Computer Science and Application 计算机科学与应用, 2021, 11(2), 370-382
Published Online February 2021 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2021.112037

Hans汉斯

基于图像的数据增强方法发展现状综述

冯晓硕[1]，沈　樾[2]，王冬琦[2*]

[1]海军研究院，北京
[2]东北大学软件学院，辽宁 沈阳
Email: 125470246@qq.com, neusheny@163.com, *wangdq@swc.neu.edu.cn

收稿日期：2021年1月17日；录用日期：2021年2月12日；发布日期：2021年2月19日

摘　要

现阶段，基于深度学习的图像处理和识别技术已经发展的十分成熟，但在某些图像识别任务中由于深度学习技术的特点，一些深度神经网络模型层数较多导致的学习能力较强，将图像数据样本中的特征学习的过于充分，使得神经网络模型在训练数据上出现过拟合现象。同时，基于深度学习的图像处理算法训练的模型的好坏与数据集的质量、规模息息相关，但由于客观原因存在获得的图像数据集小、图像质量差，样本分布不均衡等现象。针对上述问题，研究人员提出通过使用图像数据增强技术实现对模型的输入数据的规模、质量和分布情况进行优化，将数据增强后的数据集用于深度学习模型将有效降低出现过拟合现象的概率。本文的主要工作是对现有的图像数据增强技术进行讨论，从传统图像处理方法和基于深度学习数据增强方法两方面进行梳理总结，其中传统图像处理方法有几何变换、颜色变换和像素变换；基于机器学习的图像数据增强方法有自动数据增强方法、基于生成对抗网络数据增强方法和基于自动编码器和生成对抗网络组合的数据增强方法。本文着重对图像融合、信息删除以及基于生成对抗网络的图像数据增强方法等技术进行介绍，并且对文中提出的数据增强方法的思想及其优缺点进行讨论，为研究人员在不同图像任务中利用对应的数据增强方法来优化数据集从而提高模型准确率提供研究思路。

关键词

数据增强，图像数据集，图像处理，深度学习

A Survey on the Development of Image Data Augmentation

Xiaoshuo Feng[1], Yue Shen[2], Dongqi Wang[2*]

[1]Naval Research Institute, Beijing
[2]Software College, Northeastern University, Shenyang Liaoning
Email: 125470246@qq.com, neusheny@163.com, *wangdq@swc.neu.edu.cn

*通讯作者。

文章引用：冯晓硕，沈樾，王冬琦．基于图像的数据增强方法发展现状综述[J]．计算机科学与应用，2021，11(2): 370-382. DOI: 10.12677/csa.2021.112037

Received: Jan. 17th, 2021; accepted: Feb. 12th, 2021; published: Feb. 19th, 2021

Abstract

Image processing and recognition technology based on deep learning has developed very well. However, in some image recognition tasks, due to the characteristics of deep learning models, some deep neural network models have strong learning ability due to the large number of layers, and the features in the images are learned too fully, which makes the neural network model appear fitting phenomenon on the training data. At the same time, the quality of the model trained by the image processing algorithm based on deep learning is closely related to the quality and scale of the dataset. However, due to the small dataset, poor image quality and unbalanced sample distribution, etc. In order to solve the above problems, the researchers proposed to optimize the scale, quality and distribution of the input data of the model by using image data augmentation technology. Applying the augmented dataset to the deep learning model will effectively reduce the probability of over-fitting. The main contribution of this paper is to discuss the existing image data augmentation technology, and summarize the traditional image processing methods and data augmentation methods based on deep learning, among which the traditional image processing methods include geometric, color, and pixel transformation. Image data augmentation methods based on machine learning include Auto Augment, methods based on GAN and methods based on combination of AE and GAN. In this paper, the technologies of image fusion, information deletion and image data augmentation method based on GAN are introduced, and the ideas, advantages and disadvantages of the data augmentation methods proposed in this paper are discussed, which provides ideas for researchers to optimize datasets and improve the accuracy of models by using corresponding data augmentation methods in different image tasks.

Keywords

Data Augmentation, Image Dataset, Image Processing, Deep Learning

1. 引言

随着人工智能技术的不断发展，基于深度学习的图像处理技术愈发成熟，应用也越来越广泛。由数据集驱动的人工智能技术训练的模型往往需要巨大规模的数据集，但由于时间成本和金钱成本的限制，可能常会遇到图像数量少、图像质量差和类别不均衡[1]的情形，这给图像识别任务带来种种困难，数据集中图像的质量和数量极大的影响了深度学习模型的泛化能力。由于深度学习网络模型有着极强的学习能力，模型可能经常会将图像数据集上数据的一般特征提取出来作为预测某一类结果的特性，这往往会导致模型在训练集上会预测出很好的结果，而在测试集以及验证集上的有较高的错误率，模型的泛化能力低[2]。

基于图像的数据增强能够增加训练样本的多样性，如通过翻转、添加噪声等基础图像处理操作或根据现有数据生成新的样本进行数据集扩充、数据质量的增强。使用数据增强方法后的数据集训练模型，以达到提升模型的稳健性、泛化能力的效果。

本文主要研究了基于图像的数据增强技术，并对其进行归类整理，着重介绍各类技术的特点及其解

决的问题，对其存在的不足进行分析。对图像数据增强技术待解决问题进行总结，为相关研究人员提供详尽的技术发展状况报告。

2. 图像数据增强概述

2.1. 数据增强

数据增强，也叫数据增广。数据增强方法的本质实际上是在现有的有限数据的基础上，在不实际收集更多数据的前提下，而让数据产生等价于更大数据量的价值，即根据现有数据样本按照规则生成增量数据的过程。数据增强方法不仅是数据样本量的增多，更多的是数据本身特征的“增强”。样本数据是整体数据的抽样，当样本数据量足够大时，样本的分布情况和总体的分布情况应相似。但由于客观原因收集的样本数据不够完整，这时则可通过数据增强方法生成与真实数据分布更加相似的新样本的数据；深度学习网络模型拥有极强的学习能力，因此学习到的一些无用的信息特征对最终的结果会产生负面影响，而数据增强技术可实现按照需求针对数据施加约束来增加先验知识的前置过程，如将一些信息删除或补全，来减少负面影响对处理图像任务的模型性能的影响。

