软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software, [doi: 10.13328/j.cnki.jos.000000]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

面向深度学习的图像数据增强综述*

杨锁荣[1,2], 杨洪朝[1,2], 申富饶[1,3], 赵健[4]

[1](计算机软件新技术国家重点实验室(南京大学),江苏 南京 210023)
[2](南京大学 计算机科学与技术系,江苏 南京 210023)
[3](南京大学 人工智能学院,江苏 南京 210023)
[4](南京大学 电子科学与工程学院,江苏 南京 210023)

通讯作者: 申富饶, E-mail: frshen@nju.edu.cn

摘 要: 深度学习已经在许多计算机视觉任务中取得了显著的成果.然而,深度神经网络通常需要大量的训练数据以避免过拟合,但实际应用中标记数据可能非常有限.因此,数据增强已成为提高训练数据充分性和多样性的有效方法,也是深度学习模型成功应用于图像数据的必要环节.本文系统地回顾了不同的图像数据增强方法,并提出了一个新的分类方法,为研究图像数据增强提供了新的视角.我们从不同的类别出发介绍了各类数据增强方法的优势和局限性,并阐述了各类方法的解决思路和应用价值.此外,我们还介绍了语义分割、图像分类和目标检测这三种典型计算机视觉任务中常用的公共数据集和性能评价指标,并在这三个任务上对数据增强方法进行了实验对比分析.最后,我们讨论了当前数据增强所面临的挑战和未来的发展趋势.
关键词: 深度学习;图像数据增强;图像识别;泛化性能;计算机视觉
中图法分类号: TP391

中文引用格式:
英文引用格式:

Image Data Augmentation for Deep Learning: A Survey

YANG Suo-Rong[1,2], YANG Hong-Chao[1,2], SHEN Fu-Rao[1,3], ZHAO Jian[4]

[1](State Key Laboratory for Novel Software Technology (Nanjing University), Nanjing University, Nanjing 210023, China)
[2](Department of Computer Science and Technology, Nanjing 210023, China)
[3](School of Artificial Intelligence, Nanjing 210023, China)
[4](School of Electronic Science and Engineering, Nanjing 210023, China)

Abstract: Deep learning has achieved significant success in many computer vision tasks. Due to the fact that deep neural networks typically require a large amount of training data to avoid overfitting, but annotated data may be limited in practical applications, data augmentation has become an effective way to improve the adequacy and diversity of training data, and is also a necessary step for the successful application of deep learning models to image data. This paper systematically reviews different methods of image data augmentation, and proposes a new classification method to provide a fresh perspective for studying image data augmentation. We introduce the advantages and limitations of various data augmentation methods from different categories, and elaborate on the solution strategies and application values of various methods. We also introduce commonly used public datasets and performance evaluation metrics in three typical computer vision tasks, semantic segmentation, image classification, and object detection, and conduct experimental comparative analysis of data augmentation methods on these three tasks. Finally, we discuss the challenges and future development trends faced by data augmentation.

* 基金项目: 国家自然科学基金(62276127)
收稿时间:修改时间:采用时间:

Key words: deep learning; image data augmentation; image recognition; generalization; computer vision

1 引言

深度学习在计算机视觉(CV)[1]、推荐系统(RS)[2]、和自然语言处理(NLP)[3]等领域取得了惊人的进展,这些研究领域的进展和广泛应用主要得益于三个方面:深度网络架构的发展、计算能力的进步以及大数据的获取.首先,网络模型的泛化能力往往与其规模成正比,与浅层网络相比,152 层的 ResNet[4]可以从增加的深度中获得更高的精度,因此研究者们往往倾向于使用更大、更深的模型架构.然而,深度模型层数的增加意味着参数量的增加, 训练大规模的深度模型往往需要更强的计算能力,因此计算能力的发展对深度学习有很大影响.只有拥有更强的计算能力,才有可能设计出具有更深层次架构的模型.最后,训练数据的获取和标注对深度学习的发展有着很重要的影响,随着待学习参数量的增加,模型往往需要更多的训练数据来缓解模型的过拟合问题[5][6].然而,我们发现深度模型的发展、计算能力的进步和数据集的构建三个方面的发展存在一定的不平衡性.深度模型的规模和图像处理单元（GPU）得到了迅猛发展[7][8].例如,GPU 的计算性能呈现逐步上升的趋势,具体来说,每秒钟的浮点操作数（FLOPS）几乎每 3-4 年翻一倍[9].同时,模型的规模也得到了显著提升,近年来,性能较好的模型都具有很大规模的参数量,例如在 ImageNet 数据集上实现了最高 top-1 精度的 Inception-v4 模型就包含了 4260 万个参数.而目前非常热门的基于 GPT 架构（Generative Pre-trained Transformer）[10][11]的 ChatGPT,其 GPT-3 模型[12]包含了 1750 亿个参数.尽管计算性能和模型规模有着迅猛发展,但是公共数据集的发展却没有那么显著[13][14][15].其中一个重要的原因是现实世界的数据很难采集,同时数据需要人工标注,一些专业数据集,例如医疗图像数据,甚至需要专业人士进行标注.因此,构建公开数据集需要耗费大量的人力成本和时间成本.除此之外,一些特殊的数据集,例如医疗记录和医疗图像,包含有隐私信息.出于对隐私的保护,这些重要的数据很难被公开使用,即使对数据进行了脱敏操作,这一系列的处理过程也会限制数据集的进一步发展和扩增.因此,将性能更好的模型应用到现实任务中时,通常面临着训练数据不足、训练数据类别不均衡等问题[16].为了应对上述问题,数据增强技术应运而生.

数据增强是机器学习中一项用于扩充数据集的技术,可通过对原始数据进行各种变换和扩充,从而提高模型的性能和鲁棒性.其本质是通过生成新的数据来增加训练数据的充分性和多样性,从而降低模型的过拟合风险.对于图像数据,通过裁剪、翻转、颜色域变换等操作,可以将训练集扩充至原来的三倍,如果对这些方法进行组合,如裁剪+翻转,就能将数据集增强至更多倍.对于文本数据,可以进行随机替换、插入和删除等操作,生成新的文本数据.增强的数据可以看作是从接近真实数据的分布中提取的,因此增强后的数据集可以代表更全面的特征,从而训练出更具泛化性能的模型.从另一个角度来看,随着深度模型的规模越来越大,当模型复杂度超过数据集复杂度时,模型会遭遇过拟合问题.因此,为了获得更好的性能表现,我们可以在不改变模型规模大小的前提下,通过数据增强技术来平衡模型和数据集之间的复杂度差异,提高训练数据集的复杂度,缓解模型的过拟合风险.数据增强的目的就是增加训练数据的数量和多样性,以改善模型的泛化能力,减少过拟合现象的发生,并提高模型的鲁棒性,使其在面对不同的输入数据时表现更好.

本文的组织安排如下.我们首先在第二章介绍数据增强研究相关代表性工作,第三章介绍我们应用于数据增强研究的新型分类方法,第四章我们将根据我们提出的分类方法对数据增强研究分类进行介绍,第五章我们介绍了数据增强方法的评估指标,并在图像分割、图像分类和目标检测三个典型的计算机视觉任务上用多个常见的基准数据集进行了实验分析.最后,我们对数据增强研究的面临的挑战和未来的发展方向进行了讨论.

2 相关研究概述

当前数据增强研究成果众多,本章首先对数据增强具有代表性的成果以及发展脉络进行简要介绍,随后对数据增强研究的相关综述进行整理,总结了其中观点,并提出本文的不同之处和创新点.

最初的数据增强方法是通过施加低计算代价的图像操作来生成数据,如改变图像的对比度、亮度以及对图

像进行旋转等变换操作.然而,这些方法产生的数据量有限,且难以生成具有多样性的数据.随着研究的深入,研究者们提出了更多创新性的方法来扩展数据增强的范围和效果.除了 Cutout[17]和 Pairing Sample[18]这样的确定性生成方式,研究者们还利用强化学习等方法来确定数据增强操作的最优组合,从而提高数据增强的效果和生成数据的多样性[19],例如 AutoAugment[20], Fast-AutoAugment[22]等等.同时,考虑到深度模型真正用于下游任务的是从图像数据提取到的深度特征,研究者们又在特征空间中进行一系列的数据增强操作,如对特征进行旋转、缩放、裁剪等.除此之外,还有一些新型的数据增强方法被提出,如 Mixup[22]、Cutmix[23]、GridMask[24]等.这些方法都是通过在图像上添加噪声、混合、遮挡等方式来生成新的图像数据,从而增加数据的多样性和难度.同时,一些基于生成式模型的数据增强方法也被提出,如 GAN[25]、VAE[26]等,这些方法可以通过生成新的图像数据来增加训练数据的数量和质量.不过,这些方法需要大量的训练样本和计算代价,因此在实际应用中受到限制.

有不少文献回顾和总结了数据增强方面的研究工作,文献[19]探索并比较了图像分类中数据增强问题的多种解决方案,但它只涉及图像分类任务,并且只对传统的变换和 GANs 进行了实验.文献[27]从变换类型和深度学习的角度回顾了现有的人脸数据增强工作.然而,该调查只是针对人脸识别任务.文献]主要集中在基于数据扭曲和过采样的不同数据增强技术.然而,它并没有对不同的方法进行系统的回顾.虽然文献[29]介绍了一些现有的方法和有前景的数据增强发展,但它没有提供对各种实际任务的数据增强有效性的评估,并且缺乏一些新提出的方法,如 CutMix、GridMask、AugMix[31]、YOCO[32]等等.

总的来说,当前对于图像数据增强方法进行研究的综述文献通常存在一下问题: 1) 没有对不同的数据增强方法进行系统的分类和总结,没有对这些方法的优缺点进行深入分析和讨论;2)只是对不同方法进行了梳理,没有对方法的有效性按类别进行验证,缺乏实验验证评估;3)对方法的分类较为简单,图像增强数据不能完全归纳涵盖,不同类别存在重叠;4)针对不同类别的方法,没有探究其背后的可解释性特征;5)方法的总结梳理不够全面,对于一些新提出的方法没有归纳总结,可能会导致读者对该领域的最新发展了解不足.

本文旨在填补现有图像数据增强方法综述中存在的空白,通过总结最新的新型图像数据增强方法来帮助研究人员更好地应用这些方法.为此,本文提出了一个图像数据增强方法的分类法,以系统地回顾数据增强技术在目标检测、语义分割和图像分类等计算机视觉任务中的应用.在这个分类法的基础上,本文详细介绍了不同类别的数据增强方法的设计动机和可解释性原理.此外,本文还通过实验比较了不同种类的数据增强方法及其组合在各种深度学习模型以及开放的图像数据集上的表现,以便研究人员能够更好地选择和应用这些方法.最后,本文还探讨了未来图像数据增强研究的发展方向和趋势,以进一步促进数据增强技术在计算机视觉领域的应用和发展.

表 1 数据增强新型分类方法概述

基本方法	在图像层面对图像数据进行直接操作
● 图像变换	利用传统的图像操作对图像进行变换.
● 图像擦除	对图像中一部分区域进行像素擦除.
● 图像混合	将多个图像混合到一张图像中
基于深度学习方法	**利用深度学习方法或模型生成新数据**
● 自动增强	自动选择数据增强操作生成新数据
● 特征增强	利用深度模型提取特征并在特征空间中进行增强
● 深度生成式方法	利用深度生成式模型生成新数据

3 新型图像数据增强技术分类法

近年来,数据增强研究已成为深度学习领域一个极为重要的组成部分.研究者们提出了大量的数据增强方法,对这些方法进行分类已经成为一项必不可少的工作.本文提出了一种新的分类方法,图 1 展示了分类方法的框架.在新的分类方法中,首先根据引入深度学习技术的程度将数据增强方法分为基本方法和基于深度学习的方法两个大类.基本方法是指直接在图像层面,通过一些基本的图像操作来生成新的数据.这类方法计算代价小,可以直接内嵌于模型训练过程中,用确定性的方式为模型训练提供更加多样化的训练数据.基于深度学习的方法则是使用深度学习技术和模型来生成新数据,往往需要提前训练好生成式模型以保证数据的质量.在新的分类方法下,这两个大类又分别有三个子类,根据对数据操作的形式进行分类,基本数据增强方法类包括基础图像操作类、图像擦除类和图像混合类,而基于深度学习的方法则包括自动增强类、特征增强类和深度生成式模型.表 1 对这几个子类进行了简要介绍,而表 2 则对各个类别进行了进一步分析,同时还列出了具有代表性的方法示例和相关论文,从而实现了清晰、快速的分类效果.

