

反向目标干扰的图像数据增强^①

袁 姮¹, 胡 月¹, 张晟翀²

¹(辽宁工程技术大学 软件学院, 葫芦岛 125105)

²(光电信息控制和安全技术重点实验室, 天津 300308)

通信作者: 胡 月, E-mail: hyue0720@163.com



摘 要: 混合样本数据增强方法只注重模型对于图像所属类别的正向表达, 而忽略图像是否属于某一类别的反向判定. 为了解决描述图像类别方式单一而影响模型性能的问题, 提出一种反向目标干扰的图像数据增强方法. 该方法增加图像背景及目标的多样性, 防止网络模型过拟合. 其次采用反向学习机制, 让网络模型在正确辨别原图像所属类别的同时, 对填充图像不属于该类别的属性进行充分学习, 从而增强网络模型对原图像所属类别辨识的置信度. 最后, 为验证该方法的有效性, 使用不同的网络模型在 CIFAR-10、CIFAR-100 等 5 个数据集上进行大量实验. 实验结果表明, 本文方法与其他先进的数据增强方法相比较, 可以显著提高模型在复杂背景下的学习效果和泛化能力.

关键词: 图像数据增强; 混合样本; 反向学习; 目标多样化; 背景多样化

引用格式: 袁姮, 胡月, 张晟翀. 反向目标干扰的图像数据增强. 计算机系统应用, 2024, 33(6):48-57. <http://www.c-s-a.org.cn/1003-3254/9507.html>

Inverse Target Interference for Image Data Augmentation

YUAN Heng¹, HU Yue¹, ZHANG Sheng-Chong²

¹(Software College, Liaoning Technical University, Huludao 125105, China)

²(Science and Technology on Electro-optical Information Security Control Laboratory, Tianjin 300308, China)

Abstract: Mixed sample data enhancement methods focus only on the model's forward representation of the category to which the image belongs while ignoring the reverse determination of whether the image belongs to a specific category. To address the problem of uniquely describing image categories and affecting model performance, this study proposes a method of image data augmentation with inverse target interference. To prevent overfitting of the network model, the method first modifies the original image to increase the diversity of background and target images. Secondly, the idea of reverse learning is adopted to enable the network model to correctly identify the category that the original image belongs to while fully learning the attributes of the populated image that do not belong to that category to increase the confidence of the network model in identifying the category that the original image belongs to. In conclusion, to verify the method's effectiveness, the study utilizes different network models to perform many experiments on five datasets including CIFAR-10 and CIFAR-100. Experimental results show that compared to other state-of-the-art data augmentation methods, the proposed method can significantly enhance the model's learning effect and generalization ability in complex settings.

Key words: image data augmentation; mixed sample; inverse learning; target diversity; background diversity

1 引言

数据的数量和多样性对于深度学习的发展具有重要推动作用, 然而收集和标记数据是一项艰巨的任务.

因此, 为增加现有数据的多样性, 数据增强 (data augmentation, DA)^[1-3] 成为一种技术发展趋势, 其通过应用多种变换来实现. 近年来, 一系列混合样本数据增强方法

① 基金项目: 辽宁省自然科学基金 (20170540426); 辽宁省教育厅重点基金 (LJYL049)

收稿时间: 2023-12-03; 修改时间: 2023-12-29; 采用时间: 2024-01-12; csa 在线出版时间: 2024-04-19

CNKI 网络首发时间: 2024-04-23

(mixed sample data augmentation, MSDA)^[4,5]应运而生。区别于传统的单一数据增强方法,由于其充分考虑到不同类别样本间的邻近关系,在图像数据增强过程中取得了较好的增强效果。

为了提高模型性能,Zhang等人^[6]提出Mixup方法,对两幅图像根据随机权重进行全区域混合的操作来生成新样本。尽管这种简单的变换过程可以提高模型性能,但需要手动调整混合系数,操作十分不便。此外,该方法还会导致局部纹理特征模糊,引入一些不自然的伪像素信息。为减少伪像素信息对网络模型的影响,Yun等人^[7]提出CutMix方法,该方法通过将其图像的某一部分进行裁剪并填充到原图像相同位置,以此创建新样本进行训练。然而,Uddin等人^[8]认为用另一个补丁填充图像的随机区域并不能保证补丁具有丰富的信息,为解决此问题,Uddin等人^[8]提出一种名为SaliencyMix方法,该方法将补丁图像中的显著性区域填充到原图像的非显著性区域,以在生成的样本中获得多个图像的显著性特征。然而,这种方法需要额外的计算开销来寻找图像的显著性区域,实际应用效果不佳。

在2020年,Bochkovski等人^[9,10]提出一种名为Mosaic的增强方法。与现有数据增强方法不同,Mosaic通过将4幅随机切分的图像拼接在一起,生成一幅全

新的图像。同时,Mosaic方法根据4幅图像的面积比例生成新标签,并将这幅新图像和标签传入网络模型中进行训练。然而,Mosaic方法存在一个问题,即部分图像的特征区域会在切分过程中丢失,容易误导网络模型对图像类别的学习,影响网络模型性能。尽管现有的混合样本数据增强方法在网络模型中对描述图像所属某类别方面起到至关重要的作用,但它们无法提供关于其他混合图像不属于该类别的信息,不会促使网络模型对图像类别进行反向学习,从而影响模型的学习效果。

本文方法利用训练数据对原图像进行填充,以增加图像目标多样性,并弥补现有数据增强方法中目标相对简单的问题。同时,该方法通过反向学习机制,充分利用图像的类内相似性和类间差异性进行拼接,生成一幅图像将其作为训练数据的新样本,使网络模型正确学习图像的类别并反向学习填充图像不属于该类别。本文方法通过提高图像背景复杂度干扰网络模型学习,提高网络模型的学习效果和泛化能力。最后,在CIFAR-10、CIFAR-100、Fashion-MNIST、Imagenette和Imagewoof这5个数据集上进行了实验,并与多个模型进行对比,其分类准确率平均提升3.1%。不同数据增强方法生成的样本如图1所示。

