 j 个输出单元所对应的偏置量 $k=1, 2, \dots, K$ ， $w = \{w_j^k \mid j=1, 2, \dots, I_k; k=1, 2, \dots, K\}$ ， $b = \{b_j^k \mid j=1, 2, \dots, I_k; k=1, 2, \dots, K\}$ ，为各个卷积核的偏置量， $S(w, b, x)$ 为 CNN 网络当输入为 x 时的输出。通过最小化代价函数来求解 w, b 从而得到完整的网络。由于卷积核中的参数与偏置量是优化模型(1)的最优解，所以，从理论层面上分析，卷积核越大，效果越好，解释如下：由于高阶卷积核可以通过低

阶卷积核外围补 0 得到，如：

$$\begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \Rightarrow \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & b_{11} & b_{12} & b_{13} & 0 \\ 0 & b_{21} & b_{22} & b_{23} & 0 \\ 0 & b_{31} & b_{32} & b_{33} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

所以，最优的三阶卷积核置于五阶卷积核范畴中比较，其仅仅是个五阶可行解，或者可以理解为五阶卷积核当其周围全为零时的条件最优解。从而，CNN 中卷积核阶数越高，网络精度越高。但是，由于模型(1)的求解是近似的，其精度受到参数个数及算法本身的影响[13]，高阶卷积核的参数维度以平方级递增，对求解的精度和效率有严重影响，也不利于模型深度的增加，为了避免高阶卷积核的这种负面影响，实际使用时常常放弃高阶卷积核在理论上的优势，选用低阶卷积核。

3. 异形卷积核

基于以上的分析，我们提出异形卷积核的概念。在不提高参数维度的情况下，拓宽了卷积核的感受野，使得低阶卷积核能在一定程度上拥有高阶卷积核的优势，从而提高 CNN 的分类正确率。下面我们我们以 3×3 的卷积块为例， 3×3 的几种异形卷积核如图 2 所示：

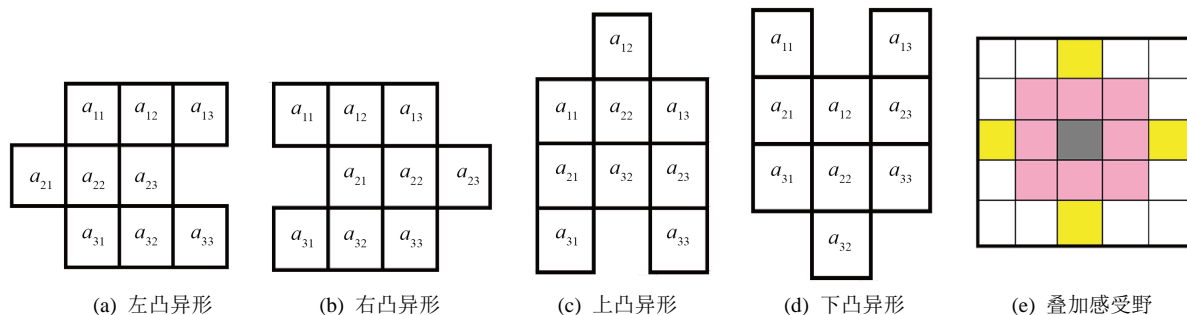


Figure 2. Shaped convolution kernel and its receptive field
图 2. 异形卷积核及其感受野

如图 2(e)所示将几种异形卷积核叠加在一起，可以扩展卷积核的感受野。从结构上看相比于矩形卷积核，异形卷积核也有利于对角特征的描述。

由上图 2 所示的卷积核的变化方式，那么还存在四种阶梯型的卷积核，若将这四种卷积核应用于卷积神经网络中只能改变其卷积的计算逻辑，无疑会增加网络的计算量。而提出异性卷积核的目的是为了在不增加网计算难度和参数数量的情况下改变其卷积核的感受野，所以这四种不适用与本文所提出的网络结构，所以不再讨论范围内。