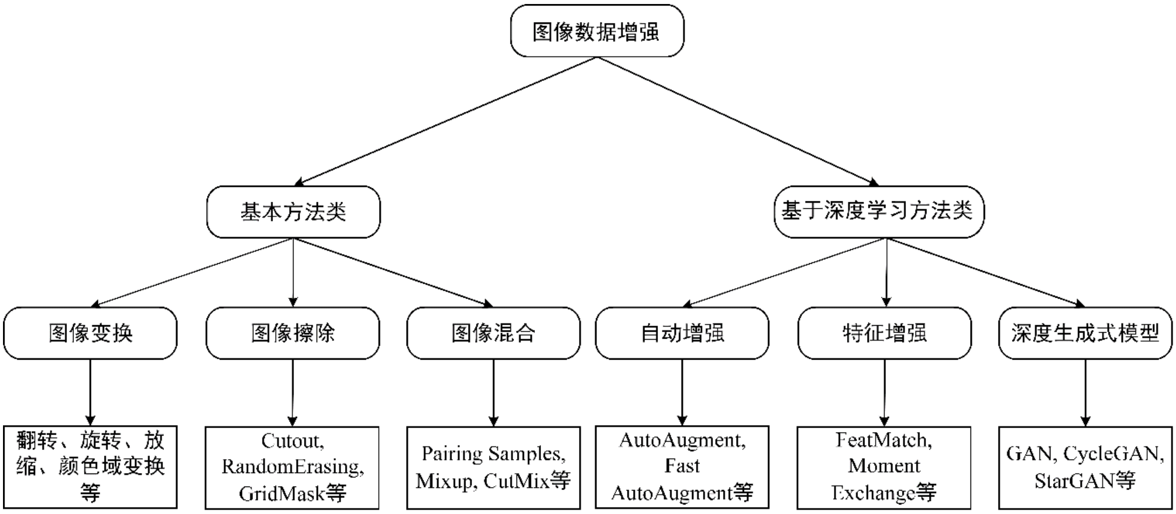


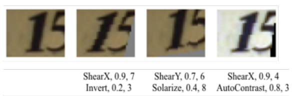
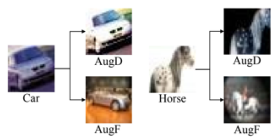
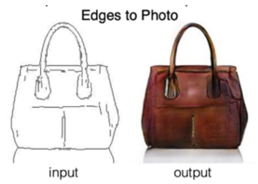
图 1 图像数据增强方法新型分类

基本方法类数据增强方法不需要引入深度学习模型,而是直接在图像空间对图像数据进行操作.这类方法具有计算复杂度小、易于操作等特点,因此目前大部分的数据增强方法都属于这一类.这类方法主要是通过图像操作来模拟现实场景中可能出现的模式.我们根据数据变换的形式将其进一步细分为图像变换、图像擦除和图像混合三个子类.其中,图像变换类可以模拟现实场景中的图像视角和颜色域的变化.例如,从表 2 中给出的示例可以看到,增强数据展现了目标以不同角度和尺寸出现的场景.近年来,图像擦除类方法取得了显著进展.这类方法可以模拟目标被遮盖的情况,进一步降低模型的敏感性并增强其泛化性能.甚至模型可以在目标被部分遮挡的情况下仍然正确识别.为了防止训练数据中的感兴趣目标被完全遮挡,表 2 中的示例给出了一种结构性遮挡的方法.图像混合类可以模拟一张图像中出现多个目标的情况,主要有两种形式,一种是像素级别的混合,直接将两张图片按像素加权叠加;另一种是将两张或多张图片拼接成一张图片.表 2 中的示例就是这种形式,通过以上方法来生成新的训练数据,可以让模型“见到”更丰富多样化的数据模式,最终提高模型的泛化性能.当前的研究中,数据增强方法面临着许多挑战和机遇,例如如何更好地保留原始数据的关键信息以及如何对不同的任务选择合适的数据增强方法等问题.因此,未来研究应该注重这些问题,并提出有效的解决方案.

基于深度学习的方法类在进行数据增强操作时引入深度学习技术,其中自动增强类是利用强化学习技术离线搜索得到最优的数据增强方法及其组合,在训练过程中使用这些最优的方法,由于组合形式的多样性,可以得到更加多样化的增强数据.以上这些方法都是基于已有的图像进行一定程度的变换得到增强数据,除此之外,

数据增强还可以借助深度生成式模型来生成“全新”的数据，以人脸数据集为例,如果数据集中只有一个人的人脸数据,传统的基于图像变换的方法仍然只能生成这一个人的数据,即使在角度或大小上有所变化,但是基于深度生成式模型的数据增强方法可以生成全新的人脸数据,或者是新的场景下的数据.除了直接在图像空间中进行增强,特征增强类方法通过在特征空间中根据特征数据生成增强数据,在训练过程中直接给模型输入特征数据,不需要特征提取过程.更多的细节和相关方法我们将在后文中进行介绍.

表 2 分类说明

分类	子分类	方法特点	论文	方法示例
基本方法	图像变换	通过旋转、颜色域变换等简单易操作的图像操作生成更加多样化的增强数据.	[29]	 [29]
	图像擦除	在图像中随机擦除单个或多个子区域的像素,以此来增强模型对于遮盖场景的鲁棒性.	[17] [34] [35] [36] [37]	 [24]
	图像混合	将两个或者多个图像按像素或拼接成单个图像,这可以强迫模型学习单个样本中出现多个目标的场景.	[18] [30] [31] [32] [38] [39]	 [18]
基于深度学习方法	自动增强	在不确定最优增强策略的问题背景下,利用强化学习搜索得到最优的数据增强策略的参数和组合形式,在训练过程中,可以使用不同的组合形式提供非常多样化的生成数据.	[40] [41] [42] [43] [44] [45] [46] [47] [48]	 [20]
	特征增强	由于模型是在提取了数据的特征之后进行下游操作的,直接在特征空间中进行数据增强操作,可以一定程度上降低在图像空间中可能引入的噪声.	[49] [51] [52] [53]	 [52]
	深度生成式方法	深度生成式方法首先利用数据训练得到一个较好的生成式模型,根据任务的不同,生成数据可能也需要生成特定的标签,自动化的形成训练数据集而不需要人工干预.	[19] [54] [55] [56] [57] [58]	 [19] [55]

4 数据增强方法概述

4.1 基本数据增强方法

基本数据增强方法的核心思想是,用尽可能低计算代价的图像空间操作生成新数据.由于在数据生成阶段不引入深度学习技术,因此这类方法计算开销比较低.因此在模型训练时,相比于梯度计算和回传等操作,数据增强过程的计算开销可以忽略不计.

4.1.1 图像处理

基础的图像处理主要集中在图像变换上,如旋转、翻转、裁剪等.这类技术的动机在于现实场景中的目标通常会以不同大小、不同角度、甚至不同场景来呈现.通过运用基础图像处理技术,可以给模型提供更加多样化的训练数据,从而可以模拟更多场景的数据,最终提高模型的泛化性能.例如,通过变换图像的亮度信息,可以模拟现实场景中光照条件发生改变的情况.这一类技术大多是在图像空间中直接对图像进行操作,易于实现.表 3 中给出了这些方法的简要描述.

表 3 基础图像操作及简要描述

方法	描述
翻转	将图像水平翻转、垂直翻转或同时翻转
旋转	以某一个角度旋转图像
缩放	增大或减小图像尺寸
噪声注入	在图像中加入噪声
颜色空间	改变图像颜色通道
对比度	改变图像对比度
锐化	修改图像清晰度
平移	将图像水平移动、垂直移动或同时移动
裁剪	裁剪图像的一个子区域

将基础图像操作用于数据增强的好处在于易于实现和计算复杂度低,除此之外,基础图像操作也可以帮助增加训练数据的多样性和复杂性.通过在训练数据集中引入旋转、平移、缩放等基本图像处理操作,可以增加模型对于不同角度、大小、形状的目标物体的识别能力,从而提升模型的泛化性能.此外,基础图像操作也可以用于矫正图像中的扭曲或畸变,提高数据质量.但是需要注意的是,过多的基础图像操作可能会导致过拟合,因此需要根据具体的任务和数据集进行选择和调整.另外,对于存在填充效应的基础图像处理方法,需要采用一些有效的填充方法,例如使用周围像素的均值或高斯噪声进行填充.同时,为了防止图像中的感兴趣对象被移出图像边界,可以采用更为精细的处理方法,如基于对象的裁剪和旋转、使用边界填充方法等.在应用基础图像操作进行数据增强时,需要根据具体的任务和数据集情况,选择合适的操作方式和参数,以提高数据增强的效果.

4.1.2 图像擦除

基于图像擦除的图像增强方法通常在图像中删除一个或多个子区域,其主要思想是将这些子区域的像素值替换为常量值或随机值.这类方法的动机是通过遮盖图中的一部分区域来增强模型对遮挡场景的鲁棒性,使得增强模型在失去一部分信息后,仍然具备在剩余信息中找到用于识别目标的显著性特征的能力.从可解释性的角度来思考,模型在实际应用时往往会遇到数据受到遮盖的情况,因此图像擦除类方法可以在模型训练阶段提供这样的数据以供训练.与此同时, 由于这类方法的操作简单且计算复杂度较低,因此不会给模型训练带来额外的开销.

DeVries 和 Taylor 提出一种简单的卷积神经网络正则化技术, 称为 Cutout[17],它在训练卷积神经网络(CNN)时随机掩盖掉输入的方形区域,去除输入图像的连续部分,有效地增加了现有样本的部分遮挡版本.该技术可以解释为输入空间中 dropout 的扩展,但与 dropout 不同的是,Cutout 作用于卷积神经网络的输入层,而不是中间特征层,同时 Cutout 删除了输入的连续部分而不是单个像素.在这种方式下,删除的区域会传播到所有后续的特征图,产生图像的最终表示.Cutout 迫使模型更多地考虑完整的图像上下文,而不是依赖于少量特定视觉特征的存在.该方法不仅非常容易实现,而且它可以与现有的数据增强方法和其他正则化器结合使用,以进一步提

高模型性能.实验结果证明这种方法能够提高卷积神经网络的鲁棒性和整体性能.

Singh 等人提出了 Hide-and-Seek(HaS)[34],它的核心思想是随机隐藏训练图像中的块,迫使网络在最有判别性的内容被隐藏时寻找其他相关的视觉内容.图 2 展示了该方法的核心思想,如果我们从图像中随机移除一些最具辨别力的块,那么狗的脸将无法被模型看到.在这种情况下,为了更好地完成分类任务,模型必须寻求其他相关部分,如尾巴和腿部.通过在每次训练过程中随机隐藏不同的块,模型可以看到图像的不同部分,并被迫关注对象的多个相关部分,而不仅仅是最具判别力的部分.由于网络在训练过程中看到的是部分隐藏的物体,因此对遮挡具有鲁棒性,这是 Hide - and - Seek 区别于随机裁剪和翻转等标准数据增强技术的关键特性,其优势对于弱监督定位任务尤为显著.同时 HaS 不仅仅局限于图像定位任务,还可以推广到视频等其他形式的视觉输入,以及图像分类、时间动作定位、语义分割、情感识别、年龄/性别估计、行人重识别等其他识别任务.