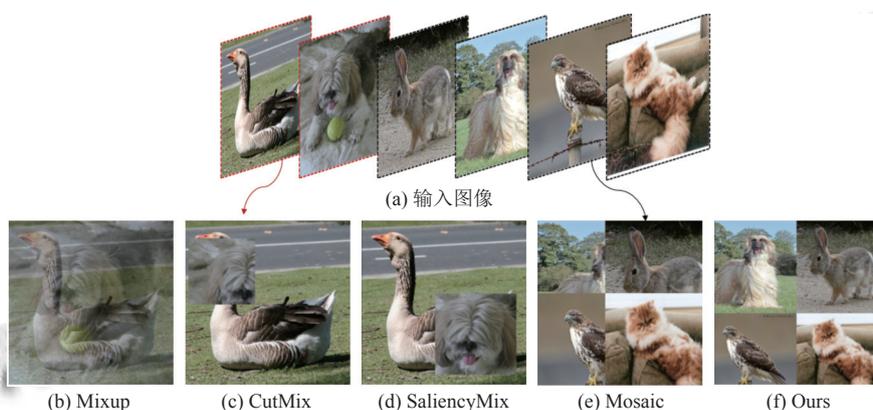


图1 不同数据增强方法样本示意图

2 相关工作

2.1 数据增强

深度卷积神经网络(convolutional neural network, CNN)^[11]在计算机视觉、自然语言处理、语音识别和信号处理等^[12]人工智能任务中获得显著成功的原因之一是因为有大量可学习参数。然而,过多的参数可能会导致网络模型过度拟合。因此,需要使用数据增强技术

增加训练数据量,避免模型过度拟合。此外,模型在处理不同分布的训练数据时,其泛化能力受到严重限制,因此数据增强技术是提高模型泛化能力的重要方法之一。

在传统数据增强领域,可以使用旋转^[13]、裁剪^[14]、颜色抖动^[15]、像素域融合^[16]等方法。这些方法会导致图像失真并存在数据高度冗余,对数据增强方法在实际应用中的发展具有一定影响。2017年,DeVries等人^[17]

提出 Cutout 方法,该方法在图像中随机选择一个裁剪区域,并将该区域内的像素内容全部去除,形成遮挡,生成与原图不同的新样本。2018年, Cubuk 等人^[18]提出一种无监督的数据增强方法 AutoAugment,该方法定义了各种增强技术的搜索空间,并能自动地设计特定的数据增强方案。

最近, Kim 等人^[19]提出 PuzzleMix 方法,该方法旨在联合优化两个目标:选择最优掩模和最优混合计划,其目的是凸显两幅图像中最显著的特征,最大限度地提高训练样本中的显著性区域。Hong 等人^[20]提出的 StyleMix 方法,解决了不区分内容和样本特征的问题,该方法将图像的内容和样本特征进行详细分离处理,并且在数据集上展现出令人印象深刻的性能。Suzuki^[21]提出基于对抗策略的 TeachAugment 数据增强方法,该方法不需要大量调参,利用教师模型生成带有关键信息的样本,取得了较好的数据增强效果。

2.2 反向学习

反向学习 (opposition-based learning, OBL)^[22]是机器学习中的一个新概念。为提高优化方法、强化学习、人工神经网络和模糊系统等计算方法的性能, OBL 的概念被广泛应用。

Ventresca 等人^[23]通过引入反传递函数的概念来改进 BPTT 算法,该算法通过使用反向传播神经网络进行学习。无论是原始传递函数还是反传递函数,网络的传递函数都是基于概率估计。根据网络的误差函数,计算每个隐藏神经元的概率。如果网络误差较小,则相同传递函数的概率增加,否则它会被还原。

Tizhoosh 等人^[24]提出一种利用反模糊集的图像阈值分割方法。该方法将窗口视为对象中心的围绕区域,并逐步增加其大小。在每次迭代中,计算当前窗口内的暗像素作为模糊集的对偶,确定两个模糊集之间最大熵差的窗口位置,并根据所选窗口的两个模糊集代表数的平均值计算图像阈值。同时,提出一种逼近第 2 类相对点的方法^[25],该方法利用 Takagi-Sugeno 或 T-S 模糊推理系统进行演化模糊推理。

差分目标传播 (DTP) 能够在复杂任务中匹配反向传播性能的学习算法,但是 DTP 权值更新和损失梯度之间的良好一致性只能在非常特定的训练体系结构条件下得到部分保证。Ernoul 等人^[26]提出了一种新的反馈权重训练方案,其通过构造将反馈算子的雅可比矩阵推向其前馈对应部分的转置,以分层的方式,从而避

免了在反馈路径中使用直接反馈连接的需要。

最近出现一种基于反向多领导者的多目标粒子群优化方法^[27],其采用反向选择方法来选择具有良好收敛性的全局最佳值和具有良好多样性的全局最佳值。反向选择方法为每个粒子匹配合适的领导者,每个粒子选择合适的学习样本,使得粒子可以在几个迭代中更接近真实情况。

近年来,越来越多的研究工作将反向学习概念应用于机器学习,以应对模式识别与图像处理等多种领域的挑战。将反向学习用于优化方法,以解决图像处理和模式识别需要参数调整等问题。在各个应用领域中,最常见的反向学习方式是基于基本概念和准对立的學習。本文采用反向学习中对立的概念,增强网络模型对图像类别的判定能力。

3 本文方法

本文方法首先在训练数据中随机选取图像对原图像进行填充,丰富原图像背景和目标的多样性,其次选取中心点进行裁剪,突显原图像的特征信息,然后将裁剪后的图像送入网络模型中进行训练,利用反向学习机制使网络模型正确辨别原图像类别,并反向学习填充图像不属于此类别。

