CN 43-1258/TP
ISSN 1007-130X

计算机工程与科学
Computer Engineering & Science

第 41 卷第 11 期 2019 年 11 月
Vol. 41, No. 11, Nov. 2019

文章编号:1007-130X(2019)11-2007-10

基于卷积神经网络的图像数据增强算法*

蒋　芸,张　海,陈　莉,陶生鑫
(西北师范大学计算机科学与工程学院,甘肃 兰州 730070)

摘　要:提升卷积神经网络的泛化能力和降低过拟合的风险是深度卷积神经网络的研究重点。遮挡是影响卷积神经网络泛化能力的关键因素之一,通常希望经过复杂训练得到的模型能够对遮挡图像有良好的泛化性。为了降低模型过拟合的风险和提升模型对随机遮挡图像识别的鲁棒性,提出了激活区域处理算法,在训练过程中对某一卷积层的最大激活特征图进行处理后对输入图像进行遮挡,然后将被遮挡的新图像作为网络的新输入并继续训练模型。实验结果表明,提出的算法能够提高多种卷积神经网络模型在不同数据集上的分类性能,并且训练好的模型对随机遮挡图像的识别具有非常好的鲁棒性。

关键词:深度学习;卷积神经网络;图像分类;数据增强
中图分类号:TP391.4　　**文献标志码**:A
doi:10.3969/j.issn.1007-130X.2019.11.015

An image data augmentation algorithm based on convolutional neural networks

JIANG Yun, ZHANG Hai, CHEN Li, TAO Sheng-xin
(College of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China)

Abstract: Improving the generalization ability and reducing the over-fitting risk is the research focus of deep convolutional neural networks. Occlusion is one of the critical factors affecting the generalization ability of convolutional neural networks. It is usually hoped that the models after complex training can have a good generalization for occlusion images. In order to reduce the over-fitting risk and improve the robustness of the model to random occlusion image recognition, this paper proposes an activation feature processing algorithm. During the training process, the input image is occluded by processing the maximum activation feature map of a convolutional layer, then the occluded new image is used as a new input to the network to go on training the model. The experimental results show that the proposed algorithm can improve the classification performance of multiple convolutional neural network models on different datasets and the trained models have excellent robustness to the identification of random occlusion images.

Key words: deep learning; convolutional neural network; image classification; data augmentation

1　引言

近年来,深度学习在计算机视觉领域取得了巨大进步,在许多具有挑战性的视觉任务中取得了最好的性能,如图像分类[1]、语义分割[2]、目标检测[3]和人体姿势识别[4]等。这些性能的大幅提升大部分可以归功于卷积神经网络 CNN(Convolutional

* 收稿日期:2019-04-29;修回日期:2019-06-04
基金项目:国家自然科学基金(61962054);2016 年甘肃省科技计划资助自然科学基金(1606RJZA047);2012 年度甘肃省高校基本科研业务费专项资金;甘肃省高校研究生导师项目(1201-16);西北师范大学第三期知识与创新工程科研骨干项目(nwnu-kjcxgc-03-67)
通信地址:730070 甘肃省兰州市安宁区西北师范大学计算机科学与工程学院
Address: College of Computer Science and Engineering, Northwest Normal University, Anning District, Lanzhou 730070, Gansu, P. R. China

Neural Network)[5],它能够学习图像的复杂层次特征。深度卷积神经网络通常包含几千万到几亿个学习参数,这些参数为复杂的图像识别任务提供必要的表示能力,但是随着网络越来越复杂以及参数的增加,其在训练过程中过拟合的风险随之增加,泛化能力也随之变差。在深度学习中,研究者提出了很多用于解决卷积神经网络模型过拟合问题的方法,包括正则化方法(Regularization)[6]、Dropout 算法[7]、数据增强(Data Augmentation)[1,8,9]、批归一化(Batch Normalization)算法[10]等。

正则化方法[6]是常用于解决神经网络过拟合的方法,包括 L_1 正则化、L_2 正则化等,这些正则化方法不仅可以控制模型的复杂度,提高模型的泛化能力,而且还可以约束模型的特性,例如稀疏、平滑等特性。在数学公式上体现为在最优化损失函数后面加上正则化项,也称为惩罚项,用于限制模型权重参数。批归一化算法[7]通过确保在深度神经网络训练过程中每一层神经网络的输入保持同分布来使得激活输入值落在非线性函数对输入比较敏感的区域,加速模型的收敛速度,也可以正则化模型,有效地防止模型过拟合。

在图像识别领域中,图像包含着各种巨大变化因素的高维数据,对训练集图像进行平移、旋转几个像素的数据增强[1,8,9]操作通常可以大大改善模型的泛化能力,降低过拟合风险,提高模型的鲁棒性。由于数据增强的有效性、可扩展性且易于实施,因此其广泛用于计算机视觉领域中。常用的数据增强方式有:旋转、翻转、裁剪、添加噪声、平移、错切变换等,通过这些方式对输入图像进行扩充,旨在通过扩充训练集图像来防止过拟合。虽然数据增强方法简单有效,但是对于不同数据集,通常需要人工设计不同的数据增强策略,因此需要丰富的实验经验来寻找一个最佳的数据增强策略。

由 Srivastava 等[7]提出的 Dropout 算法在模型训练期间以一定的丢失概率将隐藏层单元激活设置为零,在评估网络性能时保留所有激活,并根据丢失概率缩放得到输出。在全连接神经网络中使用 Dropout 算法能够得到非常好的正则化效果,有效减轻了过拟合的问题,提高了网络的鲁棒性与泛化能力,阻碍了相邻特征检测器相关性。但是,Dropout 算法对卷积神经网络的改善效果并不那么好,很大程度上归因于 2 个因素:首先,卷积层已经具有比全连接层更少的参数;其次,图像中的相邻像素共享大部分相同的信息,如果它们中的任何一个像素被丢弃,那么它们包含的信息可能仍然会从仍处于激活状态的相邻像素传递。由于这些原因,Dropout 算法虽然能够增加卷积神经网络对输入噪声的鲁棒性,但是不能对卷积神经网络起到模型平均效应。为了提高卷积层中 Dropout 算法的有效性,已经有很多研究者对标准 Dropout 算法进行改进。Tompson 等[11]提出的 SpatialDropout 通过随机丢弃整个特征图而不是单个像素,有效地绕过相邻像素传递类似信息的问题。在更有针对性的方法中,Park 等[12]提出了 max-drop,它以一定的概率降低了特征映射或通道的最大激活,虽然这种方法在某些情况下比卷积层上的标准 Dropout 算法表现更好,但是在使用批归一化的卷积神经网络中使用时,max-drop 和 SpatialDropout 都比标准 Dropout 算法的性能更差。