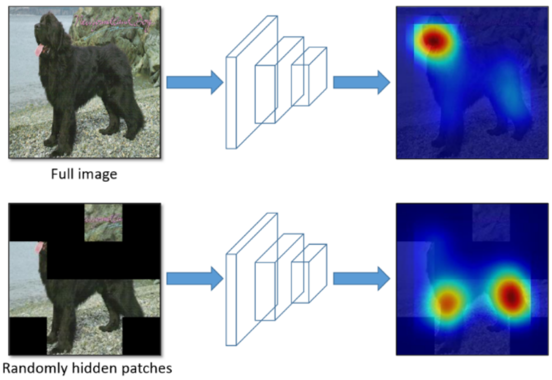


图 2 HaS 核心思想[34]

Zhong 等人提出了随机擦除法[35],在图像中随机选择一个矩形区域,用随机值或数据集的平均像素值替换其像素.训练不同遮挡程度的图像将有助于降低过拟合的风险,最终使模型更加鲁棒.随机翻转和随机裁剪也作用于图像层面,并与随机擦除密切相关,这两种技术都证明了具备提高图像识别精度的能力.与随机擦除相比,随机翻转在增强过程中不会造成信息损失.与随机裁剪不同的是,在随机擦除中,仅部分物体被遮挡,整体物体结构被保留,并且擦除区域的像素被随机值重新分配,这可以看作是给图像添加了块噪声.随机擦除方法简单可行,对不同任务中的多个图像数据集进行了合理改进,如目标检测、行人重识别等.该方法的缺点之一是目标区域可能被擦除,标签无法保留.

避免连续区域的过度删除和保留是信息删除方法的核心要求,一个成功的信息删除方法应该在删除和保留图像区域信息之间取得合理的平衡.直观上的原因有两方面.一方面,过多地删除一个或少数几个区域会导致完全的对象移除,上下文信息也会被移除,剩余信息不足以进行分类,图像更像是噪声数据.另一方面,过多的保留区域会使一些物体不受影响,它们可能成为导致网络鲁棒性降低的噪声图像.最近,Chen 等人分析了信息丢弃的要求,然后提出了一种结构化的方法 GridMask[24],它也是基于输入图像中的区域的删除.与 Cutout 和 HaS 不同,GridMask 既不删除连续区域,也不随机选择方块,删除的区域是一组空间上均匀分布的方块,其密度和大小可以控制.GridMask 很好地平衡了连续区域的删除和保留.

此外,为了平衡物体遮挡和信息保留,在模拟物体遮挡策略的基础上,提出了 FenceMask[36].FenceMask 通过增强遮挡块的稀疏性和规则性,克服了小目标增强的困难,显著提高了基线性能.FenceMask 的设计思路是进一步稀疏化和正则化遮挡块.遮挡块更稀疏意味着减少单个遮挡块的面积,增加单个遮挡块的密度.在包含许多小物体的视觉任务中,更稀疏的遮挡块可以保存更多的信息.遮挡块的规则化是指生成的遮挡块要符合特定的排列规则,防止遮挡块聚集.FenceMask 设计简单,易于实现,并且不会产生任何额外的计算.它可以应用于所有的卷积神经网络和各种计算机视觉任务,特别是在一些包含较多小物体或特征的数据集中表现出特别好的性能.除

此之外,Yang 等人提出了 AdvMask 方法[37],该方法不是随机遮盖图中的部分子区域来获得增强数据,而是利用对抗攻击方法首先学习并获得图像的关键像素,在数据增强阶段,通过遮盖一部分关键像素来迫使模型学习其它非关键像素并用其进行分类,这有效增强了模型的鲁棒性和泛化性能,使得模型可以在目标被遮挡或者部分遮挡的场景下,仍然可以进行正确分类.

图 3 展示了几种典型的图像擦除类数据增强方法的效果,这类方法的核心做法是用随机值或 0 像素值抹掉图中的一部分区域.这种方法的优点在于,通过屏蔽掉一部分像素值,模型可以学习到在其余区域中找到关键性特征的能力,同时可以增强模型对遮挡场景的鲁棒性.但是潜在缺点是,由于随机性的擦除包括位置和区域大小的随机性,有可能导致图中的感兴趣目标被完整擦除掉,因此需要根据数据集的特点来设定方法的参数以避免此情况的发生.从图中也可以看出,相比于结构化删除方法 GridMask 和 AdvMask,其它方法更容易遮盖掉图中的小目标.

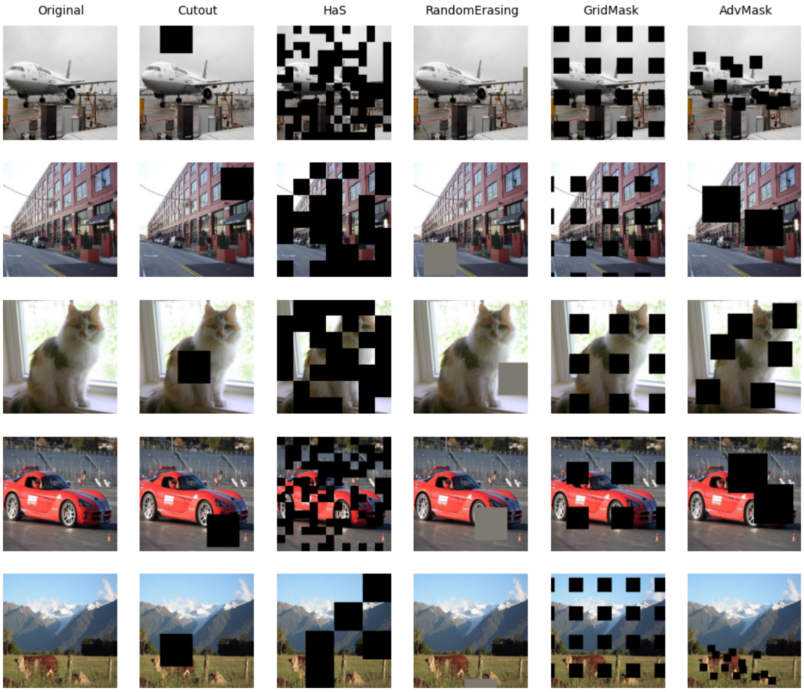


图 3 图像擦除方法对比

4.1.3 图像混合

近年来,图像混合数据增强受到越来越多的关注.这些方法主要通过将两幅或多幅图像或图像的子区域混合为一幅来完成.通过混合多个图像,或者混合多个图像中的感兴趣目标到一张图像中,可以模拟现实中可能出现的多目标场景,从而可以增强模型对这一场景的泛化性能.

Inoue 等人提出了一种数据增强方法[18],称为配对样本,通过将训练集中随机选择的两幅图像合成一幅新图像来扩大数据集.利用从训练集中随机选取的两幅图像,可以从 N 个训练样本中生成N^2个新样本,即配对样本.利用从训练集中随机选取的两幅图像,可以从N个训练样本中生成N^2个新样本.配对样本的工作流程如图 4 所示.由于使用的合成方法是将两幅图像在每个像素上的强度进行平均,所以配对样本方法的一个限制是合成的图像在视觉上没有意义.另一个局限性是这种增强技术只适用于分类任务,因为无法确定合成的混合图像的掩膜或边界框.混合图像标签遵循用于合成的第一幅图像,因此这不是一种标签保持方法.由于两幅图像在混合图像中的权重相等,除非标签 A 和标签 B 是相同的标签,否则分类器无法从混合图像中正确预测第一幅图像的标签(图 4 中标签 A).因此,训练错误率可能高于无配对样本的情况.在 CIFAR-10 等图像数据集上的实验表明,该方

法能够将验证集上的分类错误率降低 3.1 % ~ 28.8 %.

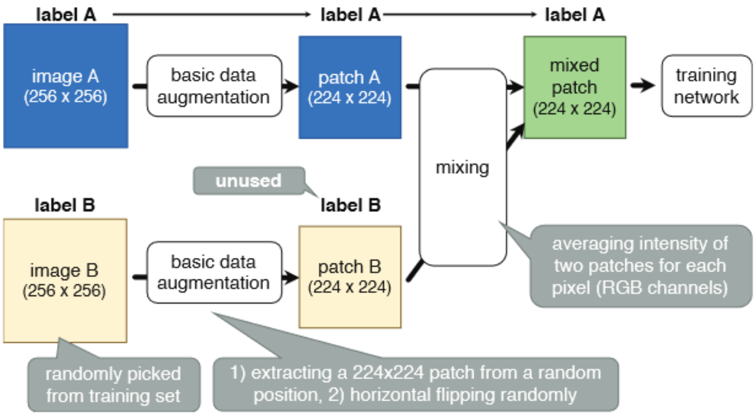


图 4 配对样本工作流程[18]

Zhang 等人讨论了一种更通用的合成方法,Mixup[22].Mixup 不只是对两幅图像的强度进行平均,而是对样本对和它们的标签进行凸组合.因此,Mixup 在数据增强和监督信号之间建立了一种线性关系,并且能够正则化神经网络,使其在训练样本之间具有简单的线性行为.这种线性行为减少了在训练样本之外进行预测时的不良振荡量.另外,从奥卡姆剃刀的角度来看,线性是一种很好的感性偏差.实验结果表明 Mixup 使得决策边界从类到类线性过渡,提供了更平滑的不确定性估计,因此 Mixup 训练的模型在模型预测和训练样本之间的梯度范数方面更加稳定.

与配对样本和 Mixup 相似,Yun 等人提出了 CutMix[23].与 Mixup 相比,CutMix 用另一幅图像的块替换掉被移除的区域,而不是简单地从训练集中移除像素或混合图像,CutMix 可以生成更自然的图像.图像的真实标签也根据组合图像的像素数量成比例混合. CutMix 具有训练过程中不存在无信息像素的特性,使得训练更高效,同时 CutMix 保留了区域 dropout 关注物体非判别部分的优势.添加的图像块通过要求模型从局部视图中识别目标来进一步增强定位能力.CutMix 与 Mixup 类似,通过对图像和标签进行插值来混合两个样本.虽然 Mixup 提高了分类性能的,但 Mixup 样本往往是非自然的.CutMix 通过将图像区域替换为另一幅训练图像的块来克服该问题.图 5 中展示了 CutMix 方法的样本示例,经过 CutMix 生成的数据可以在单个图片中包含多个样本,因此对于两个不同的目标都有显著区域.与此不同的是,Cutout 方法无法做到这一点,因为图中有一个区域是被完全抹掉的.

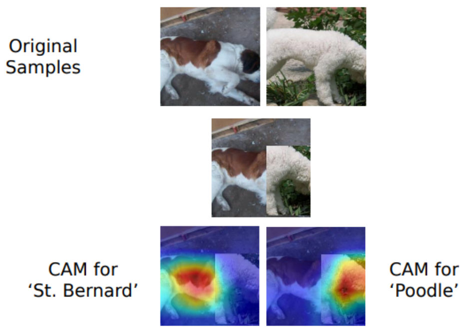


图 5 CutMix 示例及 CAM 可视化结果[30]

为了进一步防止记忆,同时以同样的方式保留数据分布,Harris 等在 CutMix 的基础上加入任意形状的掩码,

从而提出了 Fmix(Feature Mixup)[38].它是 Mixup 的一种扩展,Mixup 是一种数据增强技术,它将两个不同的输入样本混合在一起,以创建一个新的训练样本.FMix 通过混合特征而不是样本来实现这一点.具体而言,FMix 通过在输入图像中加入随机的掩码,并将其与另一幅图像的相应区域混合,从而生成一个新的图像.这个混合过程是基于一组随机生成的掩码进行的,每个掩码都与一个随机的图像匹配.通过这种方式,FMix 可以在保持图像语义完整性的同时,扰动输入图像的像素.FMix 在不增加训练时间的情况下,提高了 MixUp 和 CutMix 在一系列数据集和问题设置中的性能.