3.1 反向学习机制

现有数据增强方法针对图像类别采用正向表达,对网络模型的学习效果有一定影响,尤其在混合样本数据增强中对原图像类别的判定。为此,本文引入了反向学习机制,不仅让网络模型能够辨别原图像所属类别,还排除填充图像属于该类别的可能性。

Tizhoosh 在 2005 年提出反向学习^[28],这是机器学习中的一个概念,其灵感来源于现实世界中事物之间的相对关系,比如阴阳、冷热、干湿等。本文方法根据反向学习中相反数的定义,分析图像类间相似性和类内差异性实现对图像类别的准确区分,并有效地识别出不同的图像类别,以应对复杂的图像分类等问题。

设 $x \in [a, b]$ 为实数, x 的相反数 \bar{x} 的定义如下:

$$\bar{x} = a + b - x \quad (1)$$

利用相反数的定义,以一维空间为例,图 2 表示 x 的相反数为 \bar{x} 。



图 2 相反数

明确相反数的定义之后,将原图像 T_1 视为正样本,填充图像 T_2 、 T_3 、 T_4 视为负样本,把新生成的图像及其标签 (T_p, L_p) 送入到神经网络中,让其分辨原图像属于某一类别,判断填充图像的类别和原图像的类别是否为同一类别。

$$T_p = \{T_1, T_2, T_3, T_4\} \quad (2)$$

$$L_p = \{L_1, L_2, L_3, L_4\} \quad (3)$$

$$L_p \in L_1 \quad (4)$$

$$L_2, L_3, L_4 \notin L_1 \quad (5)$$

其中, L_1 为原图像的标签, L_2, L_3, L_4 为填充图像的标签. 新生成图像标签的组成部分为原图像标签 L_1 及填充图像 L_2, L_3, L_4 标签, 将填充图像标签信息融合到原图像标签中, 采用相反数的思想, 对填充图像标签进行反向表达, 促使网络模型更好理解原图像所属类别. 在网络模型训练过程中, 反向学习机制通过对图像标签的预测, 使用梯度下降法更新模型参数, 促使模型正确分辨原图像和填充图像之间的类别差异, 利用差异判断填充图像是否属于原图像类别, 进而加强对原图像所属类别学习的置信度并使模型一次学习到多种图像类别.

本文方法引入了更多图像来增加背景和目标的复杂度, 展示出类内相似性和类间差异性, 从而帮助网络模型更好地学习和理解某一类别的特征, 同时也能够让网络模型学习到不同类别的差异, 进而提升模型的学习效果和泛化能力.

3.2 图像混合

现有的混合样本数据增强方法只是采用两幅图像进行混合, 产生的训练样本数量及形式有限. 为了进一步提高样本丰富性, 本文方法在数据集中按批次选取图像, 每批图像包含 m 幅. 然后, 从批次中随机选取 4 幅图像进行操作. 设随机选取 4 幅图像的可能性为 N :

$$N = C_m^n \quad (6)$$

其中, n 为 4. 而随机选取两次, 两次图像相同的概率为 $P(N)$.

$$P(N) = \left(\frac{1}{C_m^n}\right)^2 \quad (7)$$

根据式 (6) 和式 (7), 本文方法利用随机选取 4 幅图像生成新样本, 增加训练样本的数量, 并且重复两次

选取到 4 幅相同图像生成样本的可能性极低, 进而避免网络模型过度拟合.

在本文方法中, 有 1 幅原图像和其他 3 幅填充图像, 其 4 幅图像按照四宫格的形式进行拼接. 为让原图像位置随机摆放, 设计一个概率因子 ρ , 其取值范围在 0 和 1 之间. 当 ρ 小于 0.25 时, 原图像所在位置为 T_1 , 摆放在四宫格的左上角; 当 ρ 在 0.25 与 0.5 之间时, 原图像所在位置为 T_2 , 摆放在右上角; 当 ρ 在 0.5 与 0.75 之间时, 原图像所在位置为 T_3 , 摆放在左下角; 当 ρ 大于 0.75 时, 原图像所在位置为 T_4 , 摆放在右下角.

本文方法通过对随机选取的 4 幅图像进行拼接, 确保生成的训练样本包含各种场景、角度和物体的组合, 让网络模型能够同时学习到不同图像, 从而加强网络模型对图像类别的判定能力; 利用其他图像丰富原图像的背景和目标, 增加网络模型对图像进行学习的干扰, 使得网络模型能够更好地适应各种较为复杂的图像场景.

3.3 中心点裁剪

由于数据集中的图像尺寸相同, 导致原图像和填充图像的信息一致. 因此, 需要对拼接后的图像进行裁剪, 突显原图像的特征信息. 首先, 设拼接后的图像尺寸为 (X, Y) , 在点 A 和点 B 规定的矩形内任意选取点 $C(x, y)$, 该点作为裁剪图像的中心点. 其次, 根据中心点 C 的位置决定裁剪图像的边长 x_i :

$$x_i = \min(x, X - x, y, Y - y) \quad (8)$$

其中, $X - x, Y - y$ 分别为点 C 到图像右侧和下侧的直线距离. 然后, 根据裁剪图像的边长 x_i 选取一个正方形, 该正方形由顶点坐标 (X_l, Y_b) 和 (X_r, Y_t) 共同确定, 顶点坐标的选取如下:

$$X_l = x - x_i, X_r = x + x_i \quad (9)$$

$$Y_b = y - x_i, Y_t = y + x_i \quad (10)$$

最后, 对选取的正方形进行规则裁剪, 得到最终的图像. 具体的中心点裁剪过程如图 3 所示.



图 3 中心点裁剪过程示意图

3.4 方法处理流程

反向目标干扰的图像数据增强方法主要处理步骤如下.

Step 1. 在总数据集中取出一个批次 (batch) 的数据 X , 在 batch 中提取一幅图像 T_1 作为原图像, 同时随机选取 3 幅图像 T_2, T_3, T_4 作为填充图像.