提升模型的泛化能力一直是深度学习领域中的研究重点。通常我们希望训练好的模型可以自动处理随机遮挡,并且在预测新数据时依旧表现良好,而不仅仅是对基础数据集数据分布的表示。遮挡是影响卷积神经网络泛化能力的关键因素之一[13]。李小薪等[14]从鲁棒分类器的设计和鲁棒特征提取 2 方面回顾了现有的遮挡人脸识别方法,并指出识别遮挡人脸的困难性主要体现在由遮挡所引发的特征损失、对准误差和局部混叠等方面。刘万军等[15]提出一种时空上下文抗遮挡视觉跟踪算法,能够用于光照变化、目标旋转、遮挡等复杂情况下的视觉目标跟踪,具有一定的实时性和高效性,尤其是在目标发生遮挡情况下具有很好的抗遮挡能力和较快的运行速度。储珺等[16]提出基于遮挡检测和时空上下文信息的目标跟踪算法,较好地解决了复杂场景下较严重的静态遮挡和动态遮挡问题。当图像的某些部分被遮挡时,强分类模型能够从整个图像结构中提取出全局特征并正确分类。然而,采用所收集的样本训练的网络模型通常在样本被遮挡方面表现出的泛化效果有限。在所有训练对象都清晰可见没有发生遮挡时,训练出来的卷积神经网络能在未被遮挡的测试图像中取得非常好的效果,但是由于卷积神经网络模型的泛化能力有限,可能无法识别部分被遮挡的对象。

降噪自动编码器[17]和上下文自动编码器[18]通过破坏输入图像并要求网络使用上下文像素来重构图像,以确定如何最好地降噪和填充空白使得模型更好地工作。这 2 种方法依赖无监督方式来学习如何从图像中更好地学习全局特征,而不仅仅是简单地学习标识特征。我们认为,模型能否利用好

图像的上下文信息(即能否提取出全局特征)和能否降低遮挡区域噪声带来的影响,是解决图像遮挡识别问题的一个方向。

受现有研究的启发,我们考虑根据上一个训练周期中网络的激活情况来自动对输入图像进行有针对的遮挡,以迫使网络使用更多的全局特征进行决策,而不是根据少量局部特征进行决策。这种方式与 Dropout 算法类似,但有 3 个重要的区别:(1)此方式仅对输入图像进行有针对的丢弃,而不是对中间层的特征像素进行随机丢弃;(2)此方式对输入图像部分连续区域进行遮挡,而不是随机对单个像素进行丢弃,有效地降低了相邻特征检测器之间的相关性;(3)我们希望能提升模型对遮挡识别的鲁棒性,而 Dropout 算法不能有效地应对遮挡问题。

本文提出了一种在训练过程中处理卷积层输出的激活特征图,然后根据特征图的激活区域来实现有针对地对输入图像进行遮挡的算法。为了评估算法的性能,将其与 ResNet(Residual Network)[19]、WRN(Wide Residual Networks)[20]、ResNeXt[21]和 Xception[22]等网络结合,在 Cifar[6]、Fashion-MNIST[23]和胎盘组织细胞图像[24]等数据集上进行了实验,取得了非常有竞争力的实验结果。

本文的主要贡献如下:(1)提出了激活区域处理算法 AR(Activation Region processing algorithm),这是一种轻量计算的算法,以非常低的内存消耗和训练时间为代价,在不增加额外训练参数和不影响测试时间的情况下,很好地提升了多种卷积神经网络模型在多个数据集上的性能;(2)结合激活区域处理算法训练出来的模型,在随机遮挡图像识别方面具有很好的鲁棒性;(3)激活区域处理算法是现有数据增强和正则化方法的补充,在与现有数据增强算法、批归一化等方法结合时,本文提出的算法能进一步提高模型的性能;(4)激活区域处理算法能够提高模型深层卷积层的激活强度,这表明算法能够鼓励模型更好地利用图像的全局特征进行决策,而不是依赖少数局部区域的激活特征来决策。

2　卷积神经网络

卷积神经网络主要由输入层、卷积层、池化层、全连接层和输出层构成,如图 1 所示。卷积层主要使用指定数量和指定感受野大小的卷积核对输入图像或上一层的输出特征进行卷积操作,计算整个卷积核和输入图像或特征图的相应位置的内积,并加上一个偏置项来提取相关图像特征图,再将提取的特征图输入至非线性激活函数上得到激活后的特征图并作为卷积层的输出。设第 i 层卷积层的输入记为 $\boldsymbol{x}^{i-1}$,输出为 $\boldsymbol{x}^i$,$\otimes$代表卷积,卷积核的参数权值为 $\boldsymbol{w}^i$,偏置项为 b^i,激活函数为 $\sigma^i(\cdot)$,则对应输出的激活特征 $\boldsymbol{x}^i$ 为:

$$\boldsymbol{x}^i = \sigma^i(\boldsymbol{x}^{i-1} \otimes \boldsymbol{w}^i + b^i)$$