与混合多个样本不同,AugMix[31]是一种混合随机生成的增强操作和使用 Jensen-Shannon 损失来增强一致性的数据处理技术,它首先将多个增强操作混合为多个增强链,每个增强链由随机选取的 1～3 个增强操作组成,然后使用 Jensen-Shannon 散度作为一致性损失,在同一输入图像的不同增强链中执行分类器的一致性嵌入,最后将多个增强链的结果以凸组合的方式混合在一起.因此,整个过程通常是将同一图像在不同增强链中产生的结果进行混合,然后再将增强链的结果和原始图像结合起来.其实现如图 6 所示. 最终的图像包含了操作的选择、操作的程度、增强链的长度和混合权重几个随机性来源.AugMix 提高了图像分类器在数据偏移下的鲁棒性和不确定性估计.

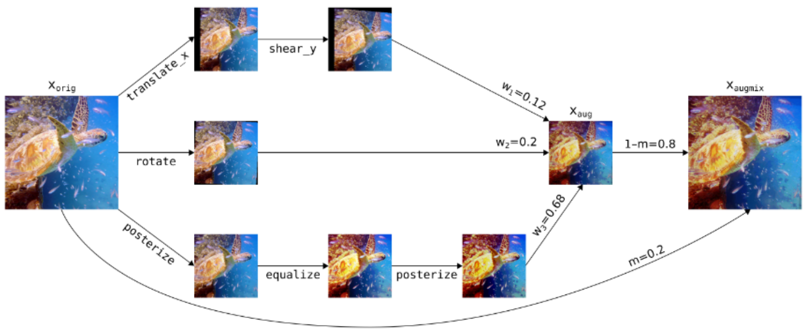


图 6 AugMix 流程[31]

与混合样本的思想类似,Junlin Han 等人提出了 You Only Cut Once (YOCO)[32],YOCO 对一幅图像在高度或宽度维度上切割成两幅相等的图像,在每一块内独立进行相同的数据增强,然后将增强块拼接在一起,形成一幅增强图像.通过应用 YOCO 可以提高每个样本生成的多样性,并鼓励神经网络从部分信息中识别对象.同时 YOCO 具有无参数、易使用、在不增加复杂度下改善所有增强操作效果的特点.

深度神经网络是计算机视觉、语音识别和语言翻译等系统的核心.然而,这些系统只有在评估与训练集非常相似的实例时才能表现良好.当在稍有不同的分布上进行评估时,神经网络常常会出现高置信度的错误预测.为了解决这个问题,Manifold Mixup[39]提出了一个新的方法.通过在随机线性组合的隐藏层特征上进行 Mixup 来实现数据增强.具体来说,它在神经网络的隐藏层特征上插入一组 Mixup 层,这些层将来自不同输入样本的隐藏层特征进行线性混合,从而创建新的隐藏层特征.这样做的目的是让模型在更广泛的特征空间上进行学习,从而提高其泛化能力.与传统的 Mixup 不同,Manifold Mixup 在隐藏层特征上进行 Mixup,而不是在输入样本上进行 Mixup.这种方法可以带来更好的鲁棒性和更好的泛化能力,尤其是在具有复杂分布的数据集上.为了得到带有软目标的混合样本,作者还在相关的 one-hot 标签中执行相同的线性插值.实验结果表明,Manifold Mixup 改进了神经网络在多层上的隐藏表示和决策边界.图 7 中展示了几种典型的图像混合类方法的示例图片,可以看到图像混合技术可以通过混合不同类别的图像,极大地提高数据的多样性,这有助于降低模型的过拟合风险.

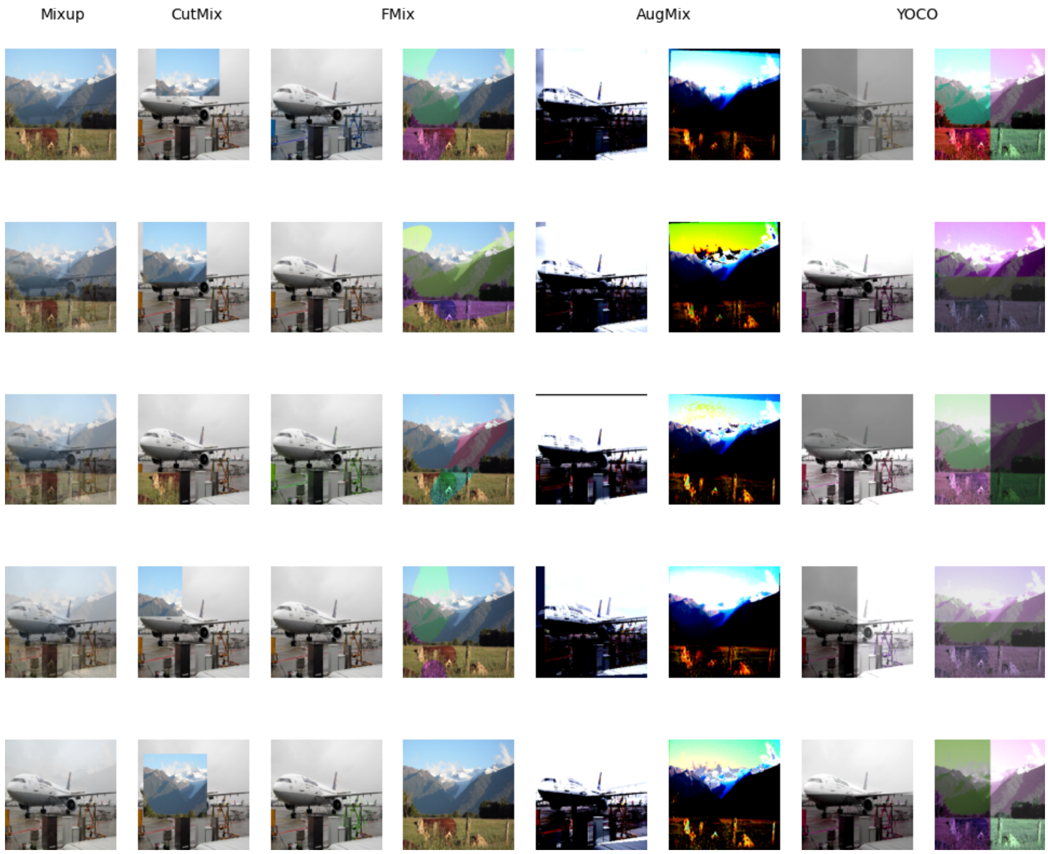


图 7 图像混合方法对比

4.2 深度学习相关方法

4.2.1 自动增强技术

与人工设计数据增强方法不同,研究者试图自动搜索增强方法以获得更好的性能.自动增强一直是深度学习研究的前沿领域,并得到了广泛的研究.自动增强基于不同的数据具有不同的特征,因此可以借助不同的数据增强方法的优势来生成更加多样化的训练集,从而比人工设计的方法带来更多的好处.但是这类方法的局限性在于,需要针对数据集搜索得到最优的数据增强操作空间和参数,因此,这类工作通常在最普遍且有代表性的图像分类任务上进行,对于目标检测和语义分割等任务,这类方法较少应用.

Cubuk 等人描述了一个称为 AutoAugment[20]的方法,相比于人工设计的数据增强方法, AutoAugment 自动地搜索可以达到最佳分类精度的数据增强策略.具体来说,AutoAugment 由搜索算法和搜索空间两部分组成.搜索算法使用强化学习技术,包含控制器和训练算法两部分,控制器在每一步使用 softmax 来预测一个决策, 并将预测作为输入嵌入到下一步的决策生成过程中,以寻找验证精度最高的最佳策略.搜索空间包含许多子策略,详细说明了各种增强操作和应用这些操作的幅度,例如旋转和平移以及这两种操作的概率和参数,而搜索算法就是要找到这些子策略的最优组合. 图 8 展示了不同批次的图片在五种不同的子策略组合下生成得到的增强数据.其中每个策略包含了两种操作,每种操作包含施加数据变换的概率和数据变换的强度两个参数.对于每个图片而言,AutoAugment 会随机选择一个子策略来生成增强图片.总的来说,自动增强方法类的一个关键挑战是如何从庞大的候选操作搜索空间中选择有效的增强策略,搜索算法往往需要上千个 GPU 小时才能搜索到有效的增强策略.

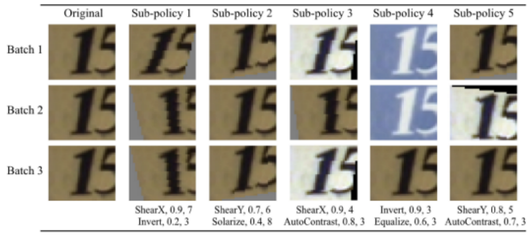


图 8 AutoAugment 示例[20]

为了降低 AutoAugment 的时间成本,Lim 等人受贝叶斯数据增强[40]的启发,提出了 Fast AutoAugment [21],通过基于密度匹配的更有效的搜索策略来寻找最优的增强策略.该方法通过学习将增强数据作为训练数据的缺失数据点的增强策略来提高给定网络的泛化性能,在策略搜索阶段通过贝叶斯优化来利用和探索一系列推理时间增强来恢复那些缺失的数据点.与 AutoAugment 相比,Fast AutoAugment 不需要对策略评估进行任何反向传播的网络训练.此外,该算法在 CIFAR - 10、CIFAR - 100、SVHN、ImageNet 等多种数据集的图像识别任务上取得了相当的性能,同时将搜索时间缩短了一个数量级.同时,Ho 等人提出了 Population Based Augmentation (PBA)[42],旨在减少 AutoAugment 的时间成本, 该方法通过生成非平稳的增强策略计划而不是固定的增强策略来实现,PBA 学习增强策略的时间表,而不是固定的策略,这种选择是 PBA 的效率提高的主要原因.与采用当前轮次数无关的相同变换的固定增强策略相反,PBA 利用基于种群的训练算法生成增强调度,为每个训练轮次定义最佳增强策略.所以 PBA 能够以较少的计算时间在多个数据集上匹配 AutoAugment 的性能.

Cubuk 等人提出了 RandAugment[44],超越了包括 AutoAugment 和 PBA 之前所有的自动增强技术.自动增强技术通常是在一个小的代理任务上对增强策略进行单独搜索,然后将搜索到的结果迁移到更大的目标任务上.这种搜索方法依赖于一个很强的假设:代理任务可以为目标任务提供一个有效的可预测性结果.但是对于数据增强研究而言,数据增强方法十分依赖于模型和数据集的大小,因此代理任务往往只能提供一个次优的结果,而不是最优的.RanAugment 通过把数据增强策略的参数整合到训练模型的超参数中去,可以移除在代理任务上单独的策略搜索,从而可以极大地减少数据增强的搜索空间, 因此简单的网格搜索足以找到一个数据增强策略.从结果来看,RandAugment 的结果优于所有采用单独搜索阶段的学习增强方法.此外,由于 RandAugment 的参数化,正则化强度可以根据不同的模型和数据集大小进行调整,使得 RandAugment 可以在不同的任务和数据集上统一使用,并且可以开箱即用.

现有的数据增强方法虽然在测试时间没有产生额外的延迟,但往往需要更多的训练次数才能有效.Emirhan Kurtulus 等人提出了一个名为 Tied -Augment 的通用框架[41],该框架前向传播过程中产生同一图像的两个增强视图,除了分类损失之外,他们还增加了一个相似度项来增强两个增强视图的特征之间的不变性,通过特征相似性来增加数据增强的有效性.实验结果表明,Tied-RandAugment 在 ImageNet 上可以比 RandAugment 提高 2.0 %.

