



# 计算机科学

COMPUTER SCIENCE

## DRSTN:深度残差软阈值化网络

曹岩, 朱真峰

引用本文

曹岩, 朱真峰. DRSTN:深度残差软阈值化网络[J]. 计算机科学, 2024, 51(6A): 230400112-7.

CAO Yan, ZHU Zhenfeng. DRSTN:Deep Residual Soft Thresholding Network[J]. Computer Science, 2024, 51(6A): 230400112-7.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于改进Efficientnetv2模型的铁矿石图像分类方法](#)

Iron Ore Image Classification Method Based on Improved Efficientnetv2

计算机科学, 2024, 51(6A): 230600212-6. <https://doi.org/10.11896/jsjcx.230600212>

[基于集成学习的MRI脑肿瘤智能诊断](#)

Intelligent Diagnosis of Brain Tumor with MRI Based on Ensemble Learning

计算机科学, 2024, 51(6A): 230600043-7. <https://doi.org/10.11896/jsjcx.230600043>

[基于快速傅里叶卷积与特征修剪坐标注意力的壁画修复](#)

Mural Inpainting Based on Fast Fourier Convolution and Feature Pruning Coordinate Attention

计算机科学, 2024, 51(6A): 230400083-9. <https://doi.org/10.11896/jsjcx.230400083>

[基于聚类优化学习的少样本图像分类](#)

Few-shot Images Classification Based on Clustering Optimization Learning

计算机科学, 2024, 51(6A): 230300227-7. <https://doi.org/10.11896/jsjcx.230300227>

[基于跨语言迁移学习及联合训练的泰语语音合成](#)

Thai Speech Synthesis Based on Cross-language Transfer Learning and Joint Training

计算机科学, 2024, 51(6A): 230500174-7. <https://doi.org/10.11896/jsjcx.230500174>

# DRSTN:深度残差软阈值化网络

曹岩 朱真峰

郑州大学计算机与人工智能学院 郑州 450001

(202022172013154@gs.zzu.edu.cn)

**摘要** 在采用深度残差等神经网络模型解决图像分类任务时,特征提取过程损失的一些重要特征会影响模型的性能。神经网络“端到端”的学习模式带来的黑盒问题,也会限制其在诸多领域的应用和发展。另外,神经网络模型往往需要较长的训练时间。为了提高深度残差网络模型的性能和训练效率,引入了模型迁移方法和软阈值化方法,提出了DRSTN(Deep Residual Soft Thresholding Network)网络,并对此网络结构进行微调,生成了不同版本的DRSTN网络。DRSTN网络的性能得益于3个方面的有机整合:1)通过梯度加权类激活映射(Gradients-weighted Class Activation Mapping, Grad-CAM)方法对网络的特征提取进行可视化,根据可视化结果挑选进一步优化的模型;2)基于模型迁移,研究人员不必全新地搭建模型,可以直接在已有的模型上进行优化,能够节省大量训练时间;3)软阈值化作为非线性变换层嵌入到深度残差网络体系结构中,以消除样本中不相关的特征。实验结果表明,在相同训练条件下,DRSTN\_KS(3 \* 3)\_RB(2;2;2)网络在CIFAR-10数据集上的分类精度相比SKNet-18, ResNet18和ConvNeXt\_tiny网络分别提高了15.5%, 8.8%和10.9%;该网络也具有一定的泛化性,在MNIST和Fashion MNIST数据集上能够达到快速的迁移效果,分类精度分别达到99.06%和93.15%。

**关键词:** 迁移学习; 残差网络; 梯度加权类激活映射; 软阈值化方法; 图像分类

**中图分类号** TP391

## DRSTN: Deep Residual Soft Thresholding Network

CAO Yan and ZHU Zhenfeng

School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, China

**Abstract** When using neural network models such as deep residuals to classify images, some important features lost during feature extraction will affect the classification performance of the model. The black box problem brought about by the “end-to-end” learning mode of neural network can also limit its application and development in many fields. In addition, neural network models often require longer training time than traditional methods. In order to improve the classification effect and training efficiency of the deep residual networks, this paper introduces the model transfer method and soft thresholding method, proposes the deep residual soft thresholding network(DRSTN) network, and fine-tunes the network structure to generate different versions DRSTN network. The performance of the DRSTN networks benefit from the organic integration of three aspects: 1) Visualize the feature extraction of the network through the gradients-weighted class activation mapping(Grad-CAM) method, and select further optimized ones based on the visualization results. 2) Based on model transfer, researchers do not need to build a model from scratch, and can directly optimize the existing models, which can save a lot of training time. 3) Soft thresholding, as a nonlinear transformation layer, is embedded into the deep residual network architecture to eliminate irrelevant features in samples. Experimental results show that under the same training conditions, the classification accuracy of the DRSTN\_KS(3 \* 3)\_RB(2;2;2) network on the CIFAR-10 dataset is 15.5%, 8.8% and 10.9% higher than that of SKNet-18, ResNet18 and ConvNeXt\_tiny networks, respectively. The network also has a certain degree of generalization. It can achieve rapid transfer on MNIST and Fashion MNIST datasets, and the classification accuracy reaches 99.06% and 93.15% respectively.