现阶段数据增强方法的使用方式主要被分为两种：离线增强和在线增强。离线增强是指对数据集执行一次性转换，该操作可成倍增加数据样本的数量。使用数据增强方法产生的样本数量为增强因子数与原始数据样本量的乘积。离线增强由于一次性处理全部数据集，因此适用于较小的数据集。在线增强使在获取批量的数据后就对其进行数据增强操作，随后增强后的数据就被送入机器学习模型进行训练，由于其批量处理的特性，因此一般适用于大数据集。

2.2. 基于图像的数据增强的分类

数据增强方法主要有作用于图像的和作用于文本的两类，本文主要介绍基于图像的数据增强方法[3]，根据是否使用机器学习技术，其被分为两个部分进行讨论：基于传统图像处理技术的图像数据增强和基于机器学习的图像数据增强技术，其中基于传统图像处理技术的图像数据增强中将介绍针对图像数据本身的几何变换、色彩变换和像素变换。基于机器学习的图像数据增强技术将介绍自动数据增强技术、基于生成对抗网络数据增强技术和基于自动编码器和生成对抗网络组合的数据增强方法，这些方法都是使用机器学习相关理论实现的图像生成、图像转换模型。本文对于现有的图像数据增强方法分类如图 1。

3. 传统的图像数据增强方法

传统的图像数据增强方法，通常使用图像处理技术[4]来完成数据集的扩充和图像质量优化，大致分为几何变换、色彩变换、像素变换三大类。

3.1. 几何变换

针对数据集进行空间几何变换，常常会存在改变图像原始的标签信息或者增加一些不相关数据的情况，这称之为不安全的转换。例如对文字的识别任务中，对图像进行翻转操作是没有意义的。但对于存在位置偏差的数据集，用几何变换技术可以很好解决问题。但在真实情况下，训练集与测试集的数据的差异十分复杂，除了移位旋转等操作外，还包括其他复杂变换。因此几何变换的应用范围相对有限。

3.1.1. 图像翻转与旋转

图像翻转操作包括对图片进行垂直和水平翻转，其中垂直翻转实现需要水平翻转后再图像进行 180˚旋转处理，水平翻转比垂直翻转应用更为广泛。这种技术的优点是易于实现，此外图像翻转在 CIFAR-10 数据集上具有较好的效果，但对文本识别的数据集，如 MNIST，使用图像翻转会更改其标签信息[5]。

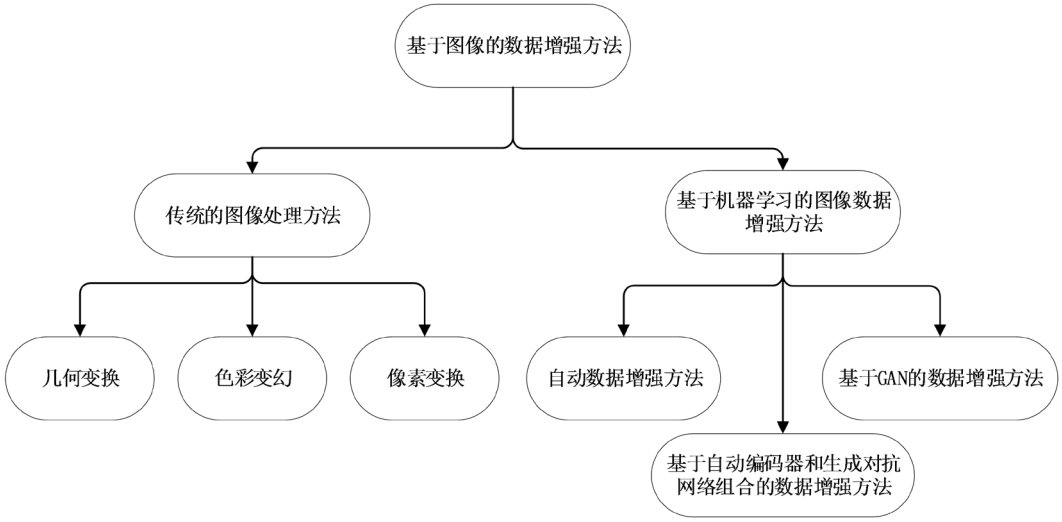

Figure 1. Classification of image data augmentation methods
图 1. 基于图像的数据增强方法分类

进行图像旋转后得到的图像与原始图像的维数是否相同取决于图像的旋转度数以及原始图像的形状。当长方形的图像旋转 180˚或者正方形的图像旋转 90˚、180˚、270˚时，旋转后的图像与原始图像能够保持一致的维数。与图像翻转操作一样，在特定的数据集上，例如文本识别数据集 MNIST，其图像变换的安全性取决于图像的旋转度数，随旋转度数的增加，转换后的图片不在保留原标签信息。

3.1.2. 图像剪裁与缩放

图像的随机剪裁可视为从原始图像上进行随机抽样，再将抽样获得的图像数据样本恢复为原始图像大小。

图像缩放分为向外缩放和向内缩放。图像缩放与图像剪裁不同的是向外缩放会得到比原始图像更大尺寸的图像，再从中剪裁出与原始图像大小相同的图像，向内缩放则缩小原始图像的大小，并针对其超出边界的部分进行填充操作从而获得与原始图像尺寸一致的图像。