传统的自动增强通常应用多个图像变换操作来生成数据,TrivialAugment[45]数据增强策略用更简单的方式实现了不错性能表现的数据增强策略.使用 NAS 方法自动搜索的数据增强的方法虽然是有效的,但局限在于需要权衡搜索效率和数据增强的性能.为了解决这个问题,论文提出的 TrivialAugment 数据增强策略,相比于之前的数据增强策略是无参数的,每张图片只使用一次数据增强方式,因此相比于 AutoAugment、PBA 乃至 RandAugment,它的搜索成本几乎是可以忽略的.TrivialAugment 采用了和 RandomAugment 相同的数据增强空间,数据增强被定义为由一个数据增强函数 a 和对应的强度值 m（部分数据增强函数不使用强度值）组成,工作原理如下：它以一幅图像 x 和一组增强操作 A 作为输入.然后简单地从 A 中随机均匀地采样一个增强,并将这个增强应用于给定的强度为 m 的图像 x,其中 m 从可能的强度集合{ 0,. . .,30 }中随机均匀地采样,然后返回增强后的图像.即对于每个图像,TrivialAugment 均匀采样一个数据增强函数和一个强度值,然后返回增强后的图片.

同时,对抗数据增强策略在许多任务的模型泛化性上表现出显著的提升,但是现有的对抗数据增强方法需要复杂的参数调整来避免过度强烈的增强,过度的增强会导致对泛化至关重要的图像特征丢失.Teppei Suzuki 提出了一种新的基于对抗策略的数据增强优化方法[46],称为 TeachAugment,它通过利用教师模型,可以在不需要复杂参数调整的情况下生成信息丰富的转换图像.具体来说,对增强策略进行搜索,使增强图像对目标模型具有对抗性以及对教师模型具有可识别性.与以往的对抗数据增强方法不同,得益于教师模型,TeachAugment 不需要先验和超参数,避免了过强的增强会破坏图像的固有含义,因此,TeachAugment 不需要进行参数调优来保证变换后的图像具有可识别性.Teppei Suzuki 还提出了使用神经网络进行数据增强,简化了搜索空间的设计,并允许使用梯度方法对数据增强进行更新.

由于目前底层的增强方法大多依赖于手工设计的操作.此外,对一个数据集有用的增强策略可能不能很好地迁移到其他数据集.以上自动增强类的方法的增强策略对所使用的数据集不具有自适应性,阻碍了这些方法的有效性.Tsz-Him Cheung 等人提出了一种新颖的 AutoDA 方法[47],称为 AdaAug,以类依赖和潜在实例依赖的方式有效地学习自适应增强策略.AdaAug 使用一个识别模型来学习每个数据实例的底层增强策略,并通过一个交替的利用和探索过程来使用可微的工作流更新增强策略.在数据增强阶段,AdaAug 训练一个分类器,经过若干个步骤,然后通过探索阶段对分类器进行验证,并更新策略以最小化验证损失.实验结果表明该方法取得了不错的实验结果.

除了自动化的搜索增强策略之外,自动增强方法类中的方法还被用于改善其它类方法中存在的潜在问题,例如,数据增强过程可能会引入有噪声的增强实例,尤其是图像擦除类方法,该类方法中随机化的擦除可能会将感兴趣目标完全移除,给推理带来负面影响.因此,Gong 等人提出了 KeepAugment[48],使用特征图(Saliency map)来检测原始图像上的重要区域,然后在增强过程中保留这些信息区域不受干扰.KeepAugment 采用两种方式达到上述目标,如图 9 所示:一种是在图像变换的过程中,显著性区域不受干扰;另一种是将显著性区域直接附加在变换后的图像上.另外,在 Tian 等人的工作中[49],作者观察到训练后期的增强操作影响更大,并提出了增强权重共享(Augmentation-wise Weight Sharing, AWS)策略作为数据增强搜索方法的验证过程.与 AutoAugment 相比,这项工作显著地提高了效率,并且可以直接在大规模数据集上进行搜索.与自动增强方法以离线方式搜索策略不同,Lin 等人[50]将增强策略制定为参数化的概率分布,分布的参数被视为超参数,进一步提出了一个双层框架,允许在网络训练的同时优化分布参数,称为 OHL-Auto-Aug. OHL-Auto-Aug 消除了重新训练的需要,大幅降低了整个搜索过程的成本,同时获得了比基线模型显著的精度提升.与 AutoAugment 相比,OHL-Auto-Aug 在 CIFAR-10 上实现了 60 倍的速度提升,在 ImageNet 上实现了 24 倍的速度提升.与上述思想类似,Shiqi Lin 等人提出了一种有效的方法,称为 SelectAugment[43].SelectAugment 根据样本内容和网络训练状态,以确定的和在线的方式选择待增强的样本.具体来说,在每个批次中,首先确定增强比例,然后在该比例下决定是否对每个训练样本进行增广.他们将这一过程建模为两步马尔可夫决策过程,并采用分层强化学习(Hierarchical Reinforcement Learning,HRL)来学习增强策略.通过这种方式,可以有效地缓解随机选择样本进行增广所带来的负面影响,提高数据增强的有效性.

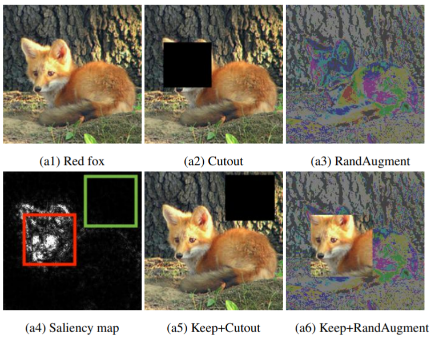


图 9 KeepAugment 示例图[48]

4.2.2　特征增强

传统的数据增强方法只是应用在图像空间中,对原始数据集进行操作得到增强数据.基于特征数据的数据增强方法不是在输入空间中进行增强,而是在学习得到的特征空间中进行数据生成.这类方法可以有效地增加训练数据的多样性,从而提高模型的鲁棒性和泛化能力.此外,由于在特征空间中进行的数据增强不涉及像素级别的变换,因此可以更加高效地进行,还可以避免一些由于传统数据增强方法所引入的失真和噪声.该类方法的动机是模型实际上用于下游任务的数据是特征图, 而不是图像空间中的数据.虽然直接在图像空间中生成数据在人眼看来是有意义的,但很可能会引入噪声信息.相比之下,特征空间通常是一个维数较小的空间,因此直接在特征空间中生成数据具有不错的应用价值.

DeVries 和 Taylor 称当沿着流形遍历时,与输入空间相比,在特征空间中遇到真实样本的可能性更大[51].因此,可以使用一些离散的数据增强操作对学习得到的特征空间中的数据向量表示进行操作,例如,添加噪声、最近邻插值和外推各种增强方法.这些方法的是领域无关的,不需要专业知识,因此可以应用于许多不同类型的问题.例如类别不平衡问题中提出了一种在特征空间中的样本之间进行插值得到新样本的方法.类似于输入空间中的数据集增强已经成为视觉识别任务的标准,特征增强类方法提出特征空间中的数据集增强作为一种通用的领域无关框架,可以提高有限标记数据下的泛化能力.

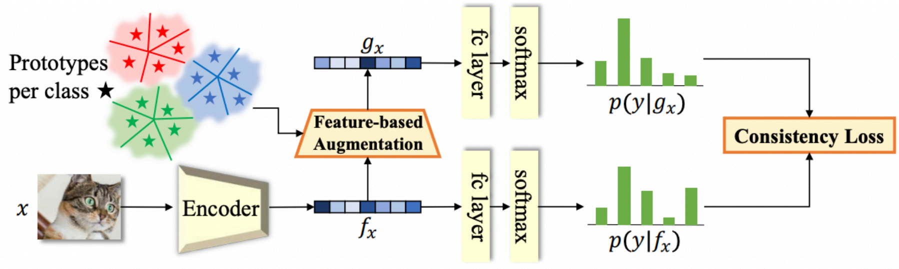


图 10 基于特征的数据增强方法[47]

最近,Kuo 等人提出了 FeatMatch[52],这是一种新颖的基于学习特征的细化和增强方法,可以产生各种复杂的变换.如图 10 所示,FeatMatch 使用了聚类提取到的类内和跨类原型表征中的信息,同时通过将特征存储在内存库中,使用跨迭代计算的特征,避免了大量的额外计算.FeatMatch 设计了一个模块,该模块通过对从数据集中其他图像的特征中提取的一小组代表性原型进行软注意力来学习改进和增强输入图像特征,由于所提出的模块是在特征空间中学习和执行的,因此可以使用十分多样的数据转换方法.

Li 等人中提出了一种隐含的数据增强方法 MoEx,称为矩交换[53],它通过鼓励模型利用潜在特征的矩信息,将一幅训练图像的学习特征的矩替换为另一幅训练图像的学习特征的矩,并对目标标签进行插值迫使模型从这些矩中提取训练信号和归一化特征.尽管 MoEx 可以交换两个图像的形状或者风格信息,但是 MoEx 不需要风格迁移预训练模型,只需要在训练过程中计算特定层中的特征均值和方差,并把它们重新注入到其他样本的特征表示中,因此该方法计算代价小,速度快,并且完全是在特征空间中的操作,可以有效地将它与现有的基于图像输入的增强方法结合起来.

但是总的来说,在特征空间中进行数据增强存在一些问题,如果数据增强技术不当,可能会生成与原始数据高度相关的变换数据,导致模型在测试数据上表现不佳.同时,特征空间往往缺乏人眼可理解的含义,所以可能会增加噪声或引入不真实的变换,使得模型在训练数据上表现良好,但在未知数据上的泛化能力却下降.此外,增强后的图像可能会变得不自然,导致模型学习到了不真实的特征,这可能会降低模型在实际应用中的性能.特征空间数据增强的计算成本也可能很高,因为它需要对输入数据进行复杂的变换和扩充.如果数据集很大,这可能会带来较大的训练代价,所以特征增强方法较少应用在实际的训练过程中.

4.2.3 深度生成模型

数据增强的最终目标是从代表数据集的生成机制的分布中抽取样本.因此,我们生成的数据分布应该与原始数据分布相同,这是深度生成模型的核心思想.在所有的深度生成模型方法中,生成对抗网络(GANs)[25]是非常具有代表性的方法.一方面,生成器可以帮助生成新的图像.另一方面,判别器确保新生成的图像与原始图像之间的差距不会太大.尽管 GAN 是一种强大的技术,可以进行无监督的生成来增强数据[19],但如何生成高质量的数据并对其进行评估仍然是一个具有挑战性的问题.在本小节中,我们将介绍一些基于 GAN 的图像数据增强技术.

Isola 等人提出了基于条件对抗网络[54]的 Pix2Pix[55],用于学习从输入图像到输出图像的映射. Pix2Pix 由生成器和判别器两部分组成.生成器G经过训练产生输出,使其不能被经过对抗性训练的判别器与"真实"图像区分开.判别器D被训练来检测伪造图像,从而增强了生成器的能力.同时,随着生成器的改进,判别器的能力也有所提高.然而,训练 Pix2Pix 需要大量的配对数据,收集配对数据具有一定的挑战性.因此, Zhu 等人中提出了 CycleGAN[56]模型,与 Pix2Pix 不同,它可以在在没有配对样本的情况下来学习图像从源域 X 到目标域 Y 的转换.图 11 展示了配对训练数据和非配对训练数据的差异,在现实世界中,获得配对的训练数据很困难,而在一些不存在配对训练数据的任务中,如风格转换,产生定性结果是非常必要的,因此应用这种技术非常有必要.