**Keywords** Transfer learning, Residual network, Gradient weighted class activation mapping, Soft thresholding method, Image classification

### 1 引言

近些年,随着人工智能领域的飞速发展,卷积神经网络(Convolutional Neural Network, CNN)<sup>[1]</sup>作为机器学习的一个重要的学习模型,已经成为该领域的核心之一,广泛应用于

图像分类<sup>[2]</sup>、智能控制<sup>[3]</sup>、语音识别<sup>[4]</sup>、人脸识别<sup>[5]</sup>和表情识别<sup>[6]</sup>等诸多领域。

为了更有效地处理呈现指数型增长的图像数据,新型网络结构从只有三层的BP网络发展到现在的几十层甚至上百层的复杂网络结构。网络结构的复杂性对硬件设备提出了更

基金项目:国家自然科学基金面上项目(62176239)

This work was supported by the National Natural Science Foundation of China(62176239).

通信作者:朱真峰(iezizhu@zzu.edu.cn)

高的要求,同时也对训练数据的效率、拟合数据的效果以及模型的泛化能力等问题带来了严峻挑战。研究人员在解决上述问题的过程中,不仅需要采用损失函数、激活函数等一些传统的技术手段,也需要尝试新的技术,以满足不同应用的需求。

为了保证网络模型的优化性能,研究人员对网络模型的深度和宽度进行改进,相继提出了多种新型网络结构: AlexNet<sup>[7]</sup>, VGGNet<sup>[8]</sup>, GoogleNet<sup>[9]</sup>, ResNet<sup>[10]</sup>, DenseNet<sup>[11]</sup> 和 SqueezeNet<sup>[12]</sup>等。与此同时,研究人员也通过相关方法对现有网络模型进行优化,例如,采用选择性内核(Selective Kernel, SK)<sup>[13]</sup>卷积改进残差网络,选择自适应核提高目标识别的效率和有效性。近年来,随着 Transformer 方法在解决图像分类领域的成功应用,也出现了多种版本,如 Vision Transformer<sup>[14]</sup>和 Swin Transformer<sup>[15]</sup>等。这些方法的出现逐渐取代了卷积神经网络在机器学习领域的主导地位。2022年, Facebook AI Research 和 UC Berkeley 提出了一种纯卷积神经网络 ConvNeXt<sup>[16]</sup>。与 Swin Transformer 相比, ConvNeXt 在准确性和处理效率两个方面都更加有效,这再次引起了卷积神经网络的研究热潮。

目前的学习方法仍然存在一些问题: 1) 网络模型的搭建和训练,需要消耗大量的时间和资源; 2) 准确识别物体的能力上还存在一定的瓶颈,处理高噪声图像数据时,因为噪声的干扰,作为局部特征提取器的卷积核不能保证检测到与图像相关的特征。为了更加有效地解决图像分类问题,本文结合梯度加权类激活映射方法、模型迁移方法和软阈值化方法提出了更加合适的模型结构和训练方法。实验结果表明,本文提出的模型及训练方法在图像分类任务上能够达到更好的性能。

## 2 相关工作

### 2.1 类激活映射

类激活映射(Class Activation Mapping, CAM)<sup>[17]</sup>是一种可视化方法,常常用于多卷积层神经网络的定位映射。如图1所示, CAM方法对卷积层的所有特征图进行全局平均池化操作,得到每个特征图的空间平均值,然后将这些值与全连接层的权重进行加权求和操作,生成网络的输出结果。类似地, CAM方法也可以通过使用全连接层的权重来加权组合卷积层的特征图,从而获得相应的类激活图。

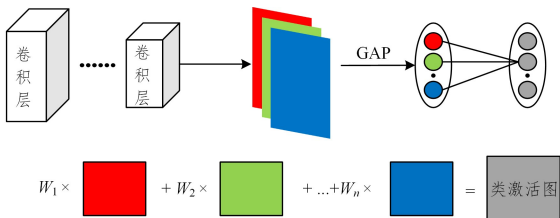


图1 类激活映射方法示意图

Fig.1 Diagram of class activation mapping method

假设最后一层卷积层产生  $K$  个特征映射,  $A^k$  代表第  $k$  个特征图,  $A^k \in R^{u \times v}$ ;  $Z$  代表特征图的尺寸大小。第  $k$  个特征图的全局平均池化(Global Average Pooling, GAP)为:

$$F^k = \frac{1}{Z} \sum_i \sum_j A_{i,j}^k \quad (1)$$

其中,  $A_{i,j}^k$  指的是特征图  $(i, j)$  位置上的激活映射。最终 CAM 计算的类输出  $Y^c$  为:

$$Y^c = \sum_k \omega_k^c \cdot F^k \quad (2)$$

其中,  $\omega_k^c$  是第  $k$  个特征图在类别  $c$  上的权重, 表示第  $k$  个特征图的重要程度。最后使用全局平均池化将这些特征图进行空间合并, 然后进行线性转换, 为每个类别  $c$  生成一个分数  $Y^c$ , 再将式(1)代入式(2)得:

$$\begin{aligned} Y^c &= \sum_k \omega_k^c \cdot \frac{1}{Z} \sum_i \sum_j A_{i,j}^k \\ &= \sum_i \sum_j \frac{1}{Z} \sum_k \omega_k^c A_{i,j}^k \end{aligned} \quad (3)$$