3.1.3. 图像移位与边缘填充

图像移位是指不改变图像的尺寸而将图像在以坐标轴为移动参考进行横轴和纵轴的移动，并针对边缘部分进行填充处理。在进行了图像移位操作后，大部分的图像数据中对于图像任务有用的部分将位于图像的边缘部分，因此深度学习模型在进行计算机视觉任务训练时会将关注焦点转移到任意位置，而不是仅仅针对图像中心区域的学习，这种操作能够有效的提高模型的鲁棒性。

在对图像数据样本进行旋转、移位、缩放等操作后，需要将变换后的图像恢复到与原始图像尺寸一致的大小，恢复的过程通过对于图像的边缘部分进行填充操作实现。常用的图像填充方法包括：常数填充，使用常数值对图像的边缘部分进行填充，这种填充方式适用于单色背景的图像数；边界值填充，在原始图像边界的外部填充原始图像的边界的像素值，此方法适用于短距离移位。

3.2. 色彩变换

3.2.1. 色彩空间

数字图像数据通过使用长、宽和通道来表示数据。

常用的色彩空间包括：

1) 通过 RGB 通道的变化和叠加得到不同颜色的 RGB 色彩空间；

2) YUV 色彩空间，其中 Y 表示亮度，UV 表示色度；

3) HSV 颜色模型，H 表示色调，S 表示饱和度，V 表示明暗程度。

此外还包括 I1I2I3，L*a*b* [6]和 YcbCr [7]，在这些颜色空间中，HSV 颜色空间是直观的，其组成部分可以很容易地与物理世界相关联[8]。

在色彩通道上进行图像亮度调节以及色度调剂是数据增强的一种有效方式。通常情况下，采集到的图像数据的亮度覆盖范围不足，为达到深度学习对亮度鲁棒性的基本要求，进行亮度转换操作成为基于色彩空间的数据增强技术中最常用的一种方法。在图像数据中，亮度偏暗的图像，亮度方差也更小，从而整体的亮度范围被压缩。Gamma 变换通过非线性变换将过亮或过暗的图片进行调整。直方图均衡化是更加高级的色彩空间增强方式，对对比度相近的图像使用该技术可增强局部的对比度而不影响整体的对比度，这种方式对过亮或过暗的图像数据能够实现有效的数据增强。

3.2.2. 色彩空间转换

色彩空间转换是色彩特征提取的一种非常有效的方式。不同的色彩空间表示形式虽各有特性，但由于其同构性，可以互相转换。图像通常位于三维 RGB 颜色空间中，但 RGB 颜色空间在感知上不均匀，颜色的接近度并不表示颜色相似性。色彩空间转换通过将图像在 RGB、HSV、LAB 等不同的颜色空间上转换，使以不同的方式对每个分量进行加权，对于不同的数据集，通常需要选择合适的颜色空间转换来提高模型的性能。

色彩空间转换的缺点除了会消耗大量内存空间和时间，其也会产生不好的效果，例如人脸识别需要的关键信息使黑白黄，但若出现大量红绿等颜色信息，则是不合理的。此外，颜色空间转换的图像增强效果是有限的，虽然比几何变换更具多样性，但不恰当的使用可能会使模型发生欠拟合。

Ze Lu 等人[9]提出一种用于面部识别任务的色彩空间框架，提出色彩空间 LuC1C2 其通过比较 RGB 系数的颜色传感器属性选择 Lu 亮度分量，通过 RGB 颜色空间的色度子空间和协方差分析来确定 C1C2 颜色分量的变换向量的方向。在 AR、Georgia Tech、FRGC 和 LFW 人脸图像数据库上实验，确定了色彩空间 LuC1C2 具有更好的人脸识别性能。并且通过将 LFW 和 FRGC 数据库上提取的 LuC1C2 颜色空间中的 CNN 特征与简单的原始像素特征相结合，显著提高面部验证性能。

3.3. 像素变换

3.3.1. 噪声

图像噪声是指在原始图像上随机叠加一些孤立的能够引起较强的视觉效果像素点或像素块，以扰乱图像的可观测信息，使其能够更好的提高卷积神经网络模型的泛化能力。常见的噪声有：椒盐噪声、高斯噪声、Coarse Dropout、Simplex Noise Alpha、Frequency Noise Alpha。它们都是以不同的方式生成以不同数值填充的不同大小像素遮掩点，再与原图混合，以扰乱原始图像的一些特征。

3.3.2. 模糊

模糊的本质上可视为对原始图像进行卷积操作，常用的方法是高斯模糊，该方法服从的卷积核矩阵服从二维正态分布，以减少各像素点值的差异从而降低细节层次，使图像数据的像素平滑化，达到模糊图片的效果。模糊半径越大，图像就越模糊。

3.3.3. 图像融合

图像融合技术，通过求两张图像的像素值的均值将两张图片混合在一起，或者是随机裁剪图像并将裁剪后的图像拼接在一起形成新图像。当混合来自整个训练集的图像而不是仅来自同一类别的实例的图

像时，可以获得更好的结果。图像融合方法从人的视角看毫无意义，但从实验的角度上观察，确实能够提升精度。

1) SMOTE [10]

采集到的数据集常存在的问题是样本类别不平衡问题，样本类别之间的较大差距会影响分类器的分类性能。SMOTE 方法提出以小样本类别合成新的样本来解决样本不平衡问题，该方法将提取的图像特征映射到特征空间，确定好采样倍率后，选取几个最相邻的样本，从中随机选取一个连线，并在连线上随机选取一点作为新样本点，重复至样本均衡。杜金华[11]在研究中提出使用基于 SMOTE 算法的上采样法分别对原始图像数据集进行增强，实验表明花岗石识别准确率有所提高。

2) MIXUP [12]