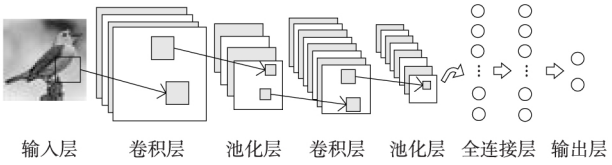


Figure 1　Convolutional neural network

图 1　卷积神经网络

通常在卷积神经网络中的连续卷积层之间周期性地插入池化层。池化层的作用主要有:(1)对激活特征图进行降维,减少网络中的参数数量和计算量;(2)保持特征尺度不变特性;(3)在一定程度上降低过拟合。池化层在输入的每个通道上独立操作,并在空间上调整其大小。池化方法主要有最大池化[1]、平均池化[25]、L_p 范数池化[26]等。记池化层的输入为 $\boldsymbol{x}^i$,则对应池化层的输出为:

$$\boldsymbol{x}^i_{\text{pool}} = subsampling(\boldsymbol{x}^i)$$

在对图像进行多次卷积池化操作后,卷积神经网络通过将三维激活特征图展开后得到的一维激活特征向量作为全连接层的输入,通过全连接层对特征进行分类,得到基于输入图像的概率分布。卷积神经网络的实质是通过多次数据变换或降维,将输入图像映射到一个新的数学模型。记卷积神经网络的输入样本为 $\boldsymbol{X}$,权值参数为 $\boldsymbol{W}$,偏置参数为 b,输出为 $\boldsymbol{Y}$,第 i 个类别标签记为 y_i,则将模型样本 x 预测为 y_i 的概率 $\bar{y}_i$ 为:

$$\bar{y}(i) = P(y_i \mid \boldsymbol{x};(\boldsymbol{W},b))$$

本文使用图像分类领域中常用的交叉熵函数作为损失函数。设损失函数为 $L(\boldsymbol{W},b)$,样本类别数为 N,则损失值的计算如式(1)所示:

$$L(\boldsymbol{W},b) = -\sum_{i=1}^{N} y_i \log \bar{y}_i \tag{1}$$

通过式(1)计算得到损失值后,卷积神经网络通过梯度下降法[27]对损失值进行反向传播,从输出层开始向输入层逐层更新卷积神经网络的可训练参数 $\boldsymbol{W}$ 和 b。设 x^i 是第 i 层输出,也是第 $i+1$ 层输入,学习率参数为 η,则由式(1)和求偏导链式法则可以得到反向传播过程中第 i 层求偏导的梯

度计算公式和参数调整公式分别如式(2)～式(4)所示:

$$\frac{\partial L(\boldsymbol{W},b)}{\partial \boldsymbol{x}^i}=\frac{\partial \boldsymbol{x}^n}{\partial \boldsymbol{x}^{n-1}}\cdots\frac{\partial \boldsymbol{x}^i}{\partial \boldsymbol{x}^{i-1}} \tag{2}$$

$$\boldsymbol{W}^i=\boldsymbol{W}^i-\eta\frac{\partial L(\boldsymbol{W},b)}{\partial \boldsymbol{W}^i} \tag{3}$$

$$b^i=b^i-\eta\frac{\partial L(\boldsymbol{W},b)}{\partial b^i} \tag{4}$$

3 激活区域处理算法 AR

设网络的输入图像为 $\boldsymbol{I}$,维度为 $C\times H\times W$。选取第 i 层卷积层的输出 $\boldsymbol{x}^i$,维度为 $C^i\times H^i\times W^i$。使用双线性插值(Bilinear Interpolation)方法对 $\boldsymbol{x}^i$ 上采样得到与输入图像相同尺寸的特征图 $\boldsymbol{x}=upsampling(\boldsymbol{x}^i)$,维度为 $C^i\times H\times W$,即得到 C^i 幅 $H\times W$ 大小的特征图,然后比较特征图的最大像素值,从中取像素值最大的特征图作为最大激活特征图 $\boldsymbol{C}_{\max}$。计算 $\boldsymbol{C}_{\max}$ 的均值 $mean(\boldsymbol{C}_{\max})$ 与标准差 $std(\boldsymbol{C}_{\max})$,设用来调整阈值大小的参数为 $\lambda\in[0,1]$,则阈值选取公式如式(5)所示:

$$T=mean(\boldsymbol{C}_{\max})+\lambda\times std(\boldsymbol{C}_{\max}) \tag{5}$$

设图像掩膜为 $\boldsymbol{M}$,维度为 $H\times W$,取最大激活特征图 $\boldsymbol{C}_{\max}$ 中的每一点 $C_{(m,n)}$ 与阈值 T 进行比较,其中,$m\in[0,H)$,$n\in[0,W)$,如果特征图中某一点的值大于阈值 T,则图像掩膜中对应点的值设为 0,如果特征图中某点的值小于阈值 T,则图像掩膜中对应点设为 1,即对图像掩膜 $\boldsymbol{M}$ 中的任意一点 $M_{(m,n)}$ 的计算公式如式(6)所示:

$$M_{(m,n)}=\begin{cases}0, C_{(m,n)}\geqslant T\\ 1, C_{(m,n)}<T\end{cases} \tag{6}$$

其中,图像掩膜中值为 0 的点组成的区域即为遮挡区域。由于 $\boldsymbol{C}_{\max}$ 中大于阈值 T 的点并不一定相邻,因此遮挡区域不一定是连续的,并且是不规则的。遮挡区域 $\boldsymbol{O}$ 可以表示如下:

$$\boldsymbol{O}=\{(m,n)\mid C_{(m,n)}\geqslant T\} \tag{7}$$

算法 1 详细描述了在卷积神经网络中使用激活区域处理算法的过程。图 2 展示了在不同数据集和不同网络结构上基于激活区域处理算法对输入图像进行遮挡的效果。受降噪自动编码器研究的启发,考虑使用不同噪声值对被遮挡区域进行填充:(1)使用 0 对遮挡区域填充,记为 Fill-0;(2)使用 1 对遮挡区域填充,记为 Fill-1;(3)对于 RGB 彩色图像,使用 ImageNet 数据集的 RGB 各通道的平均像素值[0.4902,0.4784,0.4471]对遮挡区域填充;对于灰度图像,使用 0.478 4 对遮挡区域填充,记为 Fill-I;(4)使用[0,1]中的随机噪声对遮挡区域填充,记为 Fill-R。

算法 1 激活区域处理算法

输入:输入图像 $\boldsymbol{I}$,图像的宽 W 和高 H,图像随机遮挡的概率 P,参数 λ,前一个训练周期中以图像 $\boldsymbol{I}$ 为输入的第 i 层卷积层的输出特征图 $\boldsymbol{x}_i$,当前训练周期数 e。