令  $Q_{i,j} = \frac{1}{Z} \sum_k \omega_k^c A_{i,j}^k$ , 则  $Y^c$  可以表示为:

$$Y^c = \sum_i \sum_j Q_{i,j} \quad (4)$$

由式(4)可知,  $Q_{i,j}$  表示了空间位置  $(i, j)$  处的特征信息对于类别  $c$  的重要程度。CAM方法采用全局平均池化的结构特点解释卷积神经网络的分类结果。由于全局平均池化的结构特点, CAM方法不能应用到卷积网络中的一般卷积层。因此, CAM方法不便于解释CNN中各个卷积层提取到的具体特征。

### 2.2 梯度加权类激活映射

梯度加权类激活映射(Gradients-weighted Class Activation Mapping, Grad-CAM)<sup>[17]</sup>可视化方法是CAM的一种泛化形式, 没有改变相应神经网络的结构, 直接将网络输出对特征图求导来计算类激活图。通过对式(2)的  $Y^c$  相对于  $F^k$  求偏导得出类别  $c$  相对于特征图  $F^k$  的分数梯度:

$$\frac{\partial Y^c}{\partial F^k} = \frac{\frac{\partial Y^c}{\partial A_{i,j}^k}}{\frac{\partial F^k}{\partial A_{i,j}^k}} \quad (5)$$

通过式(1)中的  $F^k$  对  $A_{i,j}^k$  求偏导可以得出  $\frac{\partial F^k}{\partial A_{i,j}^k} = \frac{1}{Z}$ , 将其代入式(5)可得:

$$\frac{\partial Y^c}{\partial F^k} = \frac{\partial Y^c}{\partial A_{i,j}^k} \cdot Z \quad (6)$$

同理, 对式(2)中的  $Y^c$  对  $F^k$  求偏导得  $\frac{\partial Y^c}{\partial F^k} = \omega_k^c$ 。再结合式(6), 可知:

$$\omega_k^c = Z \cdot \frac{\partial Y^c}{\partial A_{i,j}^k} \quad (7)$$

对特征空间中的所有像素  $(i, j)$  进行累加得:

$$\sum_i \sum_j \omega_k^c = \sum_i \sum_j Z \cdot \frac{\partial Y^c}{\partial A_{i,j}^k} \quad (8)$$

由于特征图的尺寸大小  $Z (Z = \sum_i \sum_j 1)$  和第  $k$  个特征图在类别  $c$  上的权重  $\omega_k^c$  不依赖于像素, 可以得出:

$$Z \cdot \omega_k^c = Z \sum_i \sum_j \frac{\partial Y^c}{\partial A_{i,j}^k} \quad (9)$$

最后, 第  $k$  个特征图在类别  $c$  上的权重  $\omega_k^c$  计算公式如下:

$$\omega_k^c = \sum_i \sum_j \frac{\partial Y^c}{\partial A_{i,j}^k} \quad (10)$$

CAM通过对网络的全局特征进行平均池化操作, 得到特征图每个通道的权重, 然后对这些权重进行加权求和得到网络模型注意力关注区域的热力图。Grad-CAM利用反向传播计算权重, 不用专门采用全局平均池化的结构特点, 而是将每个梯度单元的平均值作为特征图  $A^k$  的权重  $\omega_k^c$ , 等价于全局平均池化操作, 该权重就是每个特征图对于类别  $c$  的重要性权重, 如式(11)所示。

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial Y^c}{\partial A_{i,j}^k} \quad (11)$$

Grad-CAM 通过计算对每个特征图对于类别  $c$  的重要性权重  $\alpha_k^c$  与特征图  $A^k$  进行相乘累加,最后使用 ReLU 激活函数得到最终的类激活映射图,如式(12)所示。

$$L_{\text{Grad-CAM}} = \text{ReLU}(\sum_k \alpha_k^c A^k) \quad (12)$$

### 2.3 迁移学习

迁移学习旨在通过将源领域的知识迁移到目标领域中,使得目标领域的学习更加高效<sup>[18]</sup>。与传统机器学习相比,迁移学习不再要求源数据和目标数据必须具有相同分布,而是通过知识迁移,借助少量训练数据,就能够完成目标任务,同时减少训练时间。迁移学习涉及领域 Domain 和任务 Task,可以表示如下:



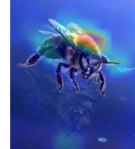
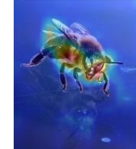

$$\text{Domain} = \{X, P(x)\}, x \in X \quad (13)$$

$$\text{Task} = \{Y, f(x)\}, y \in Y \quad (14)$$

其中,  $X$  是  $d$  维特征空间,包含所有特征向量;  $P(x)$  是边缘概率分布,代表某种特定的采样。如果  $X$  表示一个三维空间,  $P(x)$  可以是此空间中过原点的一个面。给定领域 Domain, 任务 Task 定义为由类别空间  $Y$  和预测模型  $f(x)$  组成。按统计观点,预测模型  $f(x) = P(y|x)$  解释为条件概率分布。领域 Domain 可以分为源域和目标域,任务 Task 可以分为源任务和目标任务;源域和目标域的数据分布可以相同,也可以不同。迁移学习大致可以分为 4 类:基于样本的迁移、基于特征的迁移、基于模型的迁移以及基于预训练参数的迁移。本文主要运用模型迁移方法和预训练参数迁移方法。

