

迁移学习与对抗生成网络结合的图片分类方法

孙 勇¹
SUN Yong

摘要

在深度图像分类任务中,传统方法依赖于预训练的卷积神经网络(CNN)进行特征提取并结合标准分类损失函数进行训练。然而,数据集样本不足和特征表示不足会影响模型的性能。为解决这一问题,一种结合迁移学习和对抗生成网络(GAN)的创新方法被提出。首先,利用在大规模数据集上预训练的ResNet模型提取图像的高层特征,通过微调使其适应新的分类任务;然后,训练GAN生成高质量的图像数据,增强训练数据集的多样性,从而提升模型的泛化能力。实验在CIFAR-10和ImageNet数据集上进行,结果表明,结合迁移学习和GAN的方法显著提高了图像分类的准确性和鲁棒性。所提出的方法为解决数据样本不足问题提供了有效的解决方案。

关键词

图像分类;迁移学习;对抗生成网络;卷积神经网络;特征提取

doi: 10.3969/j.issn.1672-9528.2024.11.007

0 引言

随着深度学习技术的不断进步,卷积神经网络(CNN)在图像分类任务中展现出卓越的性能。然而,模型的性能通常依赖于大量的标注数据和复杂的特征提取过程,尤其是在处理具有高度复杂性和多样性的图像数据时,传统的方法常常显得力不从心。面对数据样本不足和特征表示不足的问题,需要探索新的解决方案来提升模型的性能。

迁移学习通过使用在大规模数据集(如ImageNet)上预训练的模型,可以有效减少训练时间和计算资源消耗。这些预训练模型已经学习了丰富的特征表示,只需进行微调即可适应新的分类任务。另一方面,对抗生成网络(GAN)作为一种新兴的生成模型,能够生成高质量、具有多样性的图像数据,进一步扩展和增强训练数据集。

本文提出了一种结合迁移学习和对抗生成网络(GAN)的方法^[1],以提升图像分类模型的性能。具体方法包括:使用预训练模型提取特征,利用GAN生成高质量图像数据,结合这两种技术进行联合训练。通过引入预训练模型的强大特征提取能力和GAN^[2]生成数据的多样性,本文提出的方法提供了一种高效且实用的图像分类解决方案,为进一步研究和应用提供了新的思路和方法。

1 迁移学习

迁移学习^[3]是一种有效利用在大规模数据集上预训练的

模型来提升新任务性能的方法。在图像分类任务中,迁移学习通过提取预训练模型的特征并进行微调,可以显著提高分类模型的效果,具体流程如图1所示。本文选择ResNet模型,该模型在ImageNet数据集上经过预训练,已经学习了丰富的特征表示。在本研究中,迁移学习的主要过程包括预训练模型的选择、特征提取和模型微调。

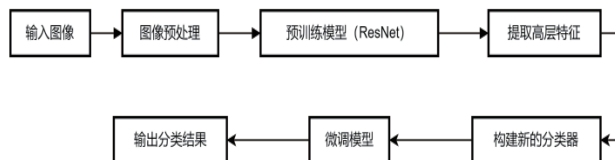


图1 迁移学习过程

首先,选择在ImageNet上预训练的ResNet模型。ResNet模型在大规模数据集上已经展示了卓越的性能,其通过引入残差学习的概念,解决了深层网络训练中的梯度消失问题。因此,ResNet模型成为特征提取的首选。在本研究中,移除ResNet模型的最后一层全连接层,仅使用中间层输出的特征向量进行特征提取。输入图像经过预处理后被输入到预训练的ResNet模型中,通过中间层提取其高层特征。设输入图像为 X ,经过预训练模型的特征提取层后,得到特征向量 $f(X)$ 公式为:

$$f(X) = \text{ResNet}(X) \quad (1)$$

式中: $f(X)$ 为高维特征向量,表示图像的深层特征。这些特征向量保留了图像中的丰富信息,为后续的分类任务提供了坚实的基础。

1. 南京审计大学 江苏南京 211815

接下来,将预训练模型提取的特征作为新的输入,构建一个新的分类器。新的分类器由若干全连接层组成,通过对新数据集进行微调,使模型参数适应特定的分类任务。在微调过程中,通过交叉熵损失函数来优化模型参数,使得模型能够准确地分类输入图像。交叉熵损失函数定义为:

$$L_{CE} = - \sum_{i=1}^N y_i \log(\hat{y}_i) \quad (2)$$

式中: y_i 是真实标签, $\hat{y}_i$ 是模型预测的概率分布, N 是训练样本数量。通过最小化交叉熵损失,能够最大化模型对正确类别的预测概率,从而提高分类准确性。