4. 融合异形卷积核和矩形卷积核的卷积神经网络

4.1. 异形卷积核的实现

以下任以 3 * 3 的卷积核为例。

假设 $A = (a_{i,j})_{3 \times 3}$ 为 3 * 3 的标准卷积核， $B = (b_{i,j})_{m \times n}$ 为图像， $C = (c_{i,j})_{m \times n}$ 为 A 卷积 B 后的图像，则 $c_{i,j} = \sum_{s=1}^3 \sum_{t=1}^3 (\alpha_{s,t} \times b_{i+s-2, j+t-2})$ ，(2) 其中 $i = 1, 2, \dots, m$ ， $j = 1, 2, \dots, n$ ，(这里规定： $i + s - 2 < 1$ ， $i + s - 2 > m$ 或 $j + s - 2 < 1$ ， $j + s - 2 > n$ 时 $b_{i+s-2, j+t-2} = 0$) 对于如图 2(a)所示的异形卷积核(其元素如图中标记)记为 \bar{A} ，令

$$\bar{c}_{i,j} = \sum_{s=1}^3 \sum_{t=1}^3 (\alpha_{s,t} \times \bar{b}_{i+s-2, j+t-2}) \tag{3}$$

$$\bar{b}_{i,j} = \begin{cases} b_{i,j} & i = 2k + 1, k = 1, 2, \dots, m - 1, j = 1, 2, \dots, n \\ b_{i,j-1} & i = 2k, k = 1, 2, \dots, m, j = 1, 2, \dots, n \end{cases} \tag{4}$$

其中 $i = 1, 2, \dots, m$ ， $j = 1, 2, \dots, n$ 若记， $\bar{B} = (\bar{b}_{i,j})_{m \times n}$ ， $\bar{C} = (\bar{c}_{i,j})_{m-2 \times n-2}$ ，则 \bar{C} 的偶数行与 \bar{A} 与图片 B 卷积后的图片偶数行相等。(我们把 \bar{B} 称为矩阵 B 的“移位矩阵”)。

分析表明，仅看偶数行 \bar{A} 与 B 卷等于 A 与 \bar{B} 的卷积。

如果删除矩阵 \bar{B} 的第一行像素记作 $\bar{B}1$ ，标准 3 * 3 卷积核 A 与矩阵 $\bar{B}1$ 进行步长为 2 的卷积，就等于图 2(b)卷积核与矩阵 B 做步长为 2 的卷积。设 P 为监督样本图像集合， $P1 = \{X|X \text{ 是 } Y \text{ 经公式(4)的变化所得图像, } Y \text{ 属于 } P\}$ ，对 $P2 = \{X|X \text{ 是 } Y \text{ 删除第一行并在最后一行补 0 得到的图像, } Y \text{ 属于 } P1\}$ 。通过对于一般的图像，删除第一行像素后(最后一行补 0)，其特征信息几乎不会变化，在获取图像特征时 P1 和 P2 几乎等价，通过深度学习得到的卷积核应该近似相同。基于上述的分析，以 P1 作为监督样本，利用 CNN 进行学习(卷积时变步长为 1)，就相当于同时用到了图 2(a)，图 2(b)所示的卷积核。同理可以用相同的思路处理图 2(c)，图 2(d)的卷积核，也就是对图像做纵向的移位，然后使用矩形卷积核进行卷积。这样可