ERM 方法会在各个类间形成明确的决策边界，而 Mixup 方法是一种基于线性过渡的数据增强的方法，使用 mixup 能够使得数据样本之间像素点是渐变的，使样本分类边界模糊化，使得非 0 即 1 的预测变为较为平滑的预测效果，抑制模型在进行预测分类时的不稳定性，增强模型的泛化能力。这种方法从训练数据中随机抽取两条数据将抽取到的图像数据的像素值进行符合 Beat 分布的融合比例的线性加权求和，同时将样本对应的 One-hot 向量标签也对应加权求和，预测生成的新样本与加权求和后的标签的损失，进行反向求导并更新参数，同时抽取批量数据并进行随机打散后进行加权求和。在 CIFAR、 ImageNet 图像分类数据集语音数据集中使用该方法能够实现模型性能的提升，并且降低模型对不完整标签的记忆。mixup 方法尽管再在程上取得很好地效果，但缺乏理论支撑，且该方法需要较长的时间才能收敛出较好的结果。ERM 方法与 Mixup 方法的对比如图 2 所示。

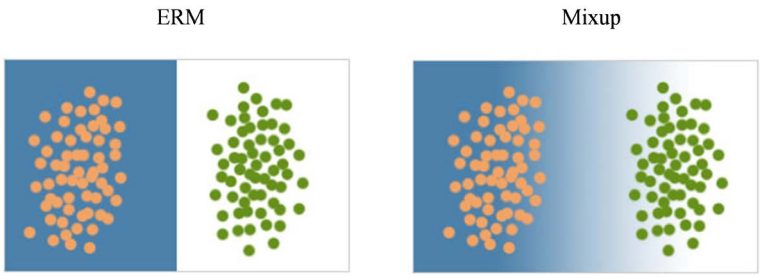

Figure 2. Comparison of ERM and Mixup
图 2. ERM 与 Mixup 对比

3) CUTMIX [13]

CutMix 是一种改进的随机擦除策略，随机擦除用一块矩形掩码覆盖原始图像，实现擦除图像上的一部分像素信息，但其缺点是减少了训练图像上的包含信息的像素比例，并且需要大量的计算，较为耗时。CutMix 方法将随机选中的区域填充其他图像的补丁区域。这种方法与 Mixup 方法相比，改进了混合图像后，人眼无法主观辨别图像标签的情形。在擦除区域添加其他样本信息，进一步增强模型定位能力。CutMix 在 CIFAR 和 ImageNet 分类任务以及对 ImageNet 上的弱监督的本地化任务领先于最新的数据增强策略，同时，在 Pascal 检测和 MS-COCO 图像字幕基准测试中获得了性能提升。这种方法改进了针对输入损坏及其模型失配检测性能的模型鲁棒性。

4) Sample Pairing [14]

Sample Pairing 方法常用于图像分类任务中的数据增强，该方法首先从训练集中随机选择两张图片，与 Mixup 方法不同的是，随机选择图像的方式是从训练集中随机抽取两张图片并分别进行基础数据增强操作(如随机翻转等)处理后，再经像素取平均值，最后叠加合成一个新的样本。而标签为

原样本标签中的一种，理论上新样本数量平方增加。这种方法能够显著提高所有测试数据集分类准确性。使用 GoogleNet 的 ILSVRC 2012 数据集的 top-1 错误率从 33.5%降低到 29.0%，而在 CIFAR-10 数据集中则从 8.22%降低到 6.93%。当训练集中的样本数量非常小时，SamplePairing 技术大大提高了模型预测的准确性。因此该技术对于训练数据量有限的任务(例如医学成像任务)更有价值。SamplePairing 方法实现简单，效率大大增加，但缺少相应的理论支撑。

3.3.4. 信息删除

1) 随机擦除[15]

随机擦除方法与添加噪声方法相似，通过随机选取图像中的矩形区域，并使用随机像素值对其遮盖。该技术可以很容易嵌入大部分卷积神经网络模型中。随机擦除的好处在于迫使模型去学习有关图像的更多描述性特征，从而防止过拟合某个特定视觉特征，确保网络关注整个图像，而不只是其中的一部分。随机擦除的缺点是不一定会保留标签(例如文本 8->6)。Zhun Zhongy 等人将随机擦除方法用于图像分类、物体检测和人员重新识别任务，并通过该方法实现了性能的提升。使用随机擦除方法进行大量实验，在 CIFAR、PASCAL VOC 2007、Fast-RCNN、re-ID、Market-1501、DukeMTMC-reID 上表现出良好的效果。

2) CUTOUT [16]

与随机擦除相似，Cutout 是在图像上的随机位置使用一定大小的正方形 path 进行 0-mask 剪裁。蒋芸等人提出了激活区域处理算法(Activation Region processing algorithm) [17]并将其嵌入到 CNN 模型，对网络卷积层的特征图进行遮盖处理，进一步提高模型的性能，降低过拟合的风险。该方法从卷积神经网络中提取出较为关键的局部特征的卷积层的素值最大的特征图，对其上采样后将像素值大于整个图像素均值的像素点使用[0,1]的随机噪声进行遮盖处理，输入到下层网络继续训练。算法在 CIFAR 和 Fashion-MNIST 数据集上得到更低的错误率。在不同的网络结构 RestNet-18、WRN-28-10、ResNext-8-64 使用 AR 算法后，与未加任何遮挡的 CNN 模型相比，得到更低的错误率。随机擦除、Cutout 和 GridMask 方法的图像增强效果如图 3 所示。