输出:被遮挡的图像 $\boldsymbol{I}'$。

1. $p=rand(0,1)$;
2. if $p\geqslant P$ or $e==0$ then
3. $\boldsymbol{I}'=\boldsymbol{I}$
4. else
5. $\boldsymbol{x}^i_{\text{up}}=upsampling(\boldsymbol{x}^i)$;
6. $\boldsymbol{C}_{\max}=\max(\boldsymbol{x}^i_{\text{up}})$;
7. $T=mean(\boldsymbol{C}_{\max})+\lambda\times std(\boldsymbol{C}_{\max})$;
8. $M_{(m,n)}=\begin{cases}0, C_{(m,n)}\geqslant T\\ 1, C_{(m,n)}<T\end{cases}$;
9. $\boldsymbol{I}'=\boldsymbol{I}\times\boldsymbol{M}$;
10. end if
11. return $\boldsymbol{I}'$

4 实验结果分析

4.1 对比实验的设置

我们在不同数据集上比较了使用和不使用激

Figure 2 Visualization of occlusion of Cifar 10 using the activation region processing algorithm while training ResNet-18

图 2 在训练 ResNet-18 时使用激活区域处理算法对 Cifar 10 图像进行遮挡的可视化

活区域处理算法训练的卷积神经网络模型。为确保实验结果的有效性,各数据集的训练集和测试集完全分开,首先使用训练集对网络进行训练,然后使用测试集进行测试。对于相同的神经网络结构和数据集,我们使用相同的参数设置、初始化方法和训练步骤进行训练,以在同等训练情形下对比加入本文算法前后的实验结果。我们通过计算模型的分类错误率来衡量模型性能,理想结果是错误率等于0。为了减少随机因素的影响和保证实验结果的准确性,本文所展示的实验结果是5次实验结果的平均错误率和标准差。实验平台、数据集、网络结构、参数设置详细介绍如下:

(1)实验平台:本文实验使用的硬件设备为Intel Xeon(R) CPU E5-2620 v3 2.40 GHz,NVIDIA Tesla K80 (12G)。本文实验所用操作系统为Ubuntu 16.04,采用Python 3.6作为编程语言,使用Facebook开源的Pytorch 1.0.0深度学习框架进行算法编码。

(2)数据集:Cifar10和Cifar100是深度学习领域常用的自然图像分类数据集之一,它们都包含50 000幅图像的训练集和10 000幅图像的测试集,每幅图像为32×32 pixel的RGB彩色图像,分别包含10和100个类别。Fashion-MNIST是一个包含10种类别的服装图像数据集,其中训练集包含60 000幅图像,测试集包含10 000幅图像,每幅图像均是28×28 pixel的灰度图像。

(3)网络结构:ResNet-18、WRN-28-10和ResNeXt-8-64。

(4)参数设置:使用Nesterov动量的随机梯度下降法(Stochastic Gradient Descent)对各网络进行端到端的训练,动量参数设置为0.9,权值衰减系数为0.000 5,小批量大小为128,使用交叉熵函数计算损失值,通过反向传播算法逐层传播损失和更新网络参数。对于相同的数据集,所有模型均采用相同的数据归一化和数据增强方法进行预处理。通过计算数据集各通道的平均值、标准差,然后将数据集图像各通道减去平均值再除以标准差得到归一化后的图像。训练200个周期,初始学习率设置为0.1,并在80,120,160个周期后依次将学习率减少5倍。在训练过程中对这3个数据集执行相同的数据增强操作,首先对图像每侧进行4个像素的0填充,然后随机裁剪出与输入图像大小相同的图像,最后进行随机水平翻转。

4.2　参数选取

由于CNN包含多个卷积块,因此我们在相同超参数设置下基于ResNet-18的不同卷积块在Cifar10上提取的特征图进行了实验。表1展现了ResNet-18的详细结构。分别选取block1、block2、block3、block4的输出作为激活区域处理算法的输入来提取最大特征图和制作图像掩膜。

Table 1　ResNet-18 network structure and experimental results of using different convolutional layer output feature maps to create image mask

表1　ResNet-18的网络结构以及选取不同卷积层输出特征图制作图像掩膜的实验结果

ResNet-18	Layer	Output size	Error rate/%
conv1	3×3,64,Stride 1	32×32×64	-
block1	$\begin{bmatrix}3\times3,64\\3\times3,64\end{bmatrix}\times2$	32×32×64	4.19±0.06
block2	$\begin{bmatrix}3\times3,128\\3\times3,128\end{bmatrix}\times2$	16×16×128	3.86±0.04
block3	$\begin{bmatrix}3\times3,256\\3\times3,256\end{bmatrix}\times2$	8×8×256	**3.63±0.07**
block4	$\begin{bmatrix}3\times3,512\\3\times3,512\end{bmatrix}\times2$	4×4×512	3.96±0.05
fc	global average pool, 10 classes,softmax	1×1	-

从表1可以看出,激活区域处理算法在以block3的特征图作为输入的情况下在Cifar10上得到了更好的结果。将block2、block3、block4输出的最大特征图上采样至32×32后进行可视化对比如图3所示。从图3中可以看出,较浅层的block2提取出的特征更多的是零散的边缘特征,较深层次的block3提取出来的是较为关键的局部特征,而更深层次的block4更多地利用了全局特征信息。在图像分类过程中,卷积神经网络中的浅层学习到的是图像中的部分边缘特征信息,而更深层根据浅层得到的边缘特征信息进一步学习,得到全局特征信息,最后进行分类决策。

对于本文提出的算法而言,block2提取的零散边缘特征信息和block4提取的全局特征信息并不利于制作图像掩膜。因为零散边缘特征信息会导致制作出来的掩膜的遮挡区域相对零散,达不到很好的遮挡效果,全局特征信息会导致制作出来的掩膜的遮挡区域过大,使得原始图像信息损失过多。我们希望遮挡图像的一部分关键的局部特征信息,以迫使网络学习更多的特征并提高网络的泛化能力,因此我们在ResNet-18网络中选取block3的输出作为算法的输入。WRN-28-10和ResNeXt-8-64均没有block4卷积模块,为了避免