Step 2. 将 T_1, T_2, T_3, T_4 这 4 幅图像进行混合, 生成一个新的图像 Img , 其中 T_1 位置由概率因子 ρ 所决定.

Step 3. 根据 Img 的特定范围, 选择一个中心点 C ,

并截取一个边长为 x_i 的正方形, 将其作为新的学习样本 IMG .

Step 4. 将 IMG 送入到神经网络模型中利用反向学习机制进行学习, 使其能够识别该样本的类别, 并排除填充图像不属于该类别.

本文方法的流程如图 4 所示.



图 4 反向目标干扰的数据增强方法的流程示意图

4 实验分析

4.1 数据集及其指标

实验在 CIFAR-10, CIFAR-100, Fashion-MNIST, Imagenette 和 Imagewoof 这 5 个数据集上进行. CIFAR-10 数据集是主要用于识别普通物体, 其中的图像尺寸是 32×32 , 包含飞机、汽车、鸟类等 10 个类别. CIFAR-100 则包含 100 个类别, 图像尺寸同为 32×32 . Fashion-MNIST 是一个用于服饰分类的数据集, 其图像为单通道, 包含 T 恤、裤子、套衫等 10 个类别. Imagenette 和 Imagewoof 是 ImageNet 的 10 个易于分类的小型数据集, 图像尺寸为 224×224 , 其主要指标见表 1.

表 1 数据集的主要指标

| 数据集 | 类别数 | 图像 (幅) | |
|---------------|-----|--------|-------|
| | | 训练数 | 测试数 |
| CIFAR-10 | 10 | 50000 | 10000 |
| CIFAR-100 | 100 | 50000 | 10000 |
| Imagenette | 10 | 9469 | 3925 |
| Imagewoof | 10 | 9035 | 3939 |
| Fashion-MNIST | 10 | 60000 | 10000 |

4.2 实验环境及参数设置

本文方法与选取的对比方法在同一实验平台下进行实验, 全部实验在 Ubuntu 18.04 的操作系统, 内存为

60 GB 的 NVIDIA Tesla P100 显卡上进行, 深度学习框架为 PyTorch, 编程软件为 PyCharm.

为验证本文方法的有效性, 本文方法分别与基线方法 (Baseline)、传统数据增强方法以及 Mixup^[6]、CutMix^[7]、Mosaic^[9]、SaliencyMix^[8]以及 SaliencyCutMix^[4]这 5 种数据增强方法进行比较. 在相同的数据集和网络模型上, 确保比较方法之间的迭代次数 (Epoch)、初始学习率 (LR) 以及批量大小 (batch-size) 等参数相同, 均使用动量为 0.9 的随机梯度下降 (stochastic gradient descent, SGD) 模型, 权重衰减为 $5E-4$. CIFAR-10 和 CIFAR-100 数据集采用 128 张图像批次训练 200 Epoch, 学习率为 0.1. Fashion-MNIST 数据集采用 256 张图像批次训练 200 Epoch, 学习率为 0.1. Imagenette 和 Imagewoof 数据集采用 16 张图像批次训练 200 Epoch, 学习率为 0.1.

4.3 实验结果分析

为比较本文提出的方法与其他数据增强方法的性能, 实验使用 ResNet-18、ResNet-34 和 WideResNet16-8 这 3 种网络模型. 对比方法 Mixup 执行数据增强的概率为 0.5, CutMix^[7]、SaliencyMix^[8]和 SaliencyCutMix^[4]方法执行数据增强的概率为 1.0. 同时, Mixup 和

CutMix 的 α 和 β 值相同且都分别定义为 1.0 和 3.0.

在 ResNet-18 网络模型及 ResNet-34 网络模型中, 本文方法及对比方法的实验结果见表 2 及表 3. 其中, Base-Tra 代表传统的数据增强方法, Ours 代表本文方法. 实验结果表明, 本文方法的分类准确率总是高于其他方法. 值得注意的是, Fashion-MNIST 数据集只有一个通道, 所以本文方法与其他方法的准确率差别不大, 但与传统的数据增强方法相比 Top-1 准确率提升大约 0.6%. 在 CIFAR-10 数据集上, 本文方法与基线方法进行对比, 其 Top-1 准确率提升 7% 左右; 与其他数据增强方法相比, 该方法准确率提升了近 1%. 在 CIFAR-100 数据集上, 该方法的 Top-1 准确率比其他方法提升了 1% 左右, 且由于该数据集中类别较多, 样本数量与实际任务相近, 更能体现该方法的优越性. 在 Imagenette 数据集和 Imagewoof 数据集上, 该方法的分类准确率与其他方法相比也有所提升. 尤其在 Imagenette 数据集上, 使用 ResNet-18 网络模型时该方法比其他方法的分类准确率提升了 1% 以上.

表 2 ResNet-18 网络模型的实验结果对比 (Top-1) (%)

| 方法 | CIFAR-10 | CIFAR-100 | Fashion-MNIST | Imagenette | Imagewoof |
|----------------|----------|-----------|---------------|------------|-----------|
| Baseline | 88.3 | 62.5 | 89.6 | 87.3 | 74.6 |
| Base-Tra | 94.7 | 77.3 | 94.9 | 89.7 | 79.1 |
| Mosaic | 94.5 | 76.3 | 94.4 | 87.1 | 76.7 |
| Mixup | 94.0 | 74.8 | 95.5 | 85.8 | 75.9 |
| CutMix | 94.4 | 76.5 | 95.4 | 91.3 | 81.5 |
| SaliencyMix | 94.6 | 75.5 | 95.2 | 90.5 | 81.3 |
| SaliencyCutMix | 94.9 | 76.3 | 95.5 | 91.2 | 81.4 |
| Ours | 95.3 | 78.6 | 95.6 | 92.1 | 81.7 |