以通过对于图片的变化使得在使用同样的卷积核的时候使其感受野增加，对于图像的改变也使得异形卷积核在 CNN 中更加容易实现。异形卷积核的实现如图 3 所示。

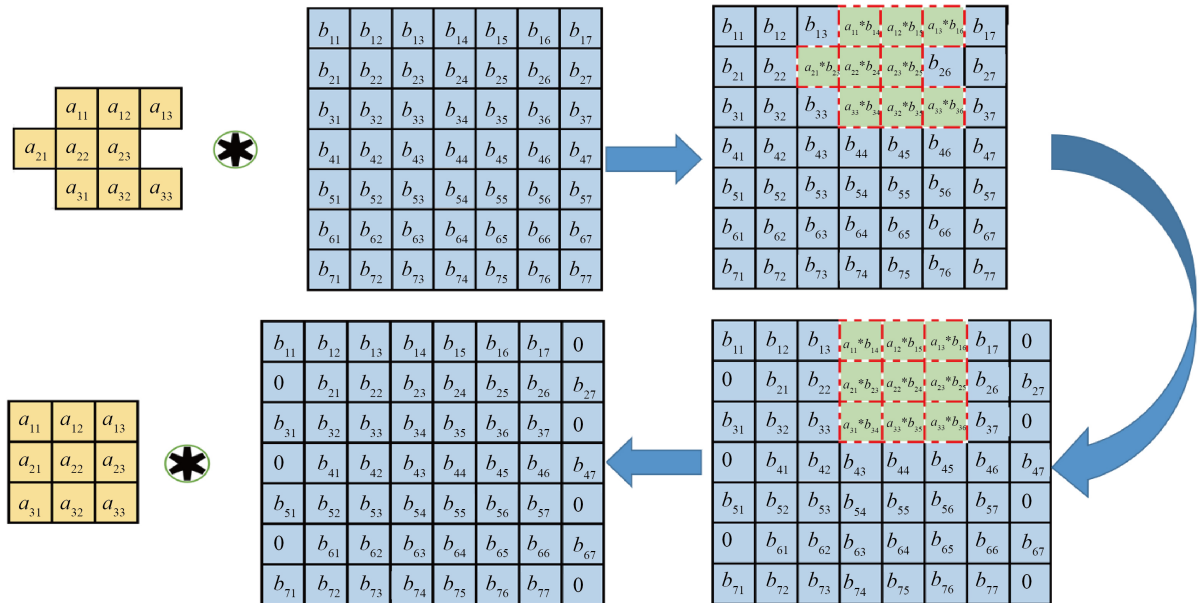


Figure 3. The realization of the special-shaped convolution kernel
图 3. 异形卷积核的实现

4.2. 融合移位图片的卷积神经网络

基于以上的分析将移位的图片输入到 CNN 中进行学习，会使得卷积神经网络第一层的卷积核变为异形卷积核，从而提高网络第一层的感受野。但是每一种移位方式(纵向或横向)只能对应训练一个网络。且通过实验发现在使用不同的移位方式训练出来的网络，会使某些类的分类精度变高而另一些类的分类精度变小，这样使得总体分类精度并没有发生变化甚至降低，所以我们提出一种联合两种异形卷积核的卷积神经网络的分类模型。这个模型中我们需要建立 3 个独立的神经网络，分别将以不同方移位的图片放入相同的网络结构中进行学习，最后将 softmax 层的概率相加和再进行分类。这样将两个网络的分类概率相加后可以使得其分到正确的类别。虽然计算量增大且网络参数变多但是可以大幅提高网络分类精度。网络结构如图 4 所示(注意网络训练的时候是分开训练的):

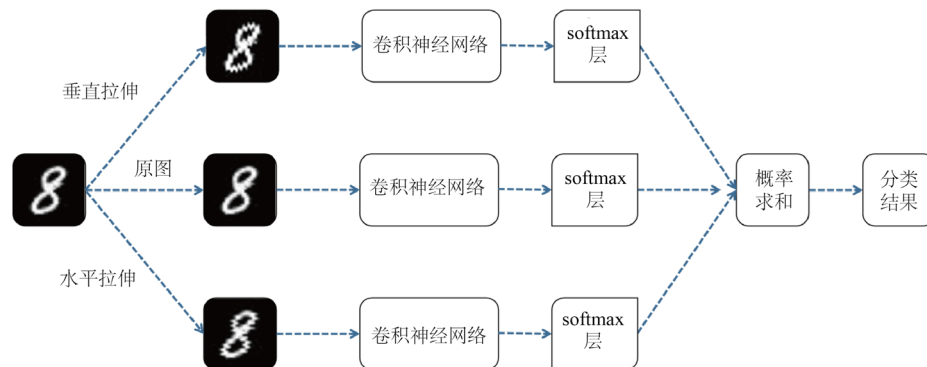


Figure 4. Converged network structure
图 4. 融合网络结构

融合网络需要对所有的训练样本图片做移位操作(横向和纵向)。分别用正常图片和移位后的图片训练各自的网络,这样可以训练出3个独立的CNN网络,将训练好的3个网络的参数保留,在其输入层前加入对图片做移位操作的函数(原图训练的网络不加),然后将其输出层舍去,把softmax分类概率相加求和构成新的输出层。这样构成一个新的网络,这个网络实际包含了三个相互独立的CNN,它们的输入是将同一张图片做不同的移位操作而得到的,所以每个独立的网络输入其实各不相同。

5. 实验

本文在MNIST、Fashion MNIST和CIFAR-10三个数据集上进行仿真实验。MNIST数据集来自美国国家标准与技术研究所National Institute of Standards and Technology(NIST)。共70,000个手写数字图片大小为 28×28 ,其中训练集60,000个,测试集10,000个。Fashion MNIST是MNIST数据集的升级版,两者格式大小相同,Fashion MNIST更具有挑战性。CIFAR-10数据集由10个类的60,000个 32×32 彩色图像组成,每个类有6000个图像,有50,000个训练图像和10,000个测试图像。使用MATLAB深度学习架构进行实验分析,为证明异形卷积核相对于传统卷积核的优势所在。构建了两种简单卷积神经网络。如图1和图5所示。