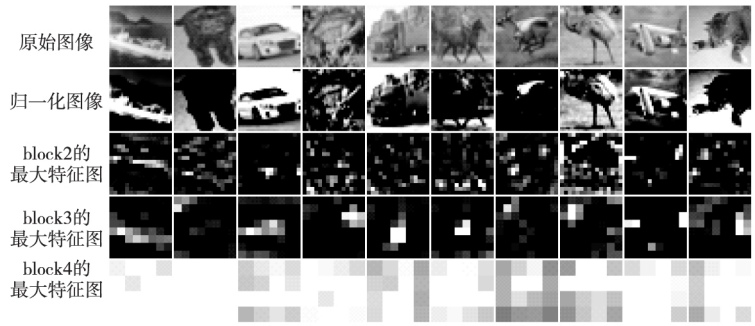


Figure 3 Visualization of different classes of Cifar10 extraction features in different convolutional layers in ResNet-18

图 3 在 ResNet-18 中不同卷积层对不同类别的 Cifar10 图像提取特征的可视化

使用最后一个卷积层的输出来制作图像掩膜,我们选取 WRN-28-10 和 ResNeXt-8-64 的 block2 的输出来作为算法的输入。

激活区域处理算法有 3 个超参数需要选取,分别是随机遮挡概率、阈值、填充方式。为了评估超参数对算法性能的影响,在不同的超参数设置下基于 ResNet-18 对 Cifar10 进行了实验。在评估其中一个参数时,其他参数保持不变。实验结果如表 2 和图 4 所示,当 $P=0.4$,$\lambda=0.3$,方式填充为 Fill-R 时,算法的性能最好。值得注意的是,算法在不同参数设置下的实验结果均优于 ResNet-18 在 Cifar10 上的基线结果。

Table 2 AR error rate using different filling methods for occlusion

表 2 AR 算法使用不同填充方式进行遮挡的错误率

Method	填充方式			
	Fill-0	Fill-1	Fill-I	Fill-R
AR Error Rate/%	3.95±0.09	3.91±0.07	3.88±0.10	**3.63±0.07**

4.3 实验结果对比

表 3 中显示了在使用数据增强的情况下,本文算法与不同网络结合后在不同数据集上的实验结果。与现有的一些研究结果相比,本文算法在 Cifar10、Cifar100 和 Fashion-MNIST 数据集上分别得到了 3.11%,17.44% 和 3.97% 的更低的错误率。对于 ResNet-18、WRN-28-10、ResNeXt-8-64 这 3 个基础网络而言,未加任何遮挡的基线模型和添加随机遮挡的模型的错误率均高于添加本文算法后的模型的错误率。同一网络模型在 Cifar10、Cifar100 和 Fashion-MNIST 上比基线最多降低了约 0.71%,1.04% 和 1.77% 的错误率。值得注意的是,本文提出的算法不仅适用于彩色图像数据集 Cifar,而且还适用于灰度图像 Fashion-MNIST。这说明本文算法能够有效地提升不同网络结构在不同数据集上的分类性能。

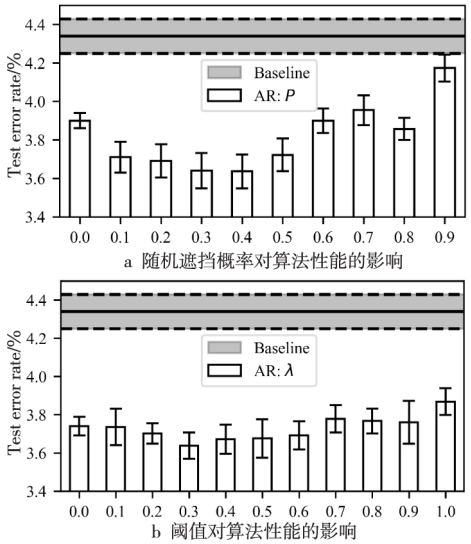


Figure 4 Effect of different hyperparameters on algorithm performance

图 4 不同超参数对算法性能的影响

为了对比采用 AR 算法训练的模型和采用随机遮挡图像方法训练的模型的性能,本文进行了以下实验:在训练过程中对图像随机添加不同大小的遮挡,随机遮挡区域面积与原始图像面积($H\times W$)比值为 $s\in[0,0.5]$,从 $[0.5,1.5]$ 范围中随机选取遮挡区域面积的宽高比 R,则遮挡区域的宽高为:$H'=\sqrt{s\times H\times W\div R}$,$W'=\sqrt{s\times H\times W\times R}$,并用 $[0,1]$ 内的随机值填充遮挡区域。实验结果如表 3 所示。

从表 3 可以看出,采用随机遮挡图像方法训练的模型也比基线模型的性能更好。与 AR 算法相比,AR 算法对模型的性能提升更加显著。由于随机遮挡方法中遮挡区域位置是随机不定的,其功能更多的是增加了训练图像的多样性,并不一定能遮挡关键区域来降低模型对局部区域的依赖。AR 算法对关键区域的遮挡可能是带来更好性能的关键。

Table 3　Error rate comparison with different baselines and current new methods

表 3　与当前一些新方法的错误率对比　%

Method	Year	Datasets		
		Cifar10	Cifar100	Fashion-MNIST
ResNet-1001[19]	2016	4.69±0.20	22.68±0.22	-
WRN-40-10[20]	2016	3.8	18.3	-
ResNeXt-8-64[21]	2017	3.65	17.77	-
Random Erasing[13]	2017	3.08±0.05	17.73±0.15	4.2±0.03
DenseNet-40-FRN[28]	2018	4.95	23.36	-
RoR optimization method[29]	2018	3.52	19.07	-
ResNet-18(基线)	2019	4.34±0.08	21.96±0.13	5.86±0.05
ResNet-18+随机遮挡	2019	3.85±0.06(↓0.49)	21.33±0.10(↓0.63)	4.45±0.07(↓1.41)
ResNet-18+AR	2019	3.63±0.07(↓0.71)	21.16±0.12(↓0.80)	4.09±0.09(↓1.77)
WRN-28-10(基线)	2019	3.70±0.11	18.48±0.21	5.50±0.08
WRN-28-10+随机遮挡	2019	3.38±0.10(↓0.32)	17.72±0.14(↓0.76)	4.23±0.09(↓1.27)
WRN-28-10+AR	2019	**3.11±0.07(↓0.59)**	**17.44±0.17(↓1.04)**	**3.97±0.07(↓1.53)**
ResNeXt-8-64(基线)	2019	3.68±0.07	18.72±0.23	5.43±0.05
ResNeXt-8-64+随机遮挡	2019	3.35±0.08(↓0.33)	18.49±0.18(↓0.23)	4.36±0.05(↓0.71)
ResNeXt-8-64+AR	2019	3.12±0.05(↓0.46)	18.33±0.17(↓0.39)	4.07±0.10(↓1.36)