表 3 ResNet-34 网络模型的实验结果对比 (Top-1) (%)

| 方法 | CIFAR-10 | CIFAR-100 | Fashion-MNIST | Imagenette | Imagewoof |
|----------------|----------|-----------|---------------|------------|-----------|
| Baseline | 89.0 | 61.5 | 94.1 | 82.4 | 66.8 |
| Base-Tra | 95.3 | 77.9 | 95.2 | 89.9 | 81.0 |
| Mosaic | 94.7 | 77.3 | 95.4 | 89.6 | 77.1 |
| Mixup | 93.8 | 73.1 | 95.5 | 86.8 | 75.9 |
| CutMix | 94.2 | 73.8 | 95.6 | 91.7 | 82.7 |
| SaliencyMix | 94.3 | 74.2 | 95.4 | 91.4 | 82.4 |
| SaliencyCutMix | 94.8 | 74.7 | 95.7 | 91.9 | 82.5 |
| Ours | 95.7 | 78.8 | 95.8 | 92.3 | 82.9 |

ResNet-18 网络模型与 ResNet-34 网络模型对不同数据集分类准确率呈现显著差异. 具体而言, 两种网络模型对 CIFAR-10 与 CIFAR-100 数据集的分类准确率提升幅度较小, 但对 Imagenette 与 Imagewoof 数据集的分类准确率提升幅度明显. 在 CIFAR-10 和 CIFAR-100

数据集中, 图像分辨率只有 32×32 像素, 而 Imagenette 和 Imagewoof 数据集中, 图像分辨率为 224×224 . 低分辨率图像导致图像细节信息的丢失或模糊, 对网络模型造成一定影响. 因此图像分辨率会对网络模型具有一定影响, 本文方法使用较高分辨率的图像可以更有效地提高模型性能.

表 4 是本文方法与对比方法在 WideResNet16-8 网络模型的实验结果. 本文方法在 CIFAR-10、CIFAR-100 数据集上的分类准确率与其他方法相比均有提升. 在 Fashion-MNIST 数据集上, CutMix 方法的分类准确率最高, 达到 96.2%. 但是, 本文方法比 CutMix 方法相比 Top-1 准确率仅下降 0.4% 左右. 在 Imagenette 和 Imagewoof 数据集上, 本文方法的分类准确率是 89.7% 和 78.7%.

表 4 WideResNet16-8 网络模型的实验结果对比 (Top-1) (%)

| 方法 | CIFAR-10 | CIFAR-100 | Fashion-MNIST | Imagenette | Imagewoof |
|----------------|----------|-----------|---------------|------------|-----------|
| Baseline | 93.1 | 73.2 | 94.4 | 80.1 | 62.2 |
| Mosaic | 94.6 | 76.9 | 95.5 | 87.0 | 71.4 |
| Mixup | 95.7 | 78.5 | 95.6 | 87.8 | 77.7 |
| CutMix | 95.2 | 76.6 | 96.2 | 88.3 | 78.0 |
| SaliencyMix | 95.8 | 76.3 | 95.4 | 87.1 | 77.9 |
| SaliencyCutMix | 95.9 | 77.7 | 95.9 | 88.6 | 78.4 |
| Ours | 96.3 | 79.3 | 95.8 | 89.7 | 78.7 |

实验结果表明, 本文方法在不同数据集的性能表现上存在明显差异. 然而, 相较于 Imagenette 数据集, 本文方法对 Imagewoof 数据集的提升效果较为有限. 在 ResNet-18 和 ResNet-34 网络中, 本文方法在 Imagenette 和 Imagewoof 数据集上与 CutMix 相比提升分别约 0.7% 和 0.2%. 在 WideResNet16-8 网络中, 本文方法相较于 CutMix 方法在 Imagenette 数据集上提升 1.4%, 而在 Imagewoof 数据集上提升 0.7%. 因为 Imagewoof 数据集图像均为犬的不同种类, 所以图像区别较小. 本文方法对多幅不同类别道德图像进行拼接, 造成网络模型对 Imagewoof 数据集进行处理时较容易对不同类别的图像进行混淆, 从而准确率提升有限. 相比较而言, Imagenette 数据集含有多种物体的类别, 本文方法使网络模型能更容易区分不同物体的图像, 从而分类准确率得到明显提升.

表 5 展示了本文方法与对比方法在 ResNet-18 和 ResNet-34 网络模型 Top-5 错误率的实验结果. 本文方法的 Top-5 错误率与其他方法相比均有下降, 说明本文方法具有提高网络模型的泛化能力.

表5 Imagenette 数据集的 Top-5 错误率对比 (%)

| 方法 | ResNet-18 | ResNet-34 |
|-------------|-----------|-----------|
| Baseline | 7.15 | 6.96 |
| Mosaic | 6.48 | 6.28 |
| Mixup | 6.33 | 6.40 |
| CutMix | 6.21 | 5.92 |
| SaliencyMix | 6.31 | 5.71 |
| Ours | 5.98 | 5.41 |

为进一步验证本文方法的有效性,使用 ResNet-50 网络模型在 CIFAR-10 数据集上进行实验,并生成分类混淆矩阵图,如图 5 所示.图中展示各类样本的预测正确和预测错误的数量以及各种方法的分类准确率.虽然本文方法的 Top-1 准确率与其他方法相比提升幅度有限,但是本文方法在 CIFAR-10 数据集上能够正确分类的样本数量明显多于其他方法,尤其是 Mixup 与 CutMix 方法. Mixup 方法的图像插值混合导致图像模糊,从而影响了正确分类样本的数量. CutMix 方法生成的新样本只包含一幅图像的特征区域,限制了模型学习能力,从而影响了预测样本正确的数量.本文方法与 SaliencyMix 方法进行对比,正确预测样本数量近乎相同且准确率提升的幅度较小,但 SaliencyMix 方法需要额外的计算开销.综上,本文方法通过随机拼接 4 幅图像,使模型能够获得更多的特征信息,并且采用反向学习机制进行学习,使网络模型能够有效地区分图像类别的同时还加强了对某一类别的判断能力.