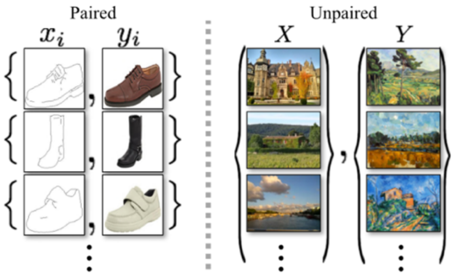


图 11 配对图像和未配对图像的例子[55]

为了得到集合 X 到集合 Y 的映射,建立双射关系是必不可少的.同时,为了防止将所有输入图像映射到 Y 中的同一幅图像,该模型包含 G 和 F 两个映射,其中 G 是从 X 到 Y 的映射,F 是从 Y 到 X 的映射.映射关系如图 12 所示.变换过程如图 13 所示.在生成器 G 的作用下,将输入 x 映射到域 Y 中的 y,然后在生成器 F 的帮助下将 y 映射到域 X 中的$\hat{x}$.这样,通过比较x和$\hat{x}$可以评估生成器的能力.同时,判别器D_x鼓励 F 将$\widehat{X}$转化为与域X不可区分的输出,D_Y鼓励 G 将 X 转化为与域 Y 不可区分的输出.

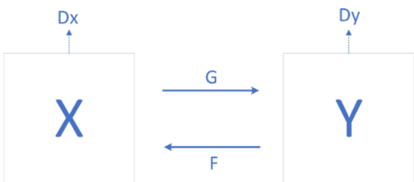


图 12 域 X 和域 Y 之间的转换[56]

因此前向循环一致性损失为:

$$x \rightarrow G(x) \rightarrow F(G(x)) \approx x$$

后向循环一致性损失为:

$$y \to F(x) \to G(F(y)) \approx y$$

CycleGAN 和 Pix2Pix 的对比实验表明,即使没有成对的训练数据,CycleGAN 也能得到令人满意的结果.

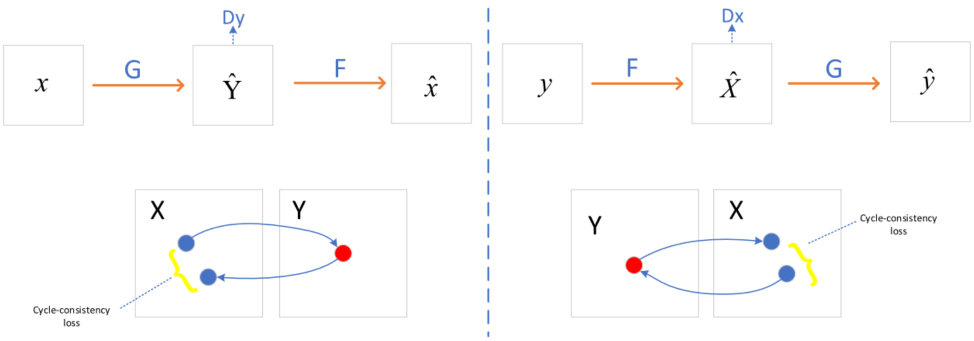


图 13 Cycle GAN 模型[56]

CycleGAN 虽然可以用于风格迁移、物体变形、季节迁移、照片增强等任务,但也存在一定的局限性.例如,如果任务是两个域之间的转换,则需要训练从域 X 到 Y 和 Y 到 X 两个 GAN 模型.如果任务是 n 个域之间的转换,则需要在每两个域之间训练 n * (n-1)个 GAN 模型.也就是说,转换只能从一个领域完成到另一个领域.为了解决 CycleGAN 的问题,Choi 等人提出了 StarGAN[57]来提高处理两个以上域的可扩展性和鲁棒性. StarGAN 通常只需要建立一个 GAN 模型,即可进行多个域之间的图像到图像的转换.在生成阶段,只需要向生成器提供源图像和表示目标域的属性标签. 判别器迫使生成器从源图像中生成一个无法区分的图像,生成的结果可以在目标域进行分类.然后,StarGAN 将域标签作为一个额外的输入,并为每个域学习一个确定性映射,这可能导致给定输入图像,每个域的输出相同.为了解决 StarGAN 的问题,Choi 等人提出了 Star GAN v2[58],它是一种可扩展的方法,可以跨多个域生成不同的图像.在这项工作中,研究人员将图像的域和风格分别定义为视觉上不同的类别组和每个图像的具体外观.例如,狗可以作为域来使用,但狗的种类很多,如拉布拉多犬、哈士奇犬等.因此,特定的犬种可以被视为图像的风格.这样,StarGAN v2 可以将一个域的图像转换为目标域的不同图像,并且支持多个域.通过这种方式,StarGAN v2 引入了两个模块:将随机高斯噪声转化为风格编码的映射网络和从图像中提取风格编码的编码器.然后,生成器在多个域上合成多样化的图像,判别器在多个域中区分真假图像.

尽管近年来深度生成模型取得了不错的发展,但是这类方法仍然很少被作为数据增强方法用于深度模型训练之中.一方面,为了获得性能不错的深度生成模型,需要耗费大量算力进行训练.另一方面,深度生成模型自身的训练需要大量的训练数据,而数据增强方法的提出本身是为了解决训练数据不足的问题,这就产生了一定的矛盾:为了增加训练数据量而使用数据增强方法,但是深度生成模型本身需要大量训练数据才能达到优异的性能.这使得深度生成模型很难应用于实际任务中.但是作为很重要的一类数据生成研究方向,我们还是介绍了其中代表性的方法和核心思想.

5 评估指标

在本节中,我们将结合具体任务来对数据增强的评估指标进行深入讨论.基于我们的数据增强分类方法,我们对数据增强技术在多个计算机视觉的核心任务上进行了评估,包括语义分割,图像分类和目标检测等三个经典任务.其中语义分割任务[59][60]的核心目标是将标签或类别与图片的每个像素关联的,它用来识别构成可区分类别的像素集合,并区分不同语义类别的目标.图像分类[61][62][63]的核心目标是根据图像的内容对图像进行分类.目标检测[64][65]的核心目标是检测数字图像中的关键目标,并对目标进行分类,简单来说就是需要回答“某处有某物”这样的问题.我们通过对这些任务的评估来展示数据增强技术的效果.

本章的组织方式如下,从 5.1 到 5.3 节,针对这三个计算机视觉任务,我们首先使用最常用的公共数据集和经典的神经网络模型来评估数据增强方法的效果,通过使用不同类别的数据增强方法来展示不同类别的数据增强方法在提高性能方面的有效性.同时,为了确保实验的公平性,我们除了数据增强方法不同之外,控制其它所有的实验设置保持完全一致.更进一步地,在 5.4 节中,我们具体到最为典型的图像分类任务上,全面展示多种最具代表性和普遍使用的数据增强方法所带来的具体的效果提升,从而可以方便读者对不同方法的性能进行比较.

5.1 语义分割

语义分割是一门涉及计算机视觉、模式识别与人工智能等多个研究领域的方向,在虚拟现实、工业自动化、自动驾驶、医学图像分析等不同领域有着广泛的应用,具有重要的研究意义[59][66].其目标是为图像中的每一个像素分配一个预先定义好的表示其语义类别的标签[67].PASCAL VOC 数据集[68]是语义分割任务最常用的数据集之一.为验证数据增强在语义分割数据集及相应模型上的有效性,我们在该任务最为常用的几个深度学习模型上进行了实验验证,包括 DeepLabV3+[2], PSPNet[69], GCNet[70], 和 ISANet[71]等,我们采用了 Intersection over Union(IoU)[72]这一指标来衡量模型的泛化性能,通过比较在有无数据增强条件下进行训练的模型表现,验证数据增强技术的有效性.

具体而言,我们的实验设置参考了[73],并按照图 1 中的分类方法应用了语义分割任务中常用的数据增强技术来评估其有效性.这些数据增强技术涵盖了基本的图像变换操作（如裁剪、旋转、平移等）、图像擦除类方法（如随机擦除、Cutout、GridMask 等）、以及图像融合方法（如 Mixup、CutMix 等）.自动增强方法由于依赖于优化得到的操作空间和参数,且现有方法通常仅中图像分类基准数据集上进行了参数调优,因此在语义分割任务中应用较少.我们将在 5.2 和 5.4 章节的实验中展示这些方法在图像分类任务中的表现.

表 4 展示了在不使用和使用数据增强技术时,几种语义分割模型的平均 IoU,其中部分结果来自[73],尽管模型架构各不相同,数据增强技术为这些模型带来了显著的性能提升.例如,GCNet 的性能提升了 1.15%,而 ISANet 的性能提升幅度高达 2.71%.值得注意的是,这些性能提升几乎没有增加网络训练的计算开销,即在相同的网络架构和训练策略下,以近乎相同的训练成本实现了更高的性能.因此,我们证明了数据增强技术能够有效提升模型训练效果,增强模型的泛化能力.

表 4 数据增强在语义分割任务上的性能提升

模型	w/o aug (%)	w/ aug (%)	改进(%)
DeepLabV3+	75.32	75.81	0.49
PSPNet	73.38	74.42	1.04
GCNet	71.86	73.01	1.15
ISANet	71.65	74.36	2.71

5.2 图像分类

图像分类任务是指给定一幅输入图像,通过某种分类算法来判断该图像所属类别.作为计算机视觉领域中最热门的研究方向之一,图像分类是实现物体检测[74][73][75]、人脸识别[76][77]、姿态估计[78][79]等众多应用的重要基础,因此其有很高的学术研究和广泛的科技应用前景[59].在本实验中,我们基于经典的图像分类深度网络模型架构和基准数据集进行实验,所使用的模型包括 WideResNet-28-10[80]、DenseNet-121[81]和 Shake ResNet[82],数据集则为 CIFAR-10/100 数据集和 SVHN[14][83].本小节的实验涵盖了基本图像处理方法、图像擦除方法、图像混合方法以及基于深度学习的自动增强类方法.

表 5 总结了使用数据增强前后的图像分类模型的实验结果,其中部分结果来自[73].可以观察到,对于 CIFAR-10/100、SVHN 数据集,无论是与 DenseNet、Wide-ResNet、Shake-ResNet 的模型架构组合,数据增强均显著提升了分类模型的准确率.总体上,实验结果符合模型基准准确率越低,数据增强的效果越显著的规律.在三个数据集中,CIFAR-100 上的性能提升最为显著,其次是 CIFAR-10,SVHN 的提升最小.

表 5 数据增强对 CIFAR-10、CIFAR-100 和 SVHN 图像分类精度的提升

数据集	模型	w/o aug (%)	w/ aug (%)	AAI (%)
CIFAR-10	DenseNet	94.15	96.32	2.17
	Wide-ResNet	93.34	96.98	3.64
	Shake-ResNet	93.70	96.86	3.16
CIFAR-100	DenseNet	74.98	79.62	4.64
	Wide-ResNet	74.46	81.91	7.45
	Shake-ResNet	73.96	80.37	6.41
SVHN	DenseNet	97.91	97.98	0.07
	Wide-ResNet	98.23	98.31	0.80
	Shake-ResNet	98.37	98.40	0.30

5.3 目标检测

目标检测是计算机视觉中非常重要的一项任务,旨在检测数字图像中的某一类视觉对象（即目标）的实例,并标注出其位置和类别.目标检测任务通常分为两个子任务:目标检测和目标定位.目标检测的目标是确定图像中是否存在目标以及它们的类别,而目标定位是找到目标的位置并用边界框进行标注.这种任务在许多应用领域中都有很广泛的应用,例如实例分割[84]、目标追踪[85]、视频监控、自动驾驶、智能家居、机器人导航、医学图像分析等.COCO 数据集[86]是最为经典的目标检测基准数据集之一,其包含超过 20 万张图片和 80 个目标类别,且每个类别都有大量的实例标注.为了验证数据增强技术在目标检测任务上的效果,我们在两个常用的目标检测深度模型 FasterRCNN[87]和 CenterNet[88]上应用了多种数据增强方法,这些方法涵盖了基本图像操作类、图像擦除类和混合类等.