4.4　不使用数据增强下的性能

由表 4 可知,在不使用数据增强的情况下,本文算法均降低了 ResNet-18、WRN-28-10 网络在 Cifar10、Cifar100 数据集上的分类错误率。结合表 3 和表 4 可知,本文算法与数据增强算法结合后的效果更好,并且可以将算法看做是一种新型的数据增强算法。

Table 4　Error rate comparison with the proposed algorithm and different baselines without using data augmentation.

表 4　在不使用数据增强的情况下,本文算法与不同基线的错误率对比　%

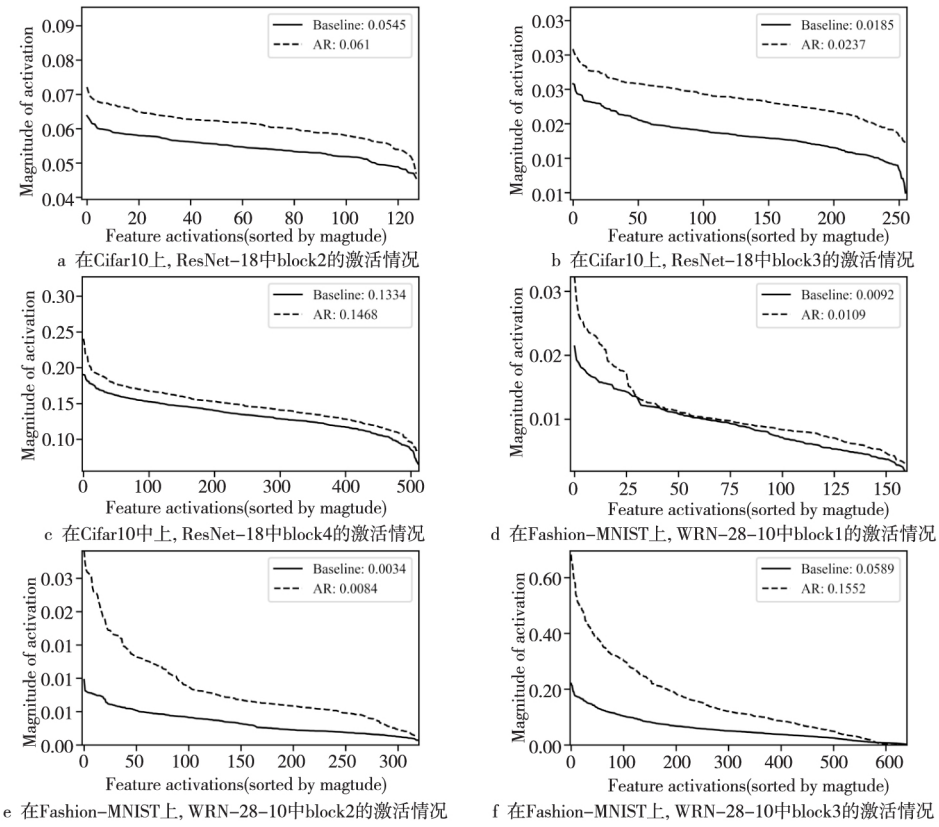
Mehtod	Datasets	
	Cifar10	Cifar100
ResNet-18	10.17±0.04	35.92±0.16
ResNet-18+AR	9.47±0.07±(↓0.70)	35.23±0.11±(↓0.69)
WRN-28-10	6.67±0.11	25.19±0.14
WRN-28-10+AR	6.16±0.14(↓0.51)	24.81±0.11(↓0.38)

4.5　对卷积神经网络的影响

为了更好地了解算法对卷积神经网络产生的影响,进行了以下对比实验:随机地从测试集中采样出 128 幅图像,作为训练好的模型的输入,从中提取出某一卷积块的输出特征图;然后计算每幅特征图的平均值作为激活值,并对于同一幅图像的激活值进行降序排序;为了减少随机采样带来的影响,最后计算这 128 幅图像对应特征激活值的平均值。对于 ResNet-18,在 Cifar10 测试集上提取 block2、block3 和 block4 这 3 个模块的输出特征图。对于 WRN-28-10,在 Fashion-MNIST 测试集上提取 block1、block2 和 block3 这 3 个模块的输出特征图。最后对比结果如图 5 所示。从图 5 可以观察到,使用 AR 算法训练的模型在不同卷积层的激活值均强于基线模型的。值得注意的是,AR 算法训练的模型的激活值与基线模型的激活值的比也随着卷积层的加深而提高。这表明 AR 算法确实鼓励模型学习更多的特征信息,并且利用更多的激活特征进行决策,而不仅仅是依赖于少量激活特征进行决策。

4.6　对随机遮挡的鲁棒性

为了验证 AR 算法训练的模型对随机遮挡图像识别的鲁棒性,本文进行了以下实验:对 Cifar10 和 Fashion-MNIST 的测试集图像随机添加不同大小的遮挡,随机遮挡区域面积与原始图像面积($H\times W$)比为 $s\in[0,0.5]$,从[0.5,1.5]范围中随机选取遮挡区域面积的宽高比 R,则遮挡区域的宽高为:$H'=\sqrt{s\times H\times W\div R}$,$W'=\sqrt{s\times H\times W\times R}$,并用[0,1]随机值填充遮挡区域,然后使用训练好的 ResNet-18 和 WRN-28-10 对遮挡测试集进行实验。对遮挡测试集进行分类后的错误率如图 6 所示。从图 6 可以看出,使用 AR 算法训练的 Res-