4.4 消融实验

Mixup、CutMix 方法采用两幅图像进行融合, Mosaic 方法与本文方法是 4 幅图像进行拼接的数据增强方法.为验证本文方法的有效性,根据各种方法的区别,使用 ResNet-34 网络模型进行消融实验,实验结果见表 6.

在 CIFAR-10 和 Imagenette 数据集上,本文方法与 Mixup 方法结合,相较于与 CutMix 方法结合分类准确率分别下降 3.4% 和 5.3%.由于 Mixup 方法通过将两幅图像按照一定比例进行融合来产生新样本,造成了图像信息的大量丢失,加大了网络模型对图像类别进行区分的难度,从而导致与 Mixup 方法相结合后的分类准确率下降.而 Mosaic 方法与 Mixup 方法结合分类准确率下降幅度更大. Mosaic 方法对图像进行随机切分后再拼接,丢失原图像中的关键特征区域,网络模型无法充分学习到图像的特征信息,还容易误导网络模型对图像类别的学习.

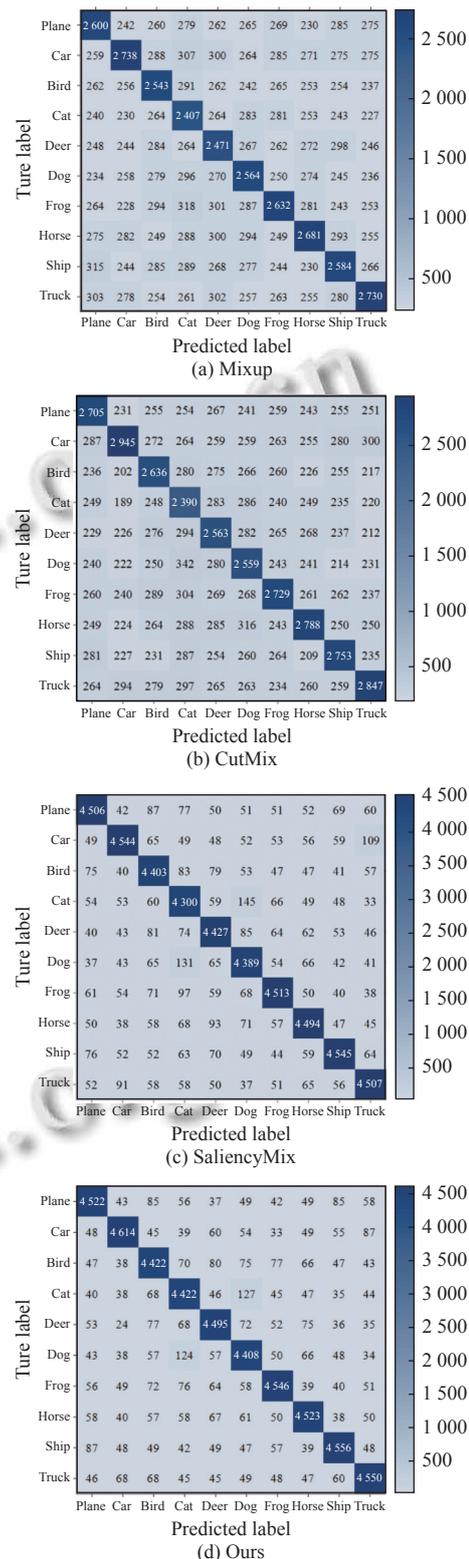


图5 方法的混淆矩阵

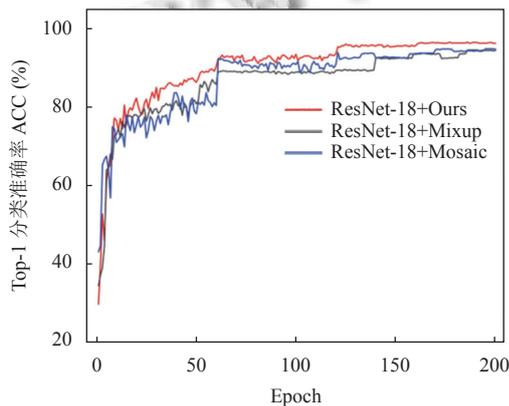
4.5 模型收敛性能分析

本文方法与对比算法 Mixup、Mosaic 在不同网络模型上的收敛性能对比如图 6 所示.在 ResNet-18 网络

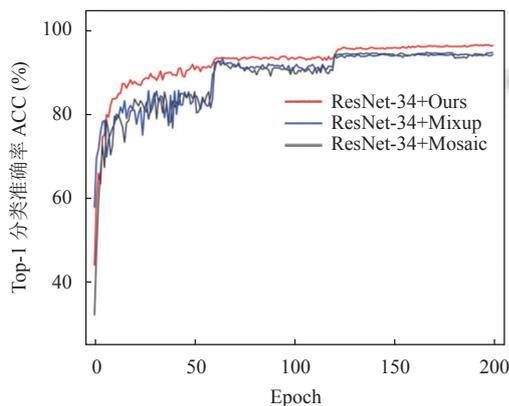
模型上,通过 200 次迭代过程观察准确率 ACC 的上升情况,并把本文方法和其他两种数据增强方法作了对比,见图 6(a)。同时还将本文方法和其他两种方法在 ResNet-34 网络模型中的分类准确率进行对比,见图 6(b)。根据图 6 所示,相较于 Mixup 和 Mosaic 方法,本文方法在准确率方面呈现出更为平稳的上升趋势。同时,本文方法在保证准确率稳定提升的前提下,还能有效减少计算时间。在训练过程中,本文方法呈现出更为优异的收敛性。