在模型微调过程中,采用分阶段训练的方法。首先,仅训练新分类器的参数,保持预训练模型的参数不变。这一步骤的目的是快速适应新任务,同时保留预训练模型的特征提取能力。接下来,逐步解冻预训练模型的部分参数,与新分类器的参数一起进行微调。通过这种逐步解冻的方法,可以有效地避免模型过拟合,同时充分利用预训练模型的特征提取能力。为了进一步增强模型的性能,在微调过程中引入了一些数据增强技术。这些技术包括随机裁剪旋转、缩放和翻转等操作,旨在增加训练数据的多样性,防止模型过拟合。例如,随机裁剪可以生成不同的图像视图,使模型能够学习到更多的图像细节;旋转和缩放则可以改变图像的几何特性,提高模型对不同姿态和尺度的鲁棒性。

在实际应用中,本研究使用 TensorFlow 和 Keras 等深度学习框架实现上述过程。首先,加载预训练的 ResNet 模型,并移除其最后一层全连接层;然后,构建新的分类器,添加若干全连接层和 SoftMax 输出层;接着,将预训练模型的输出连接到新的分类器上,形成完整的迁移学习模型;最后,通过定义交叉熵损失函数和优化器,进行模型的训练和微调。

在训练过程中,采用批量梯度下降(SGD)优化器,并设置初始学习率和动量参数。为了进一步提高模型的训练效率和稳定性,还使用了学习率调度策略。在训练的前期,采用较高的学习率快速收敛;随着训练的进行,逐步降低学习率,以细化模型参数,提高分类准确性。

通过上述迁移学习方法充分利用预训练模型的特征提取能力,并通过微调使其适应新的分类任务。

2 对抗生成网络(GAN)

对抗生成网络(GAN)是一种深度学习模型,通过两个网络——生成器(generator)和判别器(discriminator)相互对抗进行训练,从而生成逼真的图像数据。GAN的主要思想是通过生成器生成的假数据与真实数据相混合,并通过判别器来区分真假,最终使生成器能够生成与真实数据分布相近的高质量数据。

在本研究中,GAN的应用主要包括生成高质量的图像数据,增强训练数据集的多样性^[4],提高图像分类模型的泛化能力。GAN的训练过程可以分为以下几个关键步骤:生成器的设计、判别器的设计、损失函数的定义和训练过程的优化。

首先,生成器是一个深度神经网络,它接受一个随机噪声向量 z 作为输入,生成一幅假图像 $G(z)$ 。生成器的目标是通过学习,使生成的假图像尽可能逼真,以欺骗判别器。生成器的网络结构通常包括多个转置卷积层(transposed convolution)和激活函数。生成器的输出可以表示为:

$$G(z) = \text{Generator}(z) \quad (3)$$

其次,判别器也是一个深度神经网络,它接受一幅图像作为输入,输出一个概率值,表示输入图像为真实图像的概率。判别器的目标是区分真实图像和生成的假图像^[5]。判别器的网络结构通常包括多个卷积层和激活函数。判别器的输出可以表示为:

$$D(x) = \text{Discriminator}(x) \quad (4)$$

式中: x 可以是生成器生成的假图像 $G(z)$ 或真实图像,判别器的输出 $D(x)$ 表示输入图像为真实图像的概率。

GAN的训练过程是生成器和判别器相互对抗^[6]的过程。生成器希望最大化判别器对假图像的错误判别概率,而判别器希望最大化对真实图像和假图像的正确判别概率。整个训练过程可以通过以下两个损失函数来实现。