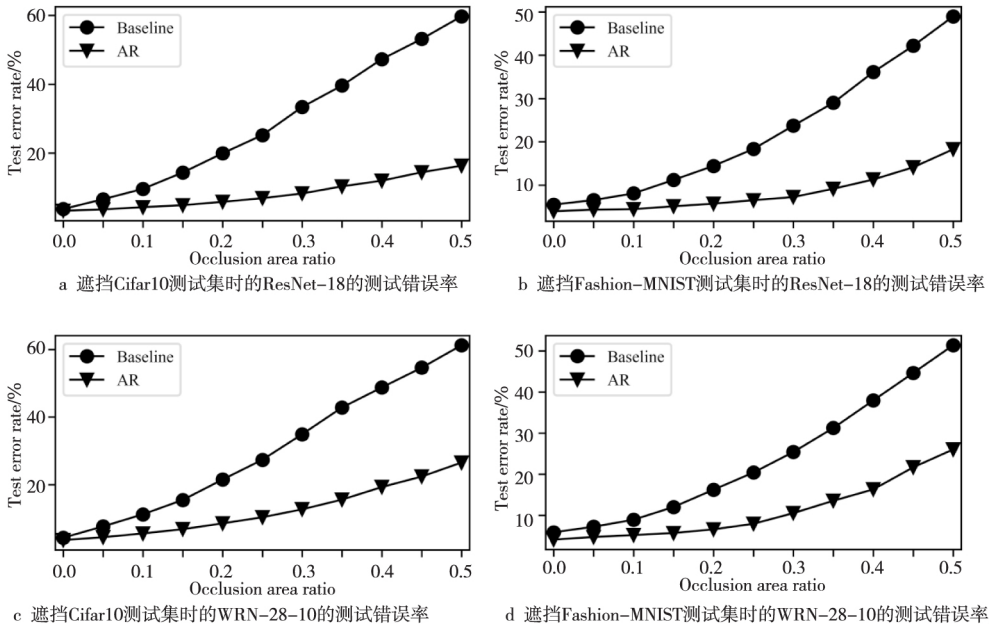
a 在Cifar10上, ResNet-18中block2的激活情况 b 在Cifar10上, ResNet-18中block3的激活情况

c 在Cifar10中上, ResNet-18中block4的激活情况 d 在Fashion-MNIST上, WRN-28-10中block1的激活情况

e 在Fashion-MNIST上, WRN-28-10中block2的激活情况 f 在Fashion-MNIST上, WRN-28-10中block3的激活情况

Figure 5 Compare different convolutional layers activation of models trained using different methods

图 5 比较使用不同方式训练的模型的不同卷积层的激活情况



a 遮挡Cifar10测试集时的ResNet-18的测试错误率 b 遮挡Fashion-MNIST测试集时的ResNet-18的测试错误率

c 遮挡Cifar10测试集时的WRN-28-10的测试错误率 d 遮挡Fashion-MNIST测试集时的WRN-28-10的测试错误率

Figure 6 Robustness of random occlusion images

图 6 对随机遮挡图像的鲁棒性

Net-18 和 WRN-28-10 模型在不同遮挡情况下的性能均优于基线模型的。

4.7 算法复杂性分析

为了提高在实际应用中的可行性和可扩展性，必须权衡算法复杂性和性能。为了说明 AR 算法的成本，以 ResNet-18 和 WRN-28-10 为例，在 Cifar10 上进行稳定训练时，比较模型的物理内存占用、GPU 内存占用和每周期的训练时间，比较结果如表 5 所示。由表 5 可知，结合 AR 算法后的模型与基线模型占用的物理内存大致相当，这是由于本

文算法的运算全部在 GPU 上。比较 GPU 内存的占用情况可知,结合 AR 算法后的模型比基线模型占用的 GPU 内存多了不到100 Mb,并且不会随着模型的变动有太大的变化。比较每周期的训练时间,结合 AR 算法后的模型比基线模型每周期的训练时间多了不到 10%。根据算法的设计,AR 算法自身并不会参与网络内部的传播运算,不涉及梯度等复杂的计算,算法额外需要的内存和计算时间取决于数据集的大小和输入特征图的尺寸大小。由于 AR 算法并不会改变模型的结构和带来额外的训练参数,因此训练好的模型与基线模型在文件大小和对测试集进行测试的时间相当。通过以上分析可知,AR 算法在空间上带来的代价是非常小的。

Table 5　Comparison of physical memory, GPU memory, and training time per epoch of the model

表 5　模型的物理内存、GPU 内用和每周期的训练时间比较

Method	Memory/Mb	GPU memory/Mb	Time/s
ResNet-18	2 464	1 543	79
ResNet-18+AR	2 477	1 615	86
WRN-28-10	2 575	4 721	412
WRN-28-10+AR	2 584	4 807	431

4.8　与预训练模型相结合

为了进一步证明本文算法具有良好的泛化性,使用预训练模型 Xception[22] 在一个全新的胎盘组织细胞图像数据集上进行微调分类实验。使用 Ferlaino 等[24] 提供的包含 5 个类别的胎盘组织细胞图像数据集,其中训练集有7 529幅图像,测试集有 1 000 幅图像(每个类别各有 200 幅),验证集有 1 000 幅图像(每个类别各有 200 幅),所有图像均为 200×200 pixel 的 RGB 彩色图像。使用训练集进行微调,验证集和测试集均不参与微调训练模型。移除原 Xception 网络的全连接层,使用全局平均池化对卷积层输出进行池化,然后构建新的全连接层:2048fc→128fc→5fc,层与层之间使用丢失率为 0.5 的 Dropout 层。使用随机梯度下降法对 Xception 进行端到端的微调,学习率为 0.01,小批量大小为 16,损失函数为交叉熵函数。将图像上采样至 299×299 pixel 以保持与原 Xception 网络的输入一致,并使用 Xception 的 block3 模块的输出(37×37×256)作为本文算法的输入。由表 6 的结果可知,使用 AR 算法能够显著改进微调 Xception 模型的性能,在验证集上获得了 1%的性能增益,在测试集上获得了1.5%的性能增益。