表 6 不同方法消融实验结果

| Mixup | CutMix | Mosaic | Ours | ACC (%) | |
|-------|--------|--------|------|----------|------------|
| | | | | CIFAR-10 | Imagenette |
| √ | × | √ | × | 91.6 | 85.0 |
| √ | × | × | √ | 93.0 | 86.1 |
| × | √ | √ | × | 95.6 | 90.5 |
| × | √ | × | √ | 96.4 | 91.4 |
| √ | √ | √ | × | 90.9 | 78.4 |
| √ | √ | × | √ | 92.1 | 79.9 |



(a) ResNet-18



(b) ResNet-34

图 6 方法的 Top-1 准确率对比

图 7 为本文方法和 Mixup、Mosaic 方法在 ResNet-34 网络模型上 Top-5 错误率的比较。与 Mixup 和 Mosaic 方法相比,本文方法的 Top-5 错误率更低,下降趋势更

加平稳,同时收敛速度在大部分迭代过程中也明显优于其他两种方法。结果表明,本文提出的方法在提高模型性能和泛化能力方面具有较大的优势。

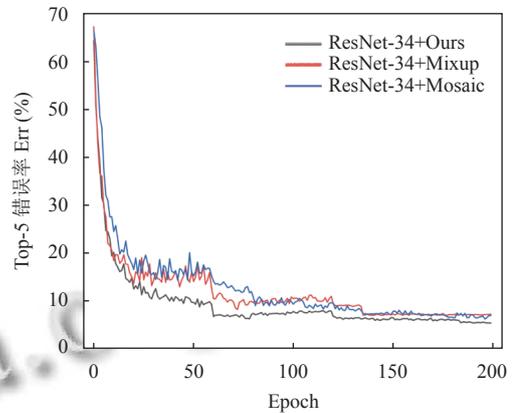


图 7 方法的 Top-5 错误率对比

图 8 为本文方法和 CutMix、SaliencyMix 方法在 ResNet-34 网络模型上损失函数的比较。图 8 (a) 为迭代 200 次时,本文方法和 CutMix 方法的损失率下降情况。本文方法在训练前期迅速收敛,达到较小的损失率。相比较而言,CutMix 方法损失率降低曲线更波动且总体变化趋势更缓慢,这表明该方法在训练时可能受部分噪声或者异常样本等因素影响而使网络模型不能对训练数据进行稳定拟合。本文方法在训练过程中能够更好地优化网络模型,使其更加准确地拟合训练数据。图 8 (b) 则展示了该方法和 SaliencyMix 方法迭代时损失率之间的对比。总体来看,本文方法损失率曲线低于 SaliencyMix 方法。由于本文方法中采用 4 幅图像进行拼接的混合策略,使得模型更加准确地学习图像的重要信息。这说明本文方法可以减少训练时的损失,损失率减小就意味着该方法对网络模型的优化程度较高,同时体现出该方法使网络模型对训练数据有较好的适应能力。

4.6 随机图像对性能影响分析

为验证 4 幅图像随机选取的作用,在 CIFAR-100 数据集上使用 ResNet-18 和 WideResNet16-8 网络模型进行对比实验,实验结果见表 7、表 8。

实验结果表明,采用随机选取 4 幅图像进行拼接的方法具有最佳性能,而随机选取一幅图像进行拼接方法的分类准确率最低。图像可能包含不同的物体、场景或角度,拼接 4 幅图像可以产生更丰富、更复杂的目标特征,有助于丰富训练样本的特征信息,使模型

能够更好地区分不同类别之间的特征差异. 同时, 还可以减小样本之间的相关性, 降低模型对于特定样本的过拟合风险. 而在选取一幅图像进行拼接的情况下, 模型可能过度依赖于该幅图像的特征, 无法充分利用其他图像包含的信息. 这说明通过随机选取 4 幅图像进行拼接可以显著增加样本的多样性, 使模型能够学习到更多的特征信息, 从而提升模型的性能. 因此, 本文方法是一种有效的方法, 可以显著提升模型的学习效果和泛化能力.

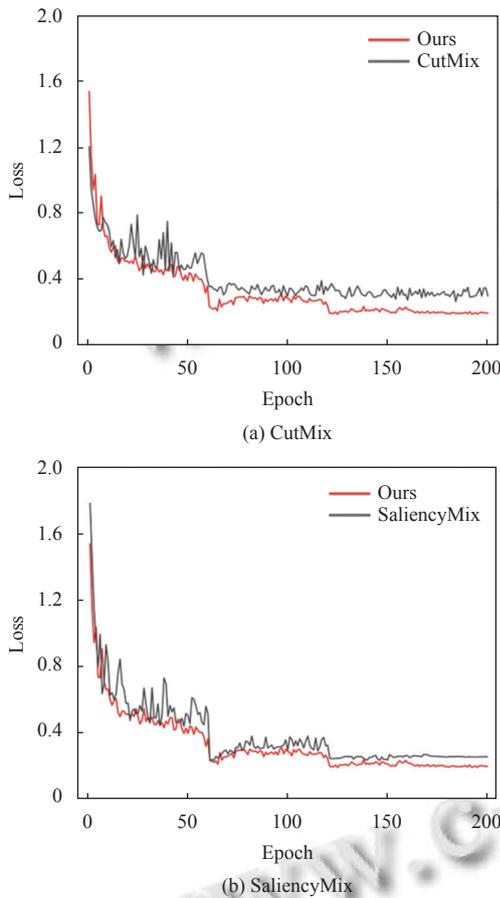


图 8 方法的损失率对比

表 7 ResNet-18 网络模型随机图像的实验 (%)

| 图像 | CIFAR-100 | Fashion-MNIST | Imagenette |
|------|-----------|---------------|------------|
| Cat1 | 77.5 | 95.4 | 86.6 |
| Cat2 | 77.8 | 95.5 | 90.9 |
| Cat3 | 78.2 | 95.5 | 91.6 |
| Cat4 | 78.6 | 95.8 | 92.3 |