表 1 网络模型的类激活映射图

Table 1 Class activation mapping plots of network models

Original	Squeezenet1_1	AlexNet	VGG-16	ResNet-18
				

由表 1 可知,不同网络模型的注意力位置(即图中的高亮区域)等也不相同。ResNet-18 网络模型能够定位整体目标,注意力集中于头部和身体,蜜蜂爪子部分的权重相对较小;VGG-16 网络模型聚焦在蜜蜂的腹部;AlexNet 网络模型集中在蜜蜂的翅膀;Squeezenet1\_1 网络模型的关注点是一部分爪子。

类激活图的对比表明,不同网络模型对于特征的关注点、特征强度以及区域的大小都有所不同,也说明了不同网络模型在特征学习方面的差异。上述对比分析表明,ResNet-18 网络模型能够更好地提取图像的全局特征。因此,本文选择残差网络结构对其进行进一步优化。

### 3.2 软阈值化方法

软阈值化(Soft Thresholding)是一种有效的信号去噪<sup>[19]</sup>方法。首先将原始信号中不重要的部分转换为接近于零的数值,然后利用软阈值化将近零的特征转换为零。小波阈值化是一种广泛应用于信号去噪的方法,通常由 3 个步骤组成:小波分解、软阈值化和小波重构。然而,设计适合特定信号的小波滤波器是非常具有挑战性的,因为它需要深入了解信号特征,同时需要具有一定的专业知识。为了解决这个问题,深度

## 3 深度残差软阈值化网络模型

本文聚焦于卷积神经网络,结合梯度加权类激活映射方法、迁移方法和软阈值化方法,设计了更加合适的模型结构和训练方法,提出了深度残差软阈值化网络模型(Deep Residual Soft Thresholding Network, DRSTN)。主要工作如下:

(1)首先将 Grad-CAM 方法应用到不同网络模型中,为每个网络模型生成一个类激活映射图,从特征注意力的角度比较多种网络模型的优劣性,最终确定采用残差网络开展进一步的模型优化工作。

(2)借鉴 ConNeXt 网络模型搭建的思想和模型迁移方法,将软阈值化方法作为非线性变换层插入到 ResNet-18 网络中,并对其进行微调,得出 DRSTN 网络模型的不同版本。

(3)通过实验验证了 DRSTN 网络模型的性能,并通过预训练参数迁移方法验证了其知识迁移效果。

### 3.1 特征图可视化挑选优化模型

Grad-CAM 是一种可解释性机器学习方法,可以帮助我们理解模型在图像分类任务中关注的实际区域,以及如何调整输入图像来影响模型的决策。通过了解模型的决策过程,可以更好地理解模型的优劣,并根据需要进行改进和优化。本文将 Grad-CAM 方法应用到不同网络模型中,为每个网络模型生成一个类激活映射图,如表 1 所列,从特征注意力的角度比较网络模型的优劣性,避免了筛选网络模型时的繁琐训练过程,节省了大量时间。

学习提供了一种新的途径,使用梯度下降方法自动学习滤波器,更好地适应不同类型信号。这种方法利用卷积神经网络的端到端的处理模式,自动学习信号的特征表示和滤波器参数,从而取代了手工设计滤波器的繁琐过程。因此,软阈值化和深度学习的集成能够有效地消除噪声,同时构建区分度比较明显的鉴别特征。软阈值化函数可表示为:

$$f(x) = \begin{cases} x - \tau, & x > \tau \\ 0, & -\tau \leq x \leq \tau \\ x + \tau, & x < -\tau \end{cases} \quad (15)$$

其中,  $x$  为输入特征,  $f(x)$  为输出特征,  $\tau$  为正的阈值。ReLU 激活函数是将负特征设置为零,而软阈值化方法是把接近于零的特征设置为零,可以保留有用的负特征。通过式(15)中的  $f(x)$  对  $x$  求导可知:

$$f'(x) = \begin{cases} 1, & x > \tau \\ 0, & -\tau \leq x \leq \tau \\ 1, & x < -\tau \end{cases} \quad (16)$$

式(16)表明,软阈值化函数的导数非 0 即 1,此属性和 ReLU 激活函数相同,说明软阈值化方法可以有效地防止梯度消失和梯度爆炸问题<sup>[18]</sup>。

### 3.3 DRSTN 网络模型搭建过程与训练流程

首先将软阈值化方法作为非线性变换层插入到 ResNet-18 网络中,得到深度残差软阈值化网络(Deep Residual Soft Thresholding Network, DRSTN)的第一个版本。然后采用模型迁移方法对该网络模型进行微调,得到不同形式的模型结构。其中包括:调整卷积核的大小和残差模块的比例。首先在第一个版本的基础上,将 Conv\_1 卷积核的大小从  $7 \times 7$  调整到  $3 \times 3$ ,得到 DRSTN 网络的第二个版本;然后在第二个版本的基础上通过改变网络中残差模块的比例,得出以下版本。首先,同比减少残差模块,由原来 Conv\_2 到 Conv\_5 堆叠的残差块次数为(2,2,2,2)改为(1,1,1,1),得到 DRSTN 网络的第