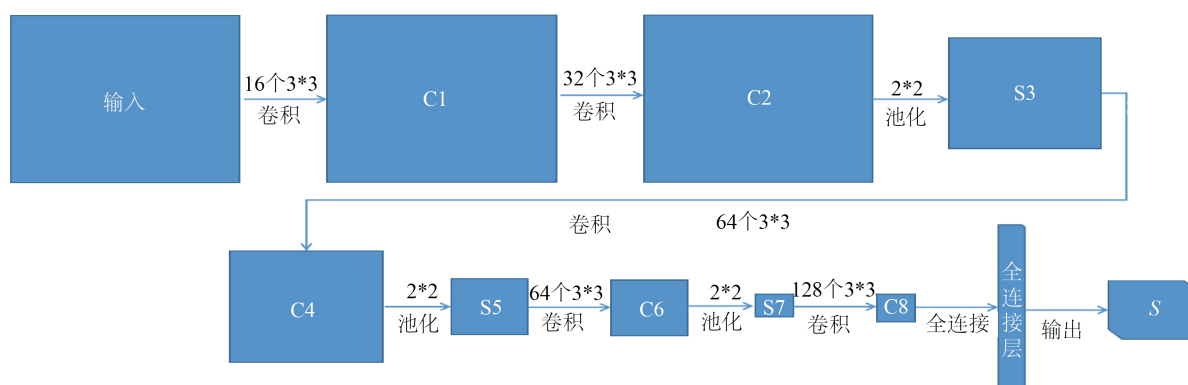


Figure 5. Network structure

图5. 网络结构

分别用移位的图片(横向和纵向)和正常图片分别训练三个神经网络(使用图1所示的网络结构,卷积层第一层用8个 3×3 的卷积块,第二层用16个 3×3 的卷积块,第三层16个 3×3 的卷积块,池化层采用 2×2 步长为2的最大池化[14]),训练过程使用具有动量的随机梯度下降法[11][12]训练网络,初始学习率为0.1。我们先取MNIST每个数字的训练集的百分之一(小样本训练集可以更加清晰看到不同卷积核对于不同的特征敏感度),在迭代到正确率基本不发生变化时停止,重复100次平均后得到以下结果:

如表1所示,在小样本训练集上可以看到正常的图片训练出的网络对数字1、4的识别正确率更高,横向拉伸图片对数字9、6识别率更高,纵向拉伸图片对7、3的识别率更高。虽然总的正确率差距不大,但是对不同的数字不同的卷积核表现出了不同正确率,若将不同的卷积核的优势融合起来便会提高网络总体的正确率。

所以使用第四章所说的方法,用水平移位、垂直移位、正常的三种图片分别训练三个网络,将各网络的输出层删除,再将softmax层的概率相加求和除以3然后进行分类(这里也可以使用加权求和的方式给每个方向不一样的权重,本文三个方向权重相同)。在训练网络前对每一张图片的最外层都补了一圈0,这样使得移位后的图片和原图片大小相同。CIFAR-10数据集使用图5网络结构,MNIST和Fashion MNIST数据集使用图1网络结构(图一所示网络卷积层第一层用16个 5×5 的卷积块,第二层用32个 3×3 的卷

积块, 第三层 64 个 3×3 的卷积块, 池化层采用 2×2 步长为 2 的最大池化), 如图所示每一层卷积层和全连接层后会加一个归一化层加快训练速度, 训练过程使用具有动量的随机梯度下降法训练网络, 初始学习率为 0.01。在迭代到正确率基本不发生变化时停止。结果如表 2 所示。

Table 1. Experimental data

表 1. 实验数据

训练图像类型	正常	横向拉伸	纵向拉伸
总正确率	84.43%	85.16%	83.82%
数字 0	93.99%	94.53%	92.54%
数字 1	95.26%	93.90%	92.75%
数字 2	83.29%	85.10%	86.35%
数字 3	81.80%	82.82%	84.16%
数字 4	88.19%	80.76%	86.87%
数字 5	81.26%	84.80%	77.34%
数字 6	91.45%	92.38%	87.58%
数字 7	81.59%	80.45%	83.87%
数字 8	74.17%	73.32%	67.97%
数字 9	72.02%	82.59%	76.84%

Table 2. Experimental results

表 2. 实验结果

数据集	不同图片训练网络的正确率%			
	正常	水平移位	垂直移位	融合网络
MNIST	99.33	99.27	99.30	99.41
Fashion-MNIST	88.31	88.07	87.67	89.63
CIFAR-10	76.90	75.62	75.39	80.30

