

文章编号: 2095-2163(2023)05-0187-04

中图分类号: TP391.41

文献标志码: A

基于深度残差注意力生成网络的数据增强图像分类

张目飞¹, 李廷¹, 苏鹏²

(1 浪潮云信息技术股份公司 服务研发部, 济南 250000; 2 山东浪潮新基建科技有限公司, 济南 250000)

摘要: 图像分类可能受到许多因素的影响, 本文提出用一个深度残差注意力生成网络生成图像数据来进行数据增强, 增加图像数据的多样性, 从而提高分类的准确率。通过与一些流行的深度学习分类方法做对比实验, 实验结果表明本文所提出的方法在分类性能上具有竞争力, 在 MNIST 和 cifar10 数据集上分别达到了 98.95% 和 92.68% 的分类准确率。

关键词: 图像分类; 残差注意力; 生成网络; 数据增强

Data-enhanced image classification based on deep residual attention generation networks

ZHANG Mufei¹, LI Yan¹, SU Peng²

(1 Service Research and Development Department, Langchao Cloud Information Technology JSC, Jinan 250000, China;

2 Shandong Langchao New Infrastructure Technology Co Ltd, Jinan 250000, China)

【Abstract】 Image classification may be affected by many factors. In this paper, we propose to generate image data with a deep residual attention generation network to perform data augmentation and increase the diversity of image data, thus improving classification accuracy. By doing comparison experiments with some popular deep learning classification methods, the experimental results show that the proposed method in this paper is competitive in classification performance, achieving 98.95% and 92.68% classification accuracy on MNIST and cifar10 datasets, respectively.

【Key words】 residual structure; generative adversarial networks; attention mechanism; image classification

0 引言

随着个人智能设备和图像相关应用的普及, 会产生大量的图像数据, 如何高效、合理地对这些图像数据进行合理的分类是一项技术难题。在过去的几年中, 深度神经网络(DNN)在计算机视觉和模式识别任务中, 如: 图像分类、语义分割、对象检测应用广泛。卷积神经网络中的卷积层能够捕获图像的局部特征, 以获得与输入维度相似的空间表示, 使用全连接层和 softmax 分类层生成概率表示, 来达到分类效果^[1]。He 等^[2]提出了深度残差网络 ResNet34, 引入了残差结构, 可以更好地学习残差信息, 并在后续层中使用这些残差信息, 提高了图像分类的性能, 为深度学习领域带来了新的思路和方法。

许多基于深度神经网络, 在网络学习过程中添加注意力机制来获得图像中感兴趣区域, 通过选择给定输入的特征通道、区域来自动提取相关特

征^[3]。Woo 等^[4]将注意力机制模块集成到 CNN 中, 提高网络的特征表达能力, 从而提高了图像分类的准确率; Wang^[5]提出了残差注意网络, 残差结构可以使网络更好地学习图像中的特征, 通过添加注意力模块来学习图像中的局部区域特征; Park 等^[6]提出了一种新的注意力机制, 可以在空间和通道维度上同时进行特征加权, 更加准确地捕捉到图像中的重要信息; Xi 等^[7]提出用残差注意模块进行特征提取, 以增强分类任务中的关键特征, 抑制无用的特征; Liang^[8]提出将自下而上和自上而下的前馈注意力残差模块用于图像分类。以上工作说明残差结构和注意力机制都可以帮助模型更好地学习图像特征, 提高图像分类的准确性。

随着数据集规模的增大和类别的增多, 训练一个高准确率的分类模型变得越来越困难。传统的数据增强方法对原始图像进行几何变换或者对图像进行随机扰动, 虽然可以增加数据集的样本量, 提高分

作者简介: 张目飞(1985-), 男, 硕士, 中级工程师, 主要研究方向: 深度学习、数据挖掘、云计算; 李廷(1984-), 男, 学士, 高级工程师, 主要研究方向: 图像处理、模式识别; 苏鹏(1983-), 男, 硕士, 高级工程师, 主要研究方向: 大数据、通信系统。

通讯作者: 张目飞 Email: zhangmufeilc@inspur.com

收稿日期: 2022-12-05

类模型的准确率,但是这些方法无法生成新的数据分布。而生成网络是一种可以学习数据分布的生成模型,可以生成新的样本,从而扩大数据集并且增加数据多样性,从而可以提高分类模型的泛化性^[9]。因此,本文提出一个深度残差注意力生成网络来生成图像数据,对数据进行必要的数据增强,利用 ResNet34 网络进行图像分类。