表 8 WideResNet16-8 网络模型随机图像的实验 (%)

| 图像 | CIFAR-100 | Imagenette |
|------|-----------|------------|
| Cat1 | 78.1 | 86.5 |
| Cat2 | 78.3 | 87.8 |
| Cat3 | 78.8 | 88.2 |
| Cat4 | 79.3 | 89.7 |

5 结论与展望

混合样本数据增强方法可以显著提高模型的性能, 然而这些方法仅在网络模型中对图像类别进行正向表达. 为更好地促进网络模型对图像类别的学习, 提出反向目标干扰的图像数据增强方法. 该方法首先利用训练数据中的其他图像丰富图像背景及目标, 然后通过反向学习机制加强网络模型对图像类内和类间的辨别能力. 实验结果表明, 本文方法提高了网络模型的学习效果和泛化能力, 降低了模型过拟合的风险, 与现有数据增强的方法相比具有了良好的性能. 然而, 在进行中心点裁剪时, 图像的小部分特征区域会丢失, 后续将对中心点裁剪方法进行优化研究.

参考文献

- Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *Journal of Big Data*, 2019, 6(1): 60. [doi: 10.1186/s40537-019-0197-0]
- Su D, Kong H, Qiao YL, et al. Data augmentation for deep learning based semantic segmentation and crop-weed classification in agricultural robotics. *Computers and Electronics in Agriculture*, 2021, 190: 106418. [doi: 10.1016/j.compag.2021.106418]
- 郭文明, 王腾亿. 类激活映射指导数据增强的细粒度图像分类. *计算机辅助设计与图形学学报*, 2021, 33(11): 1698-1704.
- 曾武, 朱恒亮, 邢树礼, 等. 显著性检测引导的图像数据增强方法. *图学学报*, 2023, 44(2): 260-270.
- Harris E, Marcu A, Painter M, et al. Fmix: Enhancing mixed sample data augmentation. arXiv:2002.12047, 2021.
- Zhang HY, Cissé M, Dauphin YN, et al. Mixup: Beyond empirical risk minimization. *Proceedings of the 6th International Conference on Learning Representations*. Vancouver: OpenReview.net, 2018.
- Yun S, Han D, Chun S, et al. CutMix: Regularization strategy to train strong classifiers with localizable features. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019. 6022-6031.
- Uddin AFMS, Monira MS, Shin W, et al. SaliencyMix: A saliency guided data augmentation strategy for better regularization. arXiv:2006.01791, 2021.
- Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934, 2020.
- 马宁, 曹云峰, 王指辉, 等. 基于 YOLOv5 网络架构的着陆

- 跑道检测算法研究. 激光与光电子学进展, 2022, 59(14): 189–195.
- 11 Koushik J. Understanding convolutional neural networks. arXiv:1605.09081, 2016.
 - 12 Kamath U, Liu J, Whitaker J. Deep Learning for NLP and Speech Recognition. Cham: Springer, 2019.
 - 13 Cui JF, Zhang XY, Xiong FB, *et al.* Pathological myopia image recognition strategy based on data augmentation and model fusion. Journal of Healthcare Engineering, 2021, 2021: 5549779.
 - 14 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778.
 - 15 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017, 60(6): 84–90. [doi: 10.1145/3065386]
 - 16 马国祥, 严传波, 张志豪, 等. 基于数据增强的 CT 图像病灶检测方法. 计算机系统应用, 2021, 30(10): 187–194. [doi: 10.15888/j.cnki.csa.008135]
 - 17 DeVries T, Taylor GW. Improved regularization of convolutional neural networks with cutout. arXiv:1708.04552, 2017.
 - 18 Cubuk ED, Zoph B, Mane D, *et al.* AutoAugment: Learning augmentation policies from data. arXiv:1805.09501, 2019.
 - 19 Kim JH, Choo W, Song HO. PuzzleMix: Exploiting saliency and local statistics for optimal mixup. Proceedings of the 37th International Conference on Machine Learning. JMLR.org, 2020. 489.
 - 20 Hong MN, Choi J, Kim G. StyleMix: Separating content and style for enhanced data augmentation. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 14857–14865.
 - 21 Suzuki T. TeachAugment: Data augmentation optimization using teacher knowledge. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 10894–10904.
 - 22 Rahnamayan S, Tizhoosh HR, Salama MMA. Opposition-based differential evolution. In: Chakraborty UK, ed. Advances in Differential Evolution. Berlin: Springer, 2008. 155–177.
 - 23 Ventresca M, Tizhoosh HR. Two frameworks for improving gradient-based learning algorithms. In: Tizhoosh HR, Ventresca M, eds. Oppositional Concepts in Computational Intelligence. Berlin: Springer, 2008. 255–284.
 - 24 Tizhoosh HR, Sahba F. Quasi-global oppositional fuzzy thresholding. Proceedings of the 2009 IEEE International Conference on Fuzzy Systems. Jeju: IEEE, 2009. 1346–1351.
 - 25 Tizhoosh HR, Rahnamayan S. Learning opposites with evolving rules. Proceedings of the 2015 IEEE International Conference on Fuzzy Systems. Istanbul: IEEE, 2015. 1–8.
 - 26 Ernoult MM, Normandin F, Moudgil A, *et al.* Towards scaling difference target propagation by learning backprop targets. Proceedings of the 39th International Conference on Machine Learning. Baltimore: PMLR, 2022. 5968–5987.
 - 27 Chen F, Liu YM, Yang J, *et al.* Multi-objective particle swarm optimization with reverse multi-leaders. Mathematical Biosciences and Engineering, 2023, 20(7): 11732–11762. [doi: 10.3934/mbe.2023522]
 - 28 Tizhoosh HR. Opposition-based learning: A new scheme for machine intelligence. Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce. IEEE, 2005. 695–701.

(校对责编: 孙君艳)