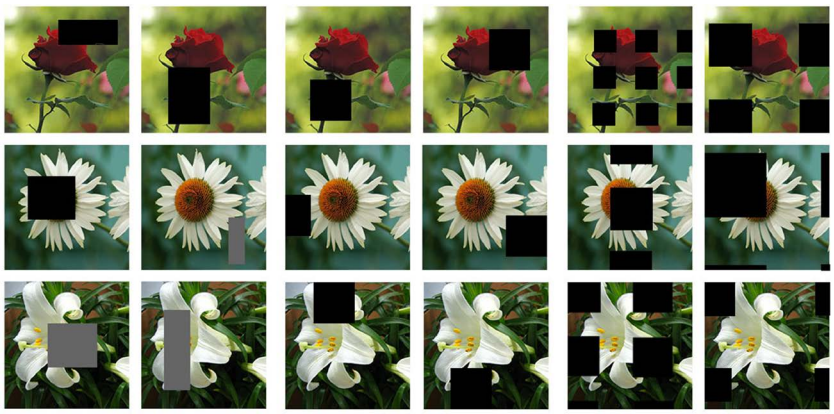

Figure 3. Comparison of Rerasure, Cutout and GridMask
图 3. 随机擦除、Cutout 和 GridMask 对比

3) GRIDMASK [18]

Pengguang 等人提出的 GridMask 的信息删除方法平衡了删除区域与保留区域的面积，其本质是对图像进行网格覆盖，优化了 cutout 和随机擦除方法过度删除问题，并保持图像区域连续，易于实现且快速灵活，与以前的方法相比，GridMask 在各种数据集上得到更优的效果，优于所有以前的无监督策略，包

括 AutoAugment [19]提出的最佳组合策略。该方法可以用作数据扩增的新基准策略。

4. 基于深度学习的数据增强方法

在进行机器学习模型训练的过程中，优化模型的目标就是尽可能的实现模型的损失降低，因此为了完成这一优化目标，往往需要大量的训练数据作为支撑。传统的数据增强技术依靠对现有数据集的微小调整，包括翻转、旋转和平移等调整方法，通过这些方法对数据集的调整，会产生大量具有微小的或者巨大差别的数据集，使用这种数据集的实验方法将会把这些调整后的数据集视为与原始数据集不同的数据，从而进行模型的训练。数据增强的作用除了能够增加训练的样本数量之外，提高模型的泛化能力之外，还可以通过增加噪声数据，从而提高模型的鲁棒性。

除了传统的数据增强技术以外，近年来，随着机器学习的快速发展和广泛应用，研究人员开始将机器学习技术用于数据增强领域的研究，并取得了一定科研成果。

4.1. 自动数据增强

从数据自身的特点出发，搜索适合不同特点数据集的数据增强策略能够从体系结构搜索的角度重新定义了一种数据增强的新模式。

谷歌大脑的研究人员提出了一种自动搜索合适的数据增强策略的方法(AutoAugment) [19]，通过设计这种不改变深度学习的网络架构的数据增强方法来实现具有更多不变性的数据增强策略，这种思想能够避免对神经网络架构进行修改而从策略搜索的角度对模型的训练过程进行性能上的优化。该方法通过创建一个搜索空间用来保存数据增强策略，并针对不同的批量任务根据搜索算法从搜索空间中选择合适的子策略，选择的子策略能够应用特定的图像处理函数进行数据增强的操作，以使这样训练出的神经网络能获得最佳的验证准确率。该算法的性能接近不使用任何无标注样本的半监督学习方法。此外，该算法能够实现策略的迁移，将学习到的策略应用到其他类似的数据增强任务上，能够得到较高的准确率，并且不需要在额外的数据上对预训练的权重进行调整。该算法中使用强化学习作为搜索算法，并提出在搜索算法的方面能够进一步研究，得到更好的实验性能。但这种方法在简化设置的情况下需要较长的训练时间。

针对计算损耗巨大的问题，谷歌大脑的研究人员又提出了一种自动数据增强的方法，称为 RandAugmentation [20]。这种方法大大缩小了数据增强所产生的样本空间，从而将数据增强的过程与深度学习模型的训练过程集成起来，而不是将数据增强作为独立的任务。该论文同时也证明了自动的选择数据增强策略的方案通常是在规模较小的数据集上训练参数量级较低的模型而实现的自动数据增强，在此基础上再将搜索到的数据增强策略应用到大规模数据集上的方法不是最优的[20]。

自动数据增强是否或者何时需要作为一个单独的搜索阶段一直是困扰着研究人员，在该方面的突破也许能够从根本上解决自动数据增强和模型的训练过程之间的关系问题。此外

Yonggang Li 等人[21]在 2020 年提出了一种新的数据增强技术，该论文提出了将可微分网络架构搜索算法应用在数据增广策略搜索任务上，该算法同样针对 AutoAugment 中的昂贵计算导致 AutoAugment 方法在适用性上表现较差的问题。DADA 算法提出通过 Gumbel-Softmax 将离散的数据增强策略选择转化为到一个可优化的问题。

AutoAugment 作为开创性的工作，提出了自动搜索策略用于数据增强，将策略的选择过程视作一个组合优化问题。但由于其需要消耗巨大的计算时间，导致其适用性较低，因此研究人员开始针对计算耗时问题提出不同的解决方案。除了上述的两种针对 Autoaugment 的改进方法之外，还有 Population Based Augmentation [22]和 Fast AutoAugment [23]等方法。

4.2. 基于生成对抗网络的数据增强方法

通过基于生成对抗网络的生成建模的方式进行数据增强是现阶段较为常用的手段。生成对抗网络应用在数据增强任务上的思想主要是其通过生成新的训练数据来扩充模型的训练样本，通过样本空间的扩充实现图像分类任务效果的提升。研究人员在原始生成对抗网络框架的基础上又提出了多种不同的改进方案，通过设计不同的神经网络架构和损失函数等手段不断提升生成对抗网络的变体的性能。