如表 2 所示通过 MNIST、Fashion MNIST 和 CIFAR-10 数据集的仿真实验证明, 联合了正常图像和拉伸图像训练出网络的 softmax 层, 融合 CNN 网络的分类正确率有较为明显的提升。其次, 经过对比发现, 采用正常图片训练所得的网络精度越小, 其融合网络对于其精度的提升就越大。所以融合网络对于在 CNN 表现不好的数据集中有着更大的使用前景。且通过表 1 所得结果, 异形卷积核对于曲线特征更加敏感, 而对于直线特征的学习能力并没有矩形卷积块强, 所以将两者结合提升了 CNN 的特征提取能力, 从而提升了准确率。通过三个数据集的仿真实验, 我们发现单独使用移位图片训练出的 CNN 网络没有正常图片训练出的 CNN 网络正确率高, 那么单独使用异形卷积核并没有矩形卷积核的效果好。这应该是异性卷积核只对某些特征比较敏感的结果。

如表 3 所示融合网络在 MNIST、Fashion MNIST 和 CIFAR-10 三种数据集中的准确率都明显优于其它算法, 且融合网络可以适用于大多数读者自己创建的 CNN, 只需要对训练数据集进行修改便可以将异形卷积核应用到网络中。这相对于对网络内部的修改有着更强可行性[15]。在实际问题中, 可以针对问题中较为明显的特征设计异形卷积核, 且对于更大的卷积核有着更多的变化方式。

Table 3. Comparison of accuracy rates of various methods
表 3. 各类方法准确率对比

数据集	识别方法	准确率%
MNIST	线性分类器	91.90
	K 近邻	95.10
	PCA	96.72
	SVM	98.60
	LeNet-5	99.20
	融合网络	99.41
Fashion-MNIST	SVM	84.69
	CNN	88.31
	融合网络	89.63
CIFAR	CNN	76.9
	K 均值	79.2
	融合网络	80.3

基金项目

陕西省自然科学基金基础研究计划资助项目(项目编号 2018JQ1038)。

参考文献

- [1] Lecun, Y. and Bottou, L. (1998) Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, **86**, 2278-2324. <https://doi.org/10.1109/5.726791>
- [2] Krizhevsky, A., Sutskever, I.I. and Hinton, G. (2012) Imagenet Classification with Deep Convolutional Neural Networks. *Proceedings of the Advances in Neural Information Processing Systems*, Lake Tahoe, 1097-1105.
- [3] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. *ICLR 2015*, San Diego, 7-9 May 2015, 4898-4906.
- [4] Szegedy, C., Liu, W., Jia, Y., et al. (2014) Going Deeper with Convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 7-12 June 2015, 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
- [5] He, K., Zhang, X., Ren, S., et al. (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [6] Huang, G., Liu, Z., Weinberger, K.Q., et al. (2016) Densely Connected Convolutional Net Works. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 2261-2269. <https://doi.org/10.1109/CVPR.2017.243>
- [7] Sabour, S., Frosst, N. and Hinton, G.E. (2017) Dynamic Routing between Capsules. *Advances in Neural Information Processing Systems*, Long Beach, 3856-3866.
- [8] Gu, J., Wang, Z., Kuen, J., et al. (2015) Recent Advances in Convolutional Neural Networks. *Computer Science*.
- [9] Zeiler, M.D. and Fergus, R. (2014) Visualizing and Understanding Convolutional Networks. In: *European Conference on Computer Vision*, Springer, Berlin, 818-833. https://doi.org/10.1007/978-3-319-10590-1_53
- [10] Cao, K. and Jain, A.K. (2015) Latent Orientation Field Estimation via Convolutional Neural Network. *International Conference on Biometrics*, Phuket, 19-22 May 2015, 349-356. <https://doi.org/10.1109/ICB.2015.7139060>
- [11] 李章维, 胡安顺, 王晓飞. 基于视觉的目标检测方法综述[J]. *计算机工程与应用*, 2020, 56(8): 1-9.
- [12] 陈超, 齐峰. 卷积神经网络的发展及其在计算机视觉领域中的应用综述[J]. *计算机科学*, 2019, 46(3): 63-73.
- [13] 唐贤伦, 杜一铭, 刘雨微, 李佳歆, 马艺玮. 基于条件深度卷积生成对抗网络的图像识别方法[J]. *自动化学报*,

2018, 44(5): 855-864.

- [14] Zhang, K., Guo, Y., Wang, X., *et al.* (2019) Multiple Feature Reweight DenseNet for Image Classification. *IEEE Access*, 7, 9872-9880. <https://doi.org/10.1109/ACCESS.2018.2890127>
- [15] 张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用[J]. 计算机学报, 2019, 42(3): 453-482.