三个版本;其次,根据残差网络的阶段性计算比例的特点,即在 Conv\_4 堆叠的残差块数量最多,本文把 Conv\_2 到 Conv\_5 堆叠的残差块次数改为(2,3,5,2),得到 DRSTN 网络的第四个版本。最后,根据残差网络的阶段性计算比例的特点,即使在 Conv\_4 堆叠的残差块数量最多,也可以猜想 Conv\_5 的残差模块对网络性能的影响不大,于是将 Conv\_5 这一阶段的残差模块去掉,就形成了 DRSTN 网络的最后一个版本。DRSTN 网络模型不同版本结构如表 2 所示,其中 KS(Kernel Size)表示卷积核尺寸, RB(Residual Block)表示残差模块, STRB(Soft Thresholding Residual Block)表示软阈值化残差模块。

DRSTN 网络模型不同版本的详细结构如表 2 所列。

表 2 DRSTN 网络模型不同版本的详细结构

Table 2 Detailed structure of different versions of DRSTN network model

	ResNet-18	DRSTN_V1	DRSTN_V2	DRSTN_V3	DRSTN_V4	DRSTN_V5
Conv_1	KS $7 \times 7$	KS $7 \times 7$	KS $3 \times 3$	KS $3 \times 3$	KS $3 \times 3$	KS $3 \times 3$
Conv_2	RB $\times 2$	STRB $\times 2$	STRB $\times 2$	STRB $\times 1$	STRB $\times 2$	STRB $\times 2$
Conv_3	RB $\times 2$	STRB $\times 2$	STRB $\times 2$	STRB $\times 1$	STRB $\times 3$	STRB $\times 2$
Conv_4	RB $\times 2$	STRB $\times 2$	STRB $\times 2$	STRB $\times 1$	STRB $\times 5$	STRB $\times 2$
Conv_5	RB $\times 2$	STRB $\times 2$	STRB $\times 2$	STRB $\times 1$	STRB $\times 2$	无

DRSTN 网络模型的工作原理为:通过在残差模块中加入软阈值化方法,将不重要的特征设置为零,增强了网络提取图像特征的能力,从而提升了网络模型在解决图像分类问题上的性能。DRSTN 网络模型具体的训练流程如图 2 所示。其中,  $K$  为卷积层中卷积核的数量,  $M$  为全连接(Full Connection, FC)网络的神经元数量;  $C, W$  和  $H$  分别为图像的通道数量、宽度和高度。在图 2 的蓝色虚框中,首先是将输入数据中特征元素取绝对值,然后经过全局平均池化操作得到一个输出特征图,记作  $A$ 。接着,将输出特征图分为两步进行操作。首先,通过两层全连接网络对输出特征图进行处理,经过批处理、归一化、激活等一系列操作,最后将输出特征图的结果

映射到一个大小为 0 到 1 的区间中,记作  $\alpha$ 。然后将输出特征图  $A$  与  $\alpha$  相乘,得到一个范围为 0 到 1 的正阈值。上述操作是使用一个非零阈值对权重进行软化,能够很好地限制权重的变化,使得模型更加稳定,并能有效防止过拟合问题。此外,不同的样本对应不同的阈值,软阈值化方法可以通过设定阈值大小来提取样本特征。其原理就是,将阈值范围内的特征视为与当前任务无关的特征,将其舍弃;与当前任务无关的特征,通过软阈值化将其设置为 0,与当前任务有关的特征,保持不变。最终通过软阈值化残差模块的叠加以及模型的微调操作,就形成了完整的 DRSTN 训练流程。