4.2.1. DCGAN

DCGAN [24]尝试将图像领域应用广泛的 CNN 与生成对抗网络 GAN 结合起来，提出了 Deep Convolutional GANs (DCGAN)，在图像分类任务上证明了其优于其他无监督算法。该算法的核心部分是对 CNN 架构进行了三处修改：(1) 使用卷积层替代了池化层。作者在 GAN 中的生成器中进行了此类修改，使得生成器能够学习其自身空间的下采样方式，而不是参数指定的下采样方式。(2) 消除了卷积特征上的全连接层。作者尝试将最高卷积特征分别直接连接到生成器和判别器的输入和输出。(3) 批量归一化[25]。使用批量标准化通过将输入标准化以使零均值和单位方差为零来稳定学习，并且能够有效解决深度生成器的所有样本坍塌到单点的问题。将该方法用到生成器的输出层和判别器的输入层会导致批量归一化模型不稳定问题，因此作者在剩余的所有层上都使用了批量归一化的操作。DCGAN 算法实现了 CNN 和 GAN 的结合，是一种有效的图像生成模型，被广泛的用于数据集样本的生成。但使用该方法中，当训练模型的时间较长时，仍然在部分模型中存在不稳定的问题。

4.2.2. CycleGAN

CycleGAN [26]作为图像转换领域的重要模型，可以实现样本数据无需配对即可进行转换，例如将一个名人转换成一个卡通人物，这种图像转换的使用能够对样本数据进行极大的扩充而保留原始图像的轮廓。CycleGAN 作为一种不对齐数据的图像转换方法现在被广泛的用于图像到图像的转换。

CycleGAN 实际上是由两个对称的生成对抗网络组成的环形网络，将该模型与 DCGAN 进行比较后发现，该模型能够控制图像生成，而 DCGAN 模型则输入一个噪声后输出一张无法控制的图片。CycleGAN 的结构如图 4 所示。

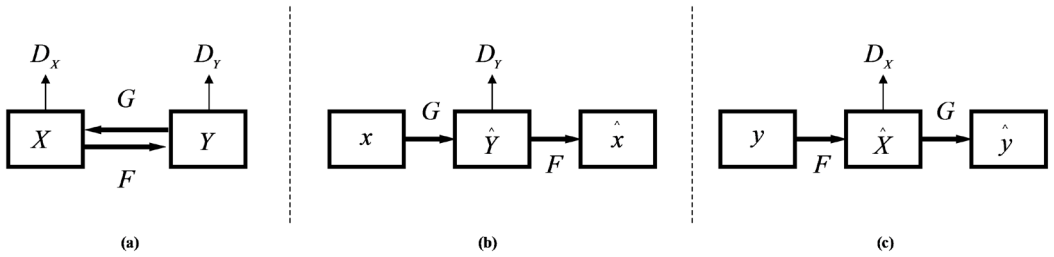

Figure 4. The structure of CycleGan: (a) This model contains two mapping functions $G: X \to Y$ and $F: Y \to X$, and associated adversarial discriminators D_Y and D_X. D_Y encourages G to translate X into outputs indistinguishable from domain Y, and vice versa for D_X, F, and X. To further regularize the mappings, we introduce two "cycle consistency losses" that capture the intuition that if we translate from one domain to the other and back again we should arrive where we started: (b) forward cycle-consistency loss: $x \to G(x) \to F(G(x)) \approx x$, and (c) backward cycle-consistency loss: $y \to F(y) \to G(F(y)) \approx y$ [26]

图 4. CycleGAN 结构：(a) 该模型由两个映射函数组成 $G: X \to Y$ 和 $F: Y \to X$，并包括两个对抗判别器 D_Y 和 D_X，D_Y 鼓励 G 将 X 转换为与 Y 无法区分的输出，D_X 则鼓励 F 将 Y 转换为与 X 无法区分的输出。为了进一步的将映射规范化，该模型定义了两个"循环一致性损失"，这两个损失函数保证了将一个域转换为另一个域并再次转换回来的时候，与原始的域尽可能保持一致。(b) 前向循环一致性损失：$x \to G(x) \to F(G(x)) \approx x$，(c) 反向循环一致性损失 $y \to F(y) \to G(F(y)) \approx y$ [26]

4.2.3. Conditional GANs

2014 年，Mehdi Mirza 等人提出了 Conditional GAN [27]，论文中提出的模型不仅仅需要较高的逼真度而且需要在一定的条件约束下完成，由于其增加了条件约束，因此生成器和判别器的设计会发生较大的改变。通过根据附加信息对模型框架进行调整，可以用于指导数据的生成过程，这种根据条件生成数据的方式对于数据增强非常有效，研究人员在原始图像上可以根据不同的需求条件生成增量数据，并将增量数据应用到下游的神经网络模型中。Conditional GAN 的结构如图 5 所示。

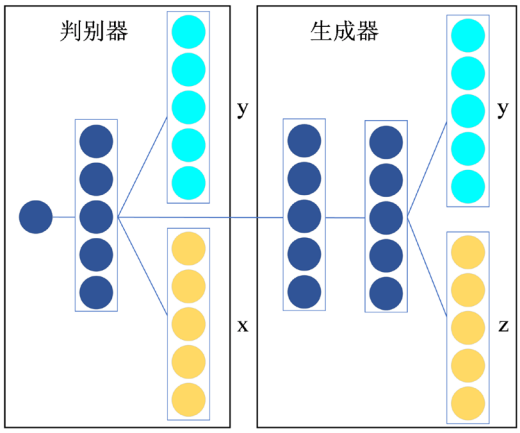

Figure 5. The structure of Conditional GAN
图 5. Conditional GAN 的结构

尽管生成对抗网络在生成图像领域被广泛的应用，但其训练的不稳定性，以及要求大量训练数据的不适用性将导致其不同的变体方法在一些时候并不能有效的实现数据增强的任务。

4.3. 基于自动编码器和生成对抗网络组合的数据增强方法

自动编码器通过将其网络结构的一般用于编码，获得图像的低维向量表示，将网络结构的另一半用于解码，获得根据低维向量表示重新构造的图像数据，这种编码解码的方式能够实现训练数据样本和噪声数据样本的容量扩充，实现了利用数据增强技术提高神经网络的泛化能力和鲁棒性。