Table 6　Experimental classification results of placental tissue cell images using pre-trained Xception model

表 6　使用预先训练的 Xception 模型对胎盘组织细胞图像的实验分类结果

Method	Validation accuracy/%	Test accuracy/%
InceptionV3[24]	90	88
InceptionResNetV2[24]	91	87
Xception[24]	91	87
Ensemble (Max)[24]	91	89
Xception+AR	92.0±0.2	90.5±0.3

5　结束语

降低网络模型过拟合的风险和提升模型在遮挡图像识别方面的泛化能力是深度学习方向卷积神经网络的研究重点。本文提出了激活区域处理算法,用于处理卷积神经网络某一层卷积层的最大激活特征图,以实现对输入图像进行遮挡,并将遮挡后的图像作为网络的新输入来继续训练网络。这种算法可以看成是数据增强算法的新形式,不仅可以降低网络模型过拟合的风险,而且还能够提升网络模型的性能。此外,使用这种算法训练的网络模型在识别随机遮挡图像方面也具有很好的鲁棒性。未来的工作中,我们会将这些方法应用于其他的计算机视觉领域中,例如图像分割、目标检测等。

参考文献:

[1] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks[C]// Proc of the International Conference on Neural Information Processing Systems, 2012: 1097-1105.

[2] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39(4): 640-651.

[3] Vinyals O, Toshev A, Bengio S, et al. Show and tell: A neural image caption generator[C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3156-3164.

[4] Toshev A, Szegedy C. DeepPose: Human pose estimation via deep neural networks[C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1653-1660.

[5] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[6] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images[R]. Toronto: University of Toronto, 2009.

[7] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A sim-

ple way to prevent neural networks from over-fitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.

[8] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[9] Wan L, Zeiler M, Zhang S, et al. Regularization of neural networks using dropconnect[C]// Proc of Machine Learning Research, 2013: 1058-1066.

[10] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]// Proc of the International Conference on Machine Learning, 2015: 1-11.

[11] Tompson J, Goroshin R, Jain A, et al. Efficient object localization using convolutional networks[C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 648-656.

[12] Park S, Kwak N. Analysis on the dropout effect in convolutional neural networks[C]// Proc of Asian Conference on Computer Vision, 2016: 189-204.

[13] Zhong Z, Zheng L, Kang G, et al. Random erasing data augmentation[J]. arXiv: 1708.04896, 2017.

[14] Li Xiao-xin, Liang Rong-hua. A review for face recognition with occlusion: From subspace regression to deep learning [J]. Chinese Journal of Computers, 2018, 41(1): 177-207. (in Chinese)

[15] Liu Wan-jun, Dong Shuai-han, Qu Hai-cheng. Anti-occlusion visual tracking algorithm based on spatio-temporal context learning[J]. Journal of Image and Graphics, 2016, 21(8): 1057-1067. (in Chinese)

[16] Chu Jun, Zhu Tao, Miao Jun, et al. Target tracking based on occlusion detection and spatio-temporal context information [J]. Pattern Recognition and Artificial Intelligence, 2017, 30(8): 718-727. (in Chinese)

[17] Vincent P, Larochelle H, Lajoie I, et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion[J]. Journal of Machine Learning Research, 2010, 11(12): 3371-3408.

[18] Pathak D, Krahenbuhl P, Donahue J, et al. Context encoders: Feature learning by inpainting[C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2536-2544.

[19] He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks[C]// Proc of the European Conference on Computer Vision, 2016: 630-645.

[20] Zagoruyko S, Komodakis N. Wide residual networks[J]. arXiv: 1605.07146, 2016.

[21] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 5987-5995.

[22] Chollet F. Xception: Deep learning with depthwise separable convolutions[J]. arXiv preprint arXiv: 1610-02357, 2016.

[23] Xiao H, Rasul K, Vollgraf R. Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms [J]. arXiv: 1708.07747, 2017.

[24] Ferlaino M, Glastonbury C A, Motta-Mejia C, et al. Towards deep cellular phenotyping in placental histology[J]. arXiv: 1804.03270, 2018.

[25] Lin M, Chen Q, Yan S. Network in network[J]. arXiv: 1312.4400, 2013.

[26] Feng J, Ni B, Tian Q, et al. Geometric-norm feature pooling for image classification[C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition, 2011: 2609-2704.

[27] Bengio Y. Learning deep architectures for AI[J]. Foundations and Trendsin Machine Learning, 2009, 2(1): 1-127.

[28] Lu B, Hu Q, Hui Y, et al. Feature reinforcement network for image classification[C]// Proc of the IEEE International Conference on Multimedia and Expo, 2018: 1-6.

[29] Zhang K, Guo L, Gao C. Optimization method of residual networks of residual networks for image classification[C]// Proc of the Big Data and Smart Computing, 2018: 321-325.

附中文参考文献:

[14] **李小薪,梁荣华. 有遮挡人脸识别综述:从子空间回归到深度学习[J]. 计算机学报**, 2018, 41(1): 177-207.

[15] **刘万军,董帅含,曲海成. 时空上下文抗遮挡视觉跟踪[J]. 中国图象图形学报**, 2016, 21(8): 1057-1067.

[16] **储珺,朱陶,缪君,等. 基于遮挡检测和时空上下文信息的目标跟踪算法[J]. 模式识别与人工智能**, 2017, 30(8): 718-727.

作者简介:

蒋芸(1970-),**女,浙江绍兴人,博士,教授,研究方向为数据挖掘,粗糙集理论及应用**。**E-mail**: jiangyun@nwnu.edu.cn

JIANG Yun, born in 1970, PhD, professor, her research interests include data mining, and rough set theory & applications.

张海(1995-),**男,江西赣州人,硕士,CCF 会员**(91466G),**研究方向为数据挖掘**。**E-mail**: haicheung1995@gmail.com

ZHANG Hai, born in 1995, MS, CCF member(91466G), his research interest includes data mining.