实验结果见表 6,其中部分结果来源于[73],从结果中可以观察到, 数据增强技术对目标检测模型的精度带来了显著提升.尤其是在 Faster R-CNN 模型上,mAP 提高了 2.40%,AP50 平均提高了 0.8%,AP75 平均提高了 0.5%.类似地,在 CenterNet 模型上,数据增强方法也同样带来了性能的提升.

表 6 数据增强方法在目标检测任务上的检测精度提升

指标		Faster R-CNN	CenterNet
mAP (%)	w/o aug	36.40	41.42
	w/ aug	36.91	41.50
	API	+2.51	+0.08
AP50 (%)	w/o aug	57.20	58.29
	w/ aug	57.97	58.37
	API	+0.77	+0.08
AP75 (%)	w/o aug	39.50	45.53
	w/ aug	40.03	45.49
	API	+0.53	-0.04

5.4 数据增强方法效果比较

本小节中,我们在多个图像分类基准数据集上应用不同的数据增强方法,从而探究各个数据增强方法对各个模型性能的具体提升效果,我们采用了几种常见的深度学习模型作为我们的基准模型,包括 ResNet、Wide-ResNet 以及 Shake-ResNet,这些模型具有不同的复杂性和参数数量,以便我们可以更全面地评估数据增强方法的效果.最终,我们用到的评测指标为测试集上的分类准确率.

如表 7 所示,我们选取了各个类中最具代表性和广泛使用的数据增强方法.实验结果表明,不同的数据增强方法在不同的数据集和模型上产生了不同的效果.相比较而言,基于深度学习的自动增强类方法通常能取得更好的性能表现,但是由于这类方法需要首先针对数据集来获取数据增强操作和参数空间,所以相比较于图像擦除类和图像混合类而言,实际场景下的应用并不是很容易.相反,图像擦除类和图像混合类由于几乎不需要先验信息,可以“即插即用”,因此当下主流的模型训练几乎都内嵌了这两类方法.从实验结果也可以看出,这两类方法同样带来了显著的效果提升.尤其是对于较为简单的模型,比如 ResNet-18,这两类方法带来的性能提升和自动

增强类方法差距不大.

表 7 数据增强方法在图像分类任务上的效果对比

数据集	方法	ResNet-18	Wide-ResNet-28-10	Shake-ResNet
CIFAR-10	Mixup[22]	96.55	96.89	96.50
	CutMix[30]	96.58	96.87	96.45
	Cutout[17]	96.01	96.92	96.96
	GridMask[24]	96.38	97.23	96.91
	AdvMask[37]	96.32	96.93	96.90
	RandomErasing[35]	95.69	96.92	96.46
	AutoAugment[20]	96.07	97.01	97.21
	Fast-AA[21]	95.89	96.77	96.42
	RandAugment[44]	96.37	96.88	97.01
	TrivialAugment[45]	96.20	97.11	97.27
	KeepAugment[48]	96.10	97.30	97.35
	SelectAugment[43]	96.18	97.33	97.38
CIFAR-100	Mixup[22]	79.33	82.28	79.20
	CutMix[30]	79.53	82.60	79.32
	Cutout[17]	78.04	81.56	79.47
	GridMask[24]	75.15	80.32	79.14
	AdvMask[37]	78.38	80.56	79.88
	RandomErasing[35]	75.95	80.50	78.89
	AutoAugment[20]	79.56	82.89	82.27
	Fast-AA[21]	79.10	82.69	81.33
	RandAugment[44]	78.33	82.90	80.02
	TrivialAugment[45]	78.70	82.70	82.10
	KeepAugment[48]	80.26	82.01	82.49
	SelectAugment[43]	81.89	83.37	85.17

6 数据增强面临的挑战和未来研究方向

尽管在图像数据增强研究方面已经做出了大量的努力来提高深度学习模型的性能,但仍有一些开放的问题尚未完全解决.

数据增强的理论研究. 关于数据增强的理论研究相对不足.目前,数据增强带来的性能提升主要归因于其通过增加训练数据集的多样性,使模型能够可以学习到更多潜在的未见样例.然而,关于数据增强有效性的深入分析仍然不足.尽管遮盖类和混合类的很多方法已被证明有效,但如本文图 3、7、9 所示,部分增强后的图像已经丢失了原有的语义信息,模型如何利用这些样本来提升泛化能力仍需进一步研究.此外, 关于训练集的增强幅度和增强尺度的选择也缺乏理论支持.例如,随机选择的增强幅度虽然可以增加样本的多样性,但同时也可能引入噪声[43]的对于数据增强的可解释性研究将有助于解决这些问题,并为数据增强方法的设计与应用提供理论指导,推动该领域的发展.

数据增强方法的评估.训练数据的数量和多样性对模型的泛化能力至关重要.然而,由于缺乏统一的衡量标准,如何评价合成图像质量仍是一个开放的问题[91].当前阶段,研究人员通过以下几种方式来评估合成数据的质

量.首先,合成数据通常由人眼进行评估,这种方法耗时耗力且主观性强.常用的方法是利用 Amazon Mechanical Turk (AMT)来评估输出的真实性,它通过要求参与者对用不同方法合成的各种图像进行投票,来评估生成图像的质量和真实性.其次,一些研究将评估与具体任务相结合,即根据数据增强方法对有数据增强和无数据增强任务度量的影响来评估数据增强方法,如具有分类精度的分类任务和具有掩码 IOU 的语义分割.然而,目前并没有针对合成数据本身的评价指标.一般来说,无论任务是什么,评价指标都是基于个体数据的多样性和整体数据分布的一致性.数据质量分析可以帮助设计评价指标.

数据增强方法的设计.虽然图像数据增强技术可以应用于目标检测[89],语义分割[90]和图像分类[91]等各种计算机视觉任务,但数据增强方法是独立于任务的,操作是同时在图像数据和标签上进行的,而不同的任务下标签类型是不同的,所以目标检测任务的数据增强方法不能直接应用于语义分割任务.这就导致了图像数据增强方法的效率低下和可扩展性差.

类不平衡. 数据不平衡或极少数据会导致数据分布严重扭曲[92][93].这种情况发生的原因在于学习过程往往会偏向于多数类样例,从而难以有效地对少数类样例进行建模.合成少数类过采样技术(SMOTE)[94]则是通过对少数类进行过采样来缓解此问题.然而,过度采样可能会饱和少数类,导致过拟合.因此,我们期望生成的数据能够在保持多样性的同时模拟出与训练数据相似的分布,以避免过度拟合的问题.最终,需要提供更加严谨和深入的理论来支持数据增强方法的选择和设计,从而更好地处理数据不平衡和少数类问题.

生成数据的数量.数据增强的一个有趣之处在于,训练数据量的增加与性能的提升并不完全成正比.当达到一定量的数据时,继续增加数据并不会带来效果的提升.这可能是因为,尽管数据数量在增加,但数据的多样性并没有提高.目前,生成的训练数据量的大小通常是根据个人经验和大量实验结果设计的.此外,当原始数据集规模较小时,可能存在悖论.我们将面临如何在极少的数据基础上生成合格数据的挑战. 因此,应该生成多少数据才尽可能高地增强模型性能,以及训练数据的数据是否存在一个上界还有待进一步探索.

数据增强的选择与组合.由于各种数据增强可以组合在一起生成新的图像数据,因此数据增强技术的选择和组合至关重要.图像识别实验表明,Pawara 等人的组合方法的结果往往优于单一方法[95].因此,在进行数据增强时,如何选择和组合方法是一个关键点.然而,从我们的评估来看,适用于不同数据集和任务的方法并不相同.因此,增强方法必须针对每一个新的任务和数据集进行精心设计、实现和测试.

7 总结

随着深度学习的发展,对训练数据集的要求越来越严格.因此,我们认为数据增强是解决有限标记图像数据不足的有效方法.本文对各种计算机视觉任务中的图像数据增强方法进行了综述,并提出了一个分类法,总结了每个类别中具有代表性的方法.然后我们在图像分割、目标检测、图像分类等三个经典的计算机视觉任务中对这些方法进行了实验比较分析,证明了数据增强方法的有效性,并在图像分类任务上进一步对比了不同数据增强方法的效果.最后,我们讨论了当前面临的挑战并强调了未来的发展前景.

References:

[1] Deep learning in computer vision: principles and applications. CRC Press, 2020.

[2] Liu B, Zeng Q, Lu L, et al. A survey of recommendation systems based on deep learning. Journal of Physics: Conference Series. IOP Publishing, 2021, 1754(1): 012148.

[3] Torfi A, Shirvani R A, Keneshloo Y, et al. Natural language processing advancements by deep learning: A survey. arXiv preprint arXiv:2003.01200, 2020.

[4] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[5] Babyak M A. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models[J]. Psychosomatic medicine, 2004, 66(3): 411-421.

[6] Ying X. An overview of overfitting and its solutions[C]//Journal of physics: Conference series. IOP Publishing, 2019, 1168: 022022.

[7] Alzubaidi L, Zhang J, Humaidi A J, et al. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions[J]. Journal of big Data, 2021, 8: 1-74.
[8] Nickolls J, Dally W J. The GPU computing era[J]. IEEE micro, 2010, 30(2): 56-69.
[9] Sun Y, Agostini N B, Dong S, et al. Summarizing CPU and GPU design trends with product data[J]. arXiv preprint arXiv:1911.11313, 2019.
[10] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding with unsupervised learning[J]. 2018.
[11] Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners[J]. OpenAI blog, 2019, 1(8): 9.
[12] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[J]. Advances in neural information processing systems, 2020, 33: 1877-1901.
[13] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge[J]. International journal of computer vision, 2015, 115: 211-252.
[14] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014: 740-755.
[15] Everingham M, Eslami S M A, Van Gool L, et al. The pascal visual object classes challenge: A retrospective[J]. International journal of computer vision, 2015, 111: 98-136.
[16] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Proceedings of the AAAI conference on artificial intelligence. 2017, 31(1).
[17] DeVries T, Taylor G W. Improved regularization of convolutional neural networks with cutout[J]. arXiv preprint arXiv:1708.04552, 2017.
[18] Inoue H. Data augmentation by pairing samples for images classification[J]. arXiv preprint arXiv:1801.02929, 2018.
[19] Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning[J]. arXiv preprint arXiv:1712.04621, 2017.
[20] Cubuk E D, Zoph B, Mane D, et al. Autoaugment: Learning augmentation strategies from data[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 113-123.
[21] Lim S, Kim I, Kim T, et al. Fast autoaugment[J]. Advances in Neural Information Processing Systems, 2019, 32.
[22] Zhang H, Cisse M, Dauphin Y N, et al. mixup: Beyond empirical risk minimization[J]. arXiv preprint arXiv:1710.09412, 2017.
[23] Yun S, Han D, Oh S J, et al. Cutmix: Regularization strategy to train strong classifiers with localizable features[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 6023-6032.
[24] Chen P, Liu S, Zhao H, et al. Gridmask data augmentation[J]. arXiv preprint arXiv:2001.04086, 2020.
[25] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
[26] Kingma D P, Welling M. Auto-encoding variational bayes[J]. arXiv preprint arXiv:1312.6114, 2013.
[27] Wang X, Wang K, Lian S. A survey on face data augmentation for the training of deep neural networks[J]. Neural computing and applications, 2020, 32(19): 15503-15531.
[28] Khosla C, Saini B S. Enhancing performance of deep learning models with different data augmentation techniques: A survey[C]//2020 International Conference on Intelligent Engineering and Management (ICIEM). IEEE, 2020: 79-85.
[29] Shorten C, Khoshgoftaar T M. A survey on image data augmentation for deep learning[J]. Journal of big data, 2019, 6(1): 1-48.
[30] Wu, Ren, et al. "Deep image: Scaling up image recognition." arXiv preprint arXiv:1501.02876 7.8 (2015): 4.
[31] Hendrycks D, Mu N, Cubuk E D, et al. Augmix: A simple data processing method to improve robustness and uncertainty[J]. arXiv preprint arXiv:1912.02781, 2019.
[32] Han, J., Fang, P., Li, W., Hong, J., Armin, M.A., Reid, I., Petersson, L. and Li, H., 2022, June. You only cut once: Boosting data augmentation with a single cut. In International Conference on Machine Learning (pp. 8196-8212). PMLR.
[33] Summers, Cecilia, and Michael J. Dinneen. "Improved mixed-example data augmentation." 2019 IEEE winter conference on applications of computer vision (WACV). IEEE, 2019.
[34] Singh K K, Yu H, Sarmasi A, et al. Hide-and-seek: A data augmentation technique for weakly-supervised localization and beyond[J]. arXiv preprint arXiv:1811.02545, 2018.