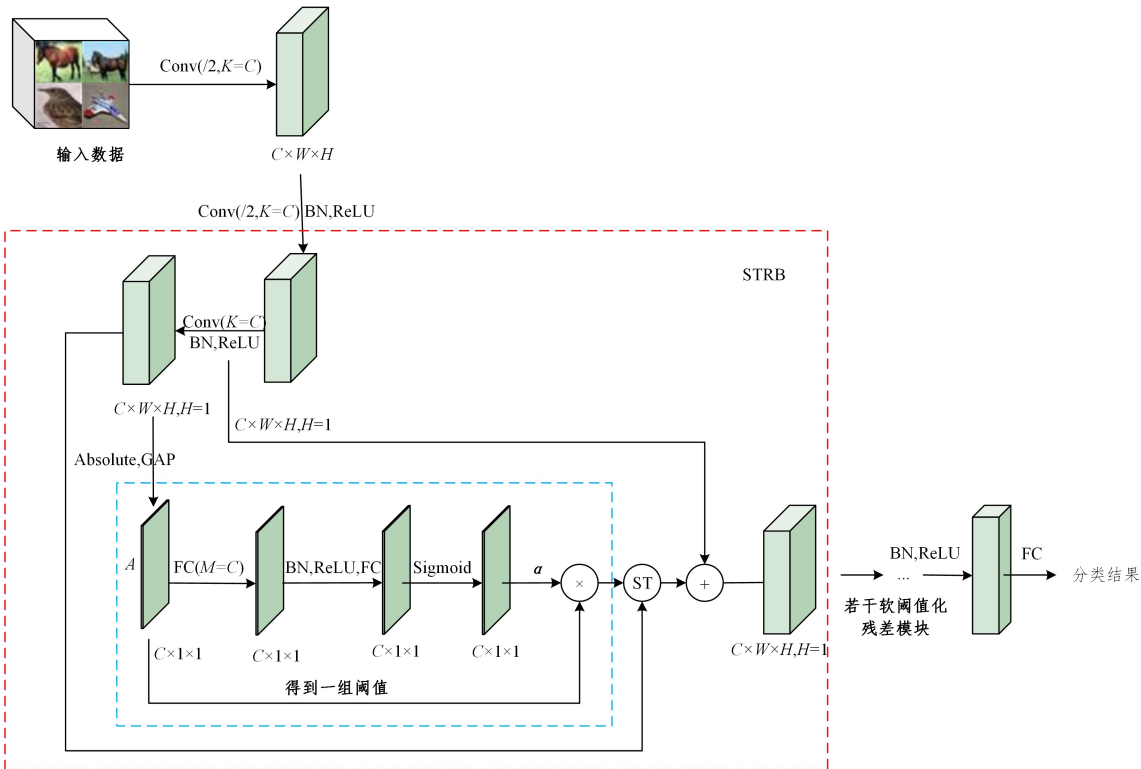


图 2 DRSTN 的训练流程图  
Fig. 2 DRSTN training flow chart

## 4 实验与分析

### 4.1 相关数据集

本文在多个数据集上进行实验,其中包括 MNIST, Fashion MNIST 和 CIFAR-10 数据集等。MNIST 数据集包含 10 个类别,其内容为手写字体 0 到 9;训练样本为 60000 张灰度图,验证样本为 10000 张灰度图,图片大小都为  $28 \times 28$ 。Fashion MNIST 是 MNIST 数据集的一种扩展,包含 10 个类别的时尚产品,一共 70000 张大小为  $28 \times 28$  的灰度图像,训练样本为 60000 张灰度图,验证样本为 10000 张灰度图,每个类别是 7000 张图像。CIFAR-10 数据集由 60000 张大小为  $32 \times 32$  的 RGB 图像组成,训练样本为 50000 张彩色图像,验证样本为 10000 张彩色图像,这些图像被标记为 10 个类别:飞机、汽车(但不是卡车或皮卡车)、鸟、猫、鹿、狗、青蛙、马、船和卡车(但不是皮卡车),每种类别有 6000 张图片。数据集的详细信息如表 3 所列。

表 3 数据集详细信息  
Table 3 Dataset details

数据集	类别数	图片类型	图片尺寸	训练样本数	验证样本数
MNIST	10	Gray	$28 \times 28$	60000	10000
Fashion MNIST	10	Gray	$28 \times 28$	60000	10000
CIFAR-10	10	RGB	$32 \times 32$	50000	10000

### 4.2 实验环境及过程

本文的所有实验均在 Windows10 操作系统的计算机上

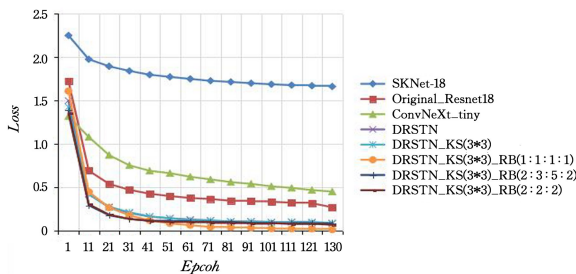


图 3 CIFAR-10 数据集上的对比实验曲线图

Fig. 3 Comparative experimental plots on CIFAR-10 dataset

由损失值变化曲线知,相比 SKNet-18, ResNet-18 和 ConvNeXt\_tiny 网络模型,本文所提出的 DRSTN 系列网络模型在图像处理的过程中能够更快收敛。最终 Loss 曲线达到相对稳定时,DRSTN 系列网络模型的损失值比 SKNet-18, ResNet-18 和 ConvNeXt\_tiny 网络模型都要小,说明 DRSTN 系列网络模型不仅能够更快地缩小预测值和真实值之间的差距,而且在 CIFAR-10 数据集上还比 SKNet-18, ResNet-18 和 ConvNeXt\_tiny 网络模型具有更好的学习效果。由准确率变化曲线知,通过少量训练就能达到相对较高的分类准确率,并且准确率曲线达到平衡时,相较于 SKNet-18, ResNet-18 和 ConvNeXt\_tiny 网络模型,DRSTN 系列网络模型拥有更高的分类准确率,这说明 DRSTN 系列网络模型在解决图像分类任务时,预测的结果和真实结果能够达到更高的相似度。损失值变化曲线和准确率变化曲线,验证了 DRSTN 系列网络模型具有较好的图像分类性能。为了能够更详细地体现 DRSTN 系列网络模型的优越性,本文记录了各网络模型在 CIFAR-10 数据集上的最优分类准确率,如表 4 所列。