在生成对抗网络被广泛应用到生成数据任务之后，研究人员开始提出生成对抗网络与自动编码器的结合形式，通过将变体的生成对抗网络和变体的自动编码器结合而形成一个通用的学习框架来生成细粒度类别的图像，这种新颖的图像生成方式同样能够有效的完成数据增强任务。Jianmin Bao 等人[28]提出的 CVAE-GAN 通过将图像建模为概率模型中图像标签和隐藏属性的组合的形式。

CVAE-GAN 首先使用编码器将给定的训练图像数据和类别标签编码为符合给定概率分布的隐变量，再通过生成器将从隐变量中采样得到的数据和对应的类别标签生成图像数据，将该图像数据输入到分类器和判别器中从而输入分类标签和判别标签，生成器和判别器构成了一个生成对抗网络，其中生成器尝试通过已经学会了区分真实样本和虚假样本的判别器提供的的梯度来学习真实数据分布。作者还在其论文中证明了均值特征匹配也可以用于条件图像生成任务中。CVAE-GAN 方法能够在多种图像任务中取得较好的性能，包括图像生成任务、图像修复任务等，但在生成位置类别的样本方面还存在一定的可探索性。

在自动编码器和生成对抗网络的组合应用方面，Yang He 等人[29]提出了一种新颖的图像生成方法，该方法可以被分类为一种随机回归方法，其学会了从单个条件输入中产生多个不同的示例。这种方法与 CVAE-GAN 方法一样结合了生成对抗网络和自动编码器的优势来完成图像生成任务。这种新提出的算法

通过使用随机回归公式为条件图像生成任务提供了一种新颖的解决方案，该模型可以生成准确且多样的样本，并且可以稳定地训练并提供具有潜在向量表示形式的抽样机制。该模型还应用了通道级别的dropout (channel-wise dropout)，从改进网络结构的角度提出解决多项选择学习思想的方法。

将图像生成技术用于数据增强任务的方法除了生成对抗网络以及自动编码器和生成对抗网络的组合形式之外，研究人员还提出了一些其他的方法。Qifeng Chen 等人[30]提出级联优化网络(cascaded refinement networks， CRN)，该模型将图像生成任务转化为回归问题，该模型还证明了可以通过合适结构的前馈网络合成图像数据，实现了将图片无缝缩放到高分辨率，并在实验中证明了模型的有效性。Justin Johnson 等人[31]提出采用感知损失函数训练前馈网络进行图像转换的任务。模型通过训练一个用于图像转换任务的前馈网络，同时不需要进行像素级别的求差值操作而构造损失函数。使用感知损失函数，从预训练好的网络中提取高级特征。该模型同样在图像转换任务中取得了不错的性能。

5. 总结

针对图像数据集的数据增强技术可以分为两种类型：对数据集进行变换操作达到扩充数据集的目的；对数据集进行过采样或欠采样达到样本与真实分布相似效果。

随着深度学习技术的不断发展，应用于计算机视觉领域的深度模型也在不断的被提出。基于深度学习的数据增强技术主要从数据扩充的角度对模型进行性能上的提升，而不是改变深度模型的网络结构。现阶段的传统图像数据增强和基于深度学习的数据增强技术都在不断的发展和提出，将数据增强技术用于计算机视觉任务也正在成为学术研究的主流做法。传统图像处理方法有几何变换、颜色变换和像素变换等，而基于深度学习的图像数据增强技术主要包括：自动数据增强通过设计一种不改变深度网络架构的数据增强方法来实现具有更多不变性的数据增强策略，通过创建一个搜索空间用来保存数据增强策略，并针对不同的任务根据搜索算法的运行进行适当子策略(例如剪裁、翻转)的选择，从而实现自动数据增强的目的；而基于生成对抗网络的数据增强主要是基于生成对抗网络的机制进行生成器和判别器的设计，以及生成对抗网络的算法框架的设计；而基于自动编码器和生成对抗网络组合形式的数据增强方法则是通过编码器、解码器、生成器和判别器的设计实现数据增强任务。本文总结了传统图像处理方法和基于深度学习数据增强方法两方面技术，讨论了不同图像数据增强技术的优缺点。

随着深度学习技术的不断革新，更多的深度模型将会被提出，而针对数据集优化的数据增强技术也会随之发展，未来研究人员将可能更多的通过改进基于深度学习的数据增强技术来适应模型算法的结构，实现模型试验效果的提升。

参考文献

[1] 王和勇, 樊泓坤, 姚正安, 李成安. 不平衡数据集的分类方法研究[J]. 计算机应用研究, 2008, 25(5): 1301-1303+1308. http://dx.chinadoi.cn/10.3969/j.issn.1001-3695.2008.05.006

[2] 张奇, 卢建斌, 刘涛, 刘齐悦. 基于 CNN 的舰船高分辨距离像目标识别[J]. 雷达科学与技术, 2020, 18(1): 27-33. http://dx.chinadoi.cn/10.3969/j.issn.1672-2337.2020.01.005

[3] Perez, L. and Wang, J. (2017) The Effectiveness of Data Augmentation in Image Classification using Deep Learning. arXiv: 1712.04621. https://arxiv.org/abs/1712.04621

[4] 朱虹. 数字图像处理基础[M]. 北京: 科学出版社, 2005.