1 深度残差注意力生成网络数据增强模型和图像分类

1.1 深度残差注意力生成网络模型结构设计

本文提出了一个深度残差注意力生成网络模型用于图像数据增强,主要结构包括生成器、判别器和残差注意力模块。生成器包含 4 个反卷积层 (DConv) 和 3 个残差注意力模型 (SPAM),残差注意力模型能够对图像的重点区域进行特别关注,以生成高质量的图像,在生成器的最后一层使用 Tanh 函

数将数据映射到 $[-1, 1]$ 的区间内;判别器包括 4 个卷积层 (Conv),能够提取图像细节特征。深度残差注意力生成网络模型结构如图 1 所示。

生成网络由生成器和判别器组成。生成器将随机向量 Z 作为输入,学习真实数据分布 $p(x)$ 从而合成逼真的图像;判别器区分生成的图像与真实的图像,其输出表示从真实分布 $p(x)$ 提取样本 y 的概率。生成网络的最终目标是让生成器生成和真实图像相同的数据分布,而判别器无法判定图像为真实图像还是生成图像,达到一个纳什平衡。在生成器和判别器相互博弈的过程中,生成网络的目标函数定义为公式(1):

$$\min_G \max_D L_{GAN} = E_{x \sim p(x)} [\log D(x)] + E_{z \sim p(z)} [\log(1 - D(G(z)))] \quad (1)$$

其中, $p(x)$ 表示真实数据分布; $p(z)$ 表示生成数据分布; $D(x)$ 表示判别器运算; $G(z)$ 表示生成器运算。

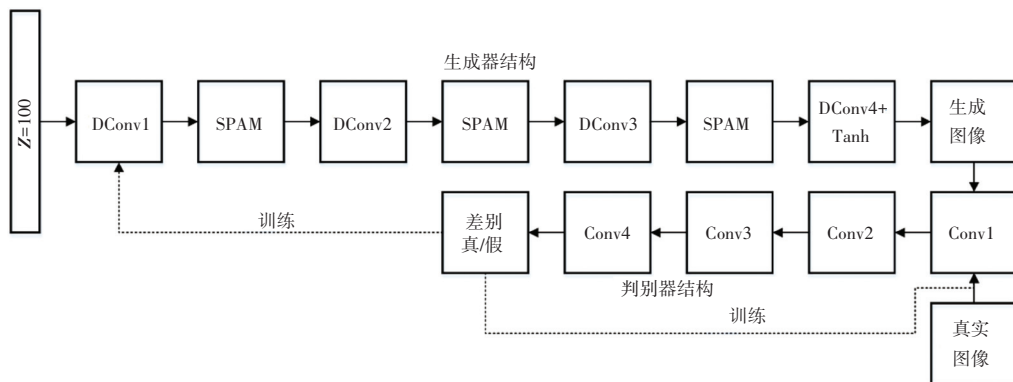


图 1 深度残差注意力生成网络模型结构

Fig. 1 Deep residual attention generation network model

本文随机选取 $Z = 100$ 维的随机数据作为生成器的输入,经过生成器生成图像;判别器网络的输入为生成图像和真实图像,判别器网络指导生成器合成图像,鼓励生成器捕捉更为精细的特征细节,使得生成器生成的图像和真实图像难以区分。

残差注意力模型使具有相似特征的区域相互增强,以突出全局视野中的感兴趣区域,残差注意力模型如图 2 所示。通过 sigmoid 函数可以得到一个 $[0, 1]$ 的系数,给每个通道或空间分配不同的权重,可以给每个特征图分配不同的重要程度。

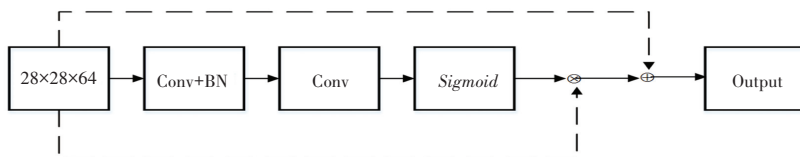


图 2 残差注意力模型

Fig. 2 Residual attention model

本文设 $C \times H \times W$ 为残差注意力模型的输入, C 为特征图的数量, H 和 W 分别表示为图像的高度和

宽度;通过卷积和批量归一化运算对输入的特征进行处理,利用 Sigmoid 函数得到空间注意系数 S ;将

输入的特征图和通过注意力模型得到的特征图利用残差结构进行融合, 得到最终的残差空间注意力特征表示, 公式(2) 和公式(3):