进行。计算机内存为 16GB,处理器为 Intel(R) Core(TM) i5-10400 CPU @ 2.90GHz。实验所用编译器为 Visual Studio Code,编译语言为 Python。首先,将 DRSTN 网络模型与 ResNet-18 和 ConvNeXt\_tiny 网络模型在 CIFAR-10 数据集上进行实验分析,采用了准确率(Accuracy)和损失值(Loss)两个指标对分类结果进行评估,以此验证 DRSTN 网络模型的性能。损失值(Loss)的大小是来自交叉熵损失函数,用来评估预测值和真实值之间的差异。具体如式(17)所示:

$$Loss = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_{i,c=1}^M y_{ic} \log(P_{ic}) \quad (17)$$

其中, $M$  代表样本的种类数量; $i$  代表每个样本; $c$  代表样本类别; $y_{ic}$  是一个指标变量, $y_{ic}$  取值为 1 代表预测的样本类别与真实的样本类别相同, $y_{ic}$  取值为 0 代表两者不同; $P_{ic}$  是观察样本  $i$  属于类别  $c$  的预测概率。

然后,采用预训练参数迁移方法,将 DRSTN 网络模型在 MNIST 和 Fashion MNIST 数据集上进行实验分析,以此验证 DRSTN 网络模型的知识迁移性能(泛化性能)。

### 4.3 实验结果及分析

图 3 给出了各网络模型在 CIFAR-10 数据集上的损失值变化和准确率变化。DRSTN 对应的是表 2 中的 DRSTN\_V1;DRSTN\_KS(3 \* 3)对应的是表 2 中的 DRSTN\_V2;DRSTN\_KS(3 \* 3)\_RB(1:1:1:1)和 DRSTN\_KS(3 \* 3)\_RB(2:3:5:2)分别对应的是表 2 中的 DRSTN\_V3 和 DRSTN\_V4;DRSTN\_KS(3 \* 3)\_RB(2:2:2)对应的是表 2 中的 DRSTN\_V5。

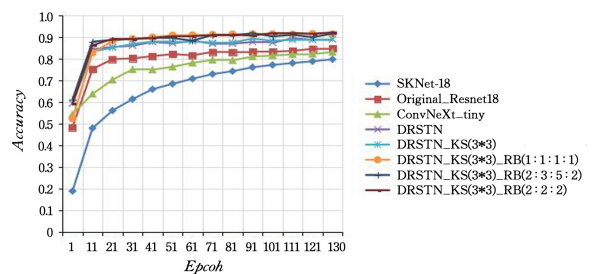


表 4 CIFAR-10 数据集上的最优准确率

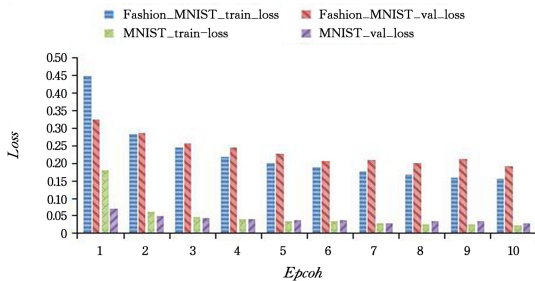
Table 4 Optimal accuracy on CIFAR-10 dataset

模型	Best_Acc
SKNet-18	0.7990
Original_Resnet18	0.8484
ConvNeXt_tiny	0.8319
DRSTN	0.8962
DRSTN_KS(3 * 3)	0.8944
DRSTN_KS(3 * 3)_RB(1:1:1:1)	0.9172
DRSTN_KS(3 * 3)_RB(2:3:5:2)	0.9119
DRSTN_KS(3 * 3)_RB(2:2:2)	0.9229

由表 4 可知,在 CIFAR-10 数据集上,SKNet-18, ResNet-18 和 ConvNeXt\_tiny 网络模型达到的最优分类准确率分别为 79.90%, 84.84% 和 83.19%, DRSTN 系列网络模型的最优分类准确率分别为 89.62%, 89.44%, 91.72%, 91.19% 和 92.29%, 相较于 SKNet-18, ResNet-18 和 ConvNeXt\_tiny 网络模型在分类性能上都提升。经计算得知,DRSTN\_KS(3 \* 3)\_RB(2:2:2)的最优分类准确率比 SKNet-18 网络模型

提高了大约 15.5%，比 ResNet-18 网络模型提高了大约 8.8%，比 ConvNeXt\_tiny 网络模型提高了大约 10.9%。

上述实验表明，本文提出的模型在处理 CIFAR-10 数据集上具有较好的效果。为了验证 DRSTN 网络模型在图像分类问题上是否真正学习到有用信息，以及是否能够将学到的知识迁移到不同的数据集上。本文选用 DRSTN 系列网络模



型中分类准确率最高的 DRSTN\_KS(3 \* 3)\_RB(2:2:2)网络模型，采用基于预训练参数迁移方法进行训练，就是将该模型在 CIFAR-10 数据集达到最优准确率的预训练参数作为训练 MNIST 数据集和 Fashion MNIST 数据集的初始化参数，以此验证 DRSTN\_KS(3 \* 3)\_RB(2:2:2)网络模型的知识迁移效果，结果如图 4 所示。