[5] Shorten, C. and Khoshgoftaar, T.M. (2019) A Survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, **6**, Article No. 60. https://doi.org/10.1186/s40537-019-0197-0

[6] Nawara, P., Jakubowski, T. and Sobol, Z. (2019) Application of the CIE L*a*b* Method for the Evaluation of the Colour of Fried Products from Potato Tubers Exposed to C Band Ultraviolet Light. *E3S Web of Conferences*, **132**, Article No. 02004. https://doi.org/10.1051/e3sconf/201913202004

[7] Keyvanpour, M. and Merrikh-Bayat, F. (2011) An Effective Chaos-Based Image Watermarking Scheme Using Fractal

coding. *Procedia Computer Science*, **3**, 89-95. https://doi.org/10.1016/j.procs.2010.12.016

[8] 霍宏涛. 数字图像处理[M]. 北京: 机械工业出版社, 2003.

[9] Lu, Z., Jiang, X. and Kot, A. (2017) Enhance Deep Learning Performance in Face Recognition. *Proceedings of* 2017 2*nd International Conference on Image*, *Vision and Computing*, Chengdu, 2-4 June 2017, 244-248. https://doi.org/10.1109/ICIVC.2017.7984554

[10] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2002) SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*, **16**, 321-357. https://doi.org/10.1613/jair.953

[11] 杜金华. 基于颜色特征和逻辑回归的饰面花岗石图像识别技术研究[D]: [硕士学位论文]. 泉州: 华侨大学, 2018.

[12] Zhang, H., Cisse, M., Dauphin, Y. and Lopez-Paz, D. (2017) Mixup: Beyond Empirical Risk Minimization. arXiv: 1710.09412. https://arxiv.org/abs/1710.09412

[13] Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J. and Yoo, Y. (2019) CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. *Proceedings of International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 6022-6031. https://doi.org/10.1109/ICCV.2019.00612

[14] Inoue, H. (2018) Data Augmentation by Pairing Samples for Images Classification. arXiv: 1801.02929. https://arxiv.org/abs/1801.02929

[15] Zhong, Z., Zheng, L., Kang, G., Li, S. and Yang, Y. (2017) Random Erasing Data Augmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 13001-13008. https://doi.org/10.1609/aaai.v34i07.7000

[16] Devries, T. and Taylor, G.W. (2017) Improved Regularization of Convolutional Neural Networks with Cutout. arXiv: 1708.04552. https://arxiv.org/abs/1708.04552

[17] 蒋芸, 张海, 陈莉, 陶生鑫. 基于卷积神经网络的图像数据增强算法[J]. 计算机工程与科学, 2019, 41(11): 2007-2016. http://dx.chinadoi.cn/10.3969/j.issn.1007-130X.2019.11.015

[18] Chen, P., Liu, S., Zhao, H. and Jia, J. (2020) GridMask Data Augmentation. arXiv: 2001.04086. https://arxiv.org/abs/2001.04086

[19] Cubuk, E., Zoph, B., Mane, D., Vasudevan, V. and Le, Q. (2019) AutoAugment: Learning Augmentation Policies from Data. 2019 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 113-123. https://doi.org/10.1109/CVPR.2019.00020

[20] Cubuk, E.D., Zoph, B., Shlens, J. and Le, Q.V. (2020) Randaugment: Practical Automated Data Augmentation with a Reduced Search Space. 2020 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition Workshops*, Seattle, 14-19 June 2020, 3008-3017. https://doi.org/10.1109/CVPRW50498.2020.00359

[21] Li, Y., Hu, G., Wang, Y., Hospedales, T., Robertson, N. and Yang, Y. (2020) DADA: Differentiable Automatic Data Augmentation. *European Conference on Computer Vision* 2020, Glasgow, 23-28 August 2020, 580-595. https://doi.org/10.1007/978-3-030-58542-6_35

[22] Ho, D., Liang, E., Stoica, I., Abbeel, P. and Chen, X. (2019) Population Based Augmentation: Efficient Learning of Augmentation Policy Schedules. *Proceedings of the* 36*th International Conference on Machine Learning*, Long Beach, 10-15 June 2019, 2731-2741.

[23] Lim, S., Kim, I., Kim, T., Kim, C. and Kim, S. (2019) Fast AutoAugment. *Proceedings of Advances in Neural Information Processing Systems*, Vancouver, 8-14 December 2019, 6665-6675.

[24] Radford, A., Metz, L. and Chintala, S. (2015) Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv: 1511.06434. https://arxiv.org/abs/1511.06434

[25] Ioffe, S. and Szegedy, C. (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. arXiv: 1502.03167. https://arxiv.org/abs/1502.03167

[26] Zhu, J.-Y., Park, T., Isola, P. and Efros, A. (2017) Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 2223-2232. https://doi.org/10.1109/ICCV.2017.244

[27] Mirza, M. and Science, S.O.J.C. (2014) Conditional Generative Adversarial Nets. arXiv: 1411.1784. https://arxiv.org/abs/1411.1784

[28] Bao, J., Chen, D., Wen, F., Li, H. and Hua, G. (2017) CVAE-GAN: Fine-Grained Image Generation through Asymmetric Training. *Proceedings of* 2017 *IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 2764-2773. https://doi.org/10.1109/ICCV.2017.299

[29] He, Y., Schiele, B. and Fritz, M. (2018) Diverse Conditional Image Generation by Stochastic Regression with Latent Drop-Out Codes. *Proceedings of* 15*th European Conference*, Munich, 8-14 September 2018, 422-437. https://doi.org/10.1007/978-3-030-01270-0_25

[30] Chen, Q. and Koltun, V. (2017) Photographic Image Synthesis with Cascaded Refinement Networks. *Proceedings of*

2017 *IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 1520-1529. https://doi.org/10.1109/ICCV.2017.168

[31] Johnson, J., Alahi, A. and Li, F.F. (2016) Perceptual Losses for Real-Time Style Transfer and Super-Resolution. 2016 *European Conference on Computer Vision*, Amsterdam, 8-16 October, 694-711. https://doi.org/10.1007/978-3-319-46475-6_43