[35] Zhong Z, Zheng L, Kang G, et al. Random erasing data augmentation[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 13001-13008.

[36] Li P, Li X, Long X. Fencemask: a data augmentation approach for pre-extracted image features[J]. arXiv preprint arXiv:2006.07877, 2020.

[37] Yang, S., Li, J., Zhang, T., Zhao, J. and Shen, F., 2023. AdvMask: A sparse adversarial attack-based data augmentation method for image classification. Pattern Recognition, 144, p.109847.

[38] Harris E, Marcu A, Painter M, et al. Fmix: Enhancing mixed sample data augmentation[J]. arXiv preprint arXiv:2002.12047, 2020.

[39] Verma V, Lamb A, Beckham C, et al. Manifold mixup: Better representations by interpolating hidden states[C]//International conference on machine learning. PMLR, 2019: 6438-6447.

[40] Tran T , Pham T , Carneiro G , et al. A Bayesian Data Augmentation Approach for Learning Deep Models[J]. 2017.

[41] Kurtuluş, E., Li, Z., Dauphin, Y. and Cubuk, E.D., 2023, July. Tied-Augment: Controlling Representation Similarity Improves Data Augmentation. In International Conference on Machine Learning (pp. 17994-18007). PMLR.

[42] Ho D, Liang E, Chen X, et al. Population based augmentation: Efficient learning of augmentation policy schedules[C]//International Conference on Machine Learning. PMLR, 2019: 2731-2741.

[43] Lin, S., Zhang, Z., Li, X. and Chen, Z., 2023, June. SelectAugment: hierarchical deterministic sample selection for data augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 2, pp. 1604-1612).

[44] Cubuk E D, Zoph B, Shlens J, et al. Randaugment: Practical automated data augmentation with a reduced search space[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 702-703.

[45] Müller, S.G. and Hutter, F., 2021. Trivialaugment: Tuning-free yet state-of-the-art data augmentation. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 774-782).

[46] Suzuki, T., 2022. Teachaugment: Data augmentation optimization using teacher knowledge. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10904-10914).

[47] Cheung, T.H. and Yeung, D.Y., 2021, October. Adaaug: Learning class-and instance-adaptive data augmentation policies. In International Conference on Learning Representations.

[48] Gong C, Wang D, Li M, et al. Keepaugment: A simple information-preserving data augmentation approach[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 1055-1064.

[49] Tian K, Lin C, Sun M, et al. Improving auto-augment via augmentation-wise weight sharing[J]. Advances in Neural Information Processing Systems, 2020, 33: 19088-19098.

[50] Lin C, Guo M, Li C, et al. Online hyper-parameter learning for auto-augmentation strategy[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 6579-6588.

[51] DeVries T, Taylor G W. Dataset augmentation in feature space[J]. arXiv preprint arXiv:1702.05538, 2017.

[52] Kuo C W, Ma C Y, Huang J B, et al. Featmatch: Feature-based augmentation for semi-supervised learning[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16. Springer International Publishing, 2020: 479-495.

[53] Li B, Wu F, Lim S N, et al. On feature normalization and data augmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 12383-12392.

[54] Mirza M, Osindero S. Conditional generative adversarial nets[J]. arXiv preprint arXiv:1411.1784, 2014.

[55] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1125-1134.

[56] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.

[57] Choi Y, Choi M, Kim M, et al. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8789-8797.

[58] Choi Y, Uh Y, Yoo J, et al. Stargan v2: Diverse image synthesis for multiple domains[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 8188-8197.

[59] Jiang F, Gu Q, Hao HZ, Li N, Guo YW, Chen DX. Survey on content-based image segmentation methods. Ruan Jian Xue Bao/ Journal of Software, 2017,28(1):160−183 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5136.htm [doi: 10. 13328/j.cnki.jos.005136]

[60] Hao S , Zhou Y , Guo Y . A Brief Survey on Semantic Segmentation with Deep Learning[J]. Neurocomputing, 2020, 406.

[61] LUO Jian-Hao, WU Jian-Xin. A Survey on Fine-grained Image Categorization UsingDeep Convolutional Features. ACTA AUTOMATICA SINICA, 2017, 43(8): 1306-1318. doi: 10.16383/j.aas.2017.c160425

[62] Fu SU,Qin LV,Renze LUO. Review of image classification based on deep learning[J]. Telecommunications Science, 2019, 35(11): 58-74.

[63] Wang W , Yang Y , Wang X , et al. Development of convolutional neural network and its application in image classification: A survey[J]. Optical Engineering, 2019, 58(4):1.

[64] WU Shuai, XU Yong, ZHAO Dongning. Survey of Object Detection Based on Deep Convolutional Network. , 2018, 31(4): 335-346.

[65] Zou Z , Shi Z , Guo Y , et al. Object Detection in 20 Years: A Survey[J]. 2019.

[66] Tian X, Wang L, Ding Q. Review of image semantic segmentation based on deep learning. Ruan Jian Xue Bao/ Journal of Software, 2019,30(2):440−468 (in Chinese). http://www.jos.org.cn/1000-9825/5659.htm

[67] Csurka G, Perronnin F. An efficient approach to semantic segmentation. Int’l Journal of Computer Vision, 2011,95(2):198−212. [doi: 10.1007/s11263-010-0344-8]

[68] Everingham M , Gool L V , Williams C , et al. The Pascal Visual Object Classes (VOC) Challenge[J]. International Journal of Computer Vision, 2010, 88(2):303-338.

[69] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.

[70] Cao Y, Xu J, Lin S, et al. Gcnet: Non-local networks meet squeeze-excitation networks and beyond[C]//Proceedings of the IEEE/CVF international conference on computer vision workshops. 2019: 0-0.

[71] Huang L, Yuan Y, Guo J, et al. Interlaced sparse self-attention for semantic segmentation[J]. arXiv preprint arXiv:1907.12273, 2019.

[72] Wang, Zifu, et al. "Revisiting evaluation metrics for semantic segmentation: Optimization and evaluation of fine-grained intersection over union." Advances in Neural Information Processing Systems 36 (2024).

[73] Yang, Suorong, et al. "Image data augmentation for deep learning: A survey." arXiv preprint arXiv:2204.08610 (2022).

[74] Ouyang W, Zeng X, Wang X, et al. DeepID-Net: Object detection with deformable part based convolutional neural networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(7): 1320-1334.

[75] Diba A, Sharma V, Pazandeh A, et al. Weakly supervised cascaded convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 914-922.

[76] Hu G, Yang Y, Yi D, et al. When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition[C]//Proceedings of the IEEE international conference on computer vision workshops. 2015: 142-150.

[77] Lawrence S, Giles C L, Tsoi A C, et al. Face recognition: A convolutional neural-network approach[J]. IEEE transactions on neural networks, 1997, 8(1): 98-113.

[78] Cao Z, Simon T, Wei S E, et al. Realtime multi-person 2d pose estimation using part affinity fields[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7291-7299.

[79] Toshev A, Szegedy C. Deeppose: Human pose estimation via deep neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1653-1660.

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

[81] 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.

[82] Gastaldi X. Shake-shake regularization[J]. arXiv preprint arXiv:1705.07485, 2017.

Krizhevsky A, Hinton G. Convolutional deep belief networks on cifar-10[J]. Unpublished manuscript, 2010, 40(7): 1-9.

[83] Netzer Y, Wang T, Coates A, et al. Reading digits in natural images with unsupervised feature learning[J]. 2011.

[84] LIANG X, LIN X, QUAN J, et al. Research on the progress of image instance segmentation based on deep learning[J]. Acta Electonica Sinica, 2020, 48(12): 2476.

[85] GAO Wen, ZHU Ming, HE Bai-gen, WU Xiao-tian. Overview of target tracking technology[J]. Chinese Optics, 2014, 7(3): 365-375.
[86] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014: 740-755.
[87] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.
[88] Duan K, Bai S, Xie L, et al. Centernet: Keypoint triplets for object detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 6569-6578.
[89] Liu L, Ouyang W, Wang X, et al. Deep learning for generic object detection: A survey[J]. International journal of computer vision, 2020, 128: 261-318.
[90] Minaee S, Boykov Y Y, Porikli F, et al. Image segmentation using deep learning: A survey[J]. IEEE transactions on pattern analysis and machine intelligence, 2021.
[91] Algan G, Ulusoy I. Image classification with deep learning in the presence of noisy labels: A survey[J]. Knowledge-Based Systems, 2021, 215: 106771.
[92] Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training gans[J]. Advances in neural information processing systems, 2016, 29.
[93] Sun Y, Wong A K C, Kamel M S. Classification of imbalanced data: A review[J]. International journal of pattern recognition and artificial intelligence, 2009, 23(04): 687-719.
[94] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of artificial intelligence research, 2002, 16: 321-357.
[95] Pawara P, Okafor E, Schomaker L, et al. Data augmentation for plant classification[C]//Advanced Concepts for Intelligent Vision Systems: 18th International Conference, ACIVS 2017, Antwerp, Belgium, September 18-21, 2017, Proceedings 18. Springer International Publishing, 2017: 615-626.

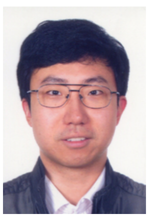
附中文参考文献:

[59] 姜枫,顾庆,郝慧珍,李娜,郭延文,陈道蓄.基于内容的图像分割方法综述.软件学报,2017,28(1):160−183. http://www.jos.org.cn/ 1000-9825/5136.htm [doi: 10.13328/j.cnki.jos.005136]
[61] 罗建豪, 吴建鑫. 基于深度卷积特征的细粒度图像分类研究综述[J]. 自动化学报, 2017, 43(8):13.
[62] 苏赋,吕沁,罗仁泽. 基于深度学习的图像分类研究综述[J]. 电信科学, 2019, 35(11): 58-74.
[64] 吴帅, 徐勇, 赵东宁. 基于深度卷积网络的目标检测综述[J]. 模式识别与人工智能, 2018, 31(4):12.
[66] 田萱王亮丁琪. 基于深度学习的图像语义分割方法综述[J]. 软件学报, 2019, 030(002):440-468.
[84] 梁新宇, 林洗坤, 权冀川, 等. 基于深度学习的图像实例分割技术研究进展[J]. 电子学报, 2020, 48(12): 2476.
[85] 高文, 朱明, 贺柏根, 等. 目标跟踪技术综述[J]. 中国光学, 2014, 7(3): 365-375.

杨锁荣(1996－),男,博士生,主要研究领域为机器学习,数据增强,计算机视觉.

杨洪朝(－),男,硕士生,主要研究领域为深度学习,数据增强,计算机视觉.

申富饶(1973－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为神经网络,机器人智能.



赵健(1979－),男,博士,副教授,IEEE 高级会员,主要研究领域为通信网络,神经计算.