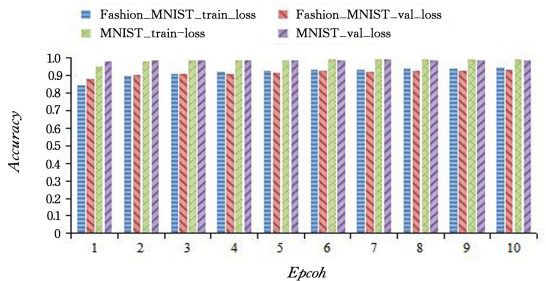


图 4 MNIST 和 Fashion MNIST 数据集上的分类结果

Fig. 4 Classification results on MNIST and Fashion MNIST datasets

由图 4 可知，损失值从一开始就处于较低的水平，并且随着训练一直在减小，并最终达到相对平稳状态，这说明 DRSTN\_KS(3 \* 3)\_RB(2:2:2)网络模型在经过 CIFAR-10 数据集训练后，能够很快地对其相似数据集进行识别处理，在短时间内缩小目标数据集预测值和真实值之间的差距，并且能够快速收敛。同时，DRSTN\_KS(3 \* 3)\_RB(2:2:2)网络模型能够一开始就达到相对较高的分类准确率，并且随着训练的进行，准确率平稳上升，最终达到平衡状态。结合损失值和准确率的变化可以得出：DRSTN\_KS(3 \* 3)\_RB(2:2:2)网络模型在 CIFAR-10 学习到的特征提取知识能够较好地迁移到 MNIST 和 Fashion MNIST 数据集上。

为了更加详细地解释 DRSTN\_KS(3 \* 3)\_RB(2:2:2)网络模型，本文记录了它在 MNIST 和 Fashion MNIST 数据集实验过程中的最好结果，如表 5 所列。由表 5 可知，DRSTN\_KS(3 \* 3)\_RB(2:2:2)网络模型经过少量训练就能在 MNIST 和 Fashion MNIST 数据集上分别达到 99% 和 93% 以上的分类准确率。结合损失值和准确率两者的结果，可以验证 DRSTN\_KS(3 \* 3)\_RB(2:2:2)网络模型具有较好的拟合能力和泛化能力，能够适用于不同数据集。

表 5 MNIST 和 Fashion-MNIST 数据集上的最优值

Table 5 Optimal values on MNIST and Fashion-MNIST datasets

	Best_train_loss	Best_val_loss	Best_train_acc	Best_val_acc
MNIST	0.0306	0.0281	0.9903	0.9906
Fashion_MNIST	0.1553	0.1905	0.9434	0.9315

**结束语** DRSTN 网络模型是一种新型网络结构，结合梯度加权类激活映射方法、迁移方法和软阈值化方法设计了更加合适的模型结构和训练方法。实验结果表明，DRSTN 网络模型相较于 SKNet-18, ResNet-18 和 ConvNeXt\_tiny 网络模型在图像分类效果上有明显的提升，并且能够适用于不同的数据集，具有一定的泛化性。

在今后的研究中，我们将在更多数据集上验证所提模型的性能，并在现阶段工作的基础上对模型做进一步的优化。

## 参考文献

[1] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature,

2015, 521(7553): 436-444.

- [2] HASSANZADEH T, ESSAM D, SARKER R. EvoDCNN: An evolutionary deep convolutional neural network for image classification[J]. Neurocomputing, 2022, 488: 271-283.
- [3] FADLULLAH Z M, TANG F, MAO B, et al. State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems[J]. IEEE Communications Surveys & Tutorials, 2017, 19(4): 2432-2455.
- [4] ZHANG Z, GEIGER J, POHJALAINEN J, et al. Deep learning for environmentally robust speech recognition: An overview of recent developments[J]. ACM Transactions on Intelligent Systems and Technology, 2018, 9(5): 1-28.
- [5] DENG J, GUO J, XUE N, et al. Arcface: Additive angular margin loss for deep face recognition[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 4690-4699.
- [6] DU J, CHEN Y H, ZHANG L, et al. Low-power expression recognition based on improved deep residual network[J]. Computer Science, 2018, 45(9): 303-307.
- [7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [8] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409. 1556, 2015.
- [9] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.
- [10] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [11] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4700-4708.
- [12] HASSANPOUR M, MALEK H. Document image classification using SqueezeNet convolutional neural network[C]// 2019 5th Iranian Conference on Signal Processing and Intelligent Sys-

tems. IEEE, 2019:1-4.

- [13] LI X, WANG W, HU X, et al. Selective kernel networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:510-519.
- [14] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv:2010.11929, 2021.
- [15] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 10012-10022.
- [16] LIU Z, MAO H, WU C Y, et al. A convnet for the 2020s[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:11976-11986.
- [17] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017:618-626.
- [18] PAN S J, YANG Q. A survey on transfer learning[J]. IEEE

Transactions on Knowledge and Data Engineering, 2010, 22(10):1345-1359.

- [19] ISOGAWA K, IDA T, SHIODERA T, et al. Deep shrinkage convolutional neural network for adaptive noise reduction[J]. IEEE Signal Processing Letters, 2017, 25(2):224-228.



**CAO Yan**, born in 1993, postgraduate. His main research interests include machine learning and transfer learning.



**ZHU Zhenfeng**, born in 1980, Ph.D, associate professor, is a member of CCF (No. 18743M). His main research interests include data mining, machine learning and computer vision.