$$S = \text{sigmoid}(\text{Conv}(X)) \quad (2)$$

$$\text{Output} = S \cdot X + X \quad (3)$$

其中, X 表示空间注意模型的输入, Conv 表示卷积运算。

1.2 数据预处理和图像分类

首先, 对输入图像进行数据预处理, 主要包括: 将图像裁剪为 28×28 的大小, 并进行随机旋转和对比度增强; 其次, 将预处理的数据送入到深度残差注意力生成网络中进行数据增强。深度残差注意力生成网络通过学习图像不变性特征, 合成高质量的数据, 注意力机制对图像的兴趣区域进行重点关注; 生成器通过学习随机数据来生成感兴趣的图像分布, 判别器学习真实样本的分布, 辨别生成器生成的图像; 同时训练生成器和判别器, 促使两者竞争, 在理想情况下, 生成器可以生成近似于真实的图像数据, 而判别器不能将真实图像与生成图像区分, 从而达到纳什平衡, 达到数据增强的目的; 最后, 利用 ResNet34 网络对增强的图像数据进行分类。

2 实验结果分析

2.1 实验参数设置

本文使用 PyTorch 深度学习框架来训练模型, GPU 为 NVIDIA Tesla V100, 显存为 32 GB。采用 Adam 算法优化损失函数, 采用小批量样本的方式训练深度学习模型, batch_size 设置为 64, 在训练的过程中采用固定步长策略调整学习率, 初始学习率设置为 0.000 1, gamma 值为 0.85, L2 正则化系数设置为 0.000 1, 迭代次数为 50 000 次。

2.2 数据集

本文采用的数据集为 MNIST 数据集和 cifar10 数据集。MNIST 数据集一共有 70 000 张图片, 其中 60 000 张作为训练集, 10 000 张作为测试集, 每张图片由 28×28 的 0~9 的手写数字图片组成; cifar10 数据集由 60 000 张 32×32 的彩色图片组成, 一共有十个类别, 每个类别有 6 000 张图片, 其中 50 000 张图片作为训练集, 10 000 张图片作为测试集。

2.3 实验结果

使用深度残差注意力生成网络分别对 MNIST 和 cifar10 数据集中的图像进行图像增强, 使得图像的特征更加多样, 对 MNIST 数据集进行数据增强的效果如图 3 所示, 对 cifar10 数据进行数据增强的

效果如图 4 所示。



图 3 MNIST 数据集数据增强的效果

Fig. 3 Effect of data enhancement of MNIST dataset

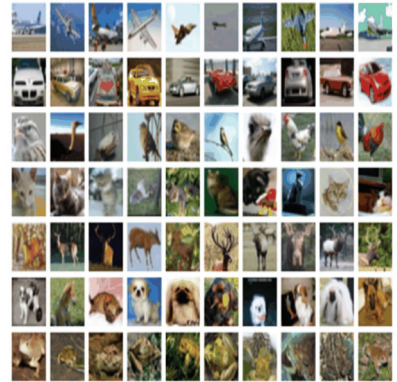


图 4 cifar10 数据集数据增强的效果

Fig. 4 Effect of data enhancement on the cifar10 dataset

从图 3 和图 4 可以看出, 使用深度残差注意力生成网络对 MNIST 和 cifar10 数据集进行数据增强, 具有很强的视觉可读性, 同时也具有较清晰的纹理特征, 实现了数据增强, 扩充了数据集。

为了验证本文模型数据增强后的 MNIST 以及 cifar10 数据在分类方面的效果, 选择 CNN、ResNet18、ResNet34、ResNet50 和 ResNet101 作为分类网络做对比实验。第一组测试增强数据的分类准确率; 第二组, 测试原始数据的分类准确率; 第三组, 将增强数据和原始数据各拿出 50% 组成新的数据集进行测试, 实验结果见表 1 和表 2。

表 1 MNIST 数据集分类准确率实验结果 (%)

Tab. 1 Experimental results of classification accuracy of MNIST dataset (%)

模型	生成数据	原始数据	生成+原始
CNN	96.49	97.84	98.36
ResNet18	96.93	97.92	98.52
ResNet34	97.26	98.02	98.95
ResNet50	96.43	97.65	98.35
ResNet101	96.73	97.17	97.96