生成器的损失函数:

$$L_G = \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (5)$$

判别器的损失函数:

$$L_D = -\mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] - \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (6)$$

式中: $P_{\text{data}}(x)$ 是真实数据的分布, $p_z(z)$ 是随机噪声的分布。在训练过程中,交替优化生成器和判别器的损失函数,使生成器生成的图像越来越逼真,判别器的判别能力也越来越强。

通过以上步骤,GAN能够生成与真实数据分布相近的高质量图像,这些生成的图像可以用于扩展训练数据集,增强模型的泛化能力。特别是在数据样本不足的情况下,GAN生成的数据可以显著提升分类模型的性能。

3 实验

本实验旨在验证结合迁移学习和对抗生成网络(GAN)的方法在图像分类任务中的有效性。整个实验过程包括数据集选择与预处理、模型训练与微调,以及性能评估三个主要步骤。

首先,选择 CIFAR-10 和 ImageNet 数据集作为实验数据,这些数据集包含多种类的图像,适合作为分类任务的标准数

数据集。在数据预处理中，对图像进行标准化处理，并采用数据增强技术（如随机裁剪、旋转、缩放和翻转）增加训练数据的多样性，从而提高模型的泛化能力。接下来，采用迁移学习方法，利用在 ImageNet 上预训练的 ResNet 模型进行特征提取。具体步骤包括：加载预训练的 ResNet 模型，移除其最后一层全连接层，提取中间层的高维特征向量。设输入图像为 X ，经过特征提取层后得到特征向量 $f(X)$ 。然后，构建新的分类器，并使用提取的特征进行微调训练。新的分类器通过最小化交叉熵损失函数来优化参数。与此同时，训练 GAN 模型生成高质量的图像数据，以扩展和增强训练数据集。GAN 的训练过程包括交替优化生成器和判别器的损失函数。通过交替训练生成器和判别器，GAN 能够生成逼真的图像，扩展训练数据集，提高模型的分类性能。本研究的核心思想是结合迁移学习提取的高层特征与 GAN 生成的高质量图像数据，形成一个更为全面的训练集，以提升分类模型的泛化能力和准确性。在微调阶段，将迁移学习提取的特征与 GAN 生成的图像结合，通过一个新的损失函数进行联合优化。新的联合损失函数包括分类损失和对抗损失：

$$L_{\text{joint}} = \lambda L_{\text{CE}} + (1 - \lambda)L_{\text{GAN}} \tag{7}$$

式中： L_{CE} 是交叉熵损失； L_{GAN} 是对抗损失； λ 是权重参数，用于平衡两部分损失。

通过上述实验流程如图 2 所示，可以系统地验证结合迁移学习和 GAN 的方法在图像分类任务中的性能提升效果。核心思想是利用迁移学习提取的丰富特征和 GAN 生成的多样性数据，形成一个更强大的分类模型，提高其在不同数据集上的泛化能力和准确性。

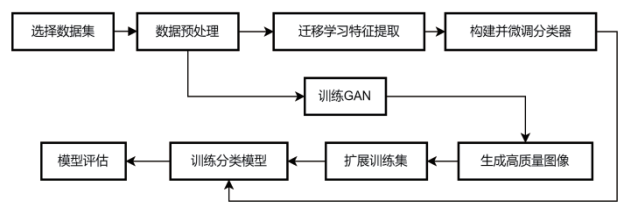


图 2 结合迁移学习和对抗生成网络（GAN）

3.1 CIFAR-10 数据集实验

为了评估结合迁移学习与对抗生成网络（GAN）的方法在图像分类任务中的有效性，使用 CIFAR-10 数据集进行实验。CIFAR-10 数据集是公认的图像分类任务基准数据集，包含 10 个类别的 60 000 张 32 像素 × 32 像素的彩色图像，其中 50 000 张用于训练，10 000 张用于测试。每个类别有 6000 张图像，涵盖了飞机、汽车、鸟类、猫、鹿、狗、青蛙、马、船和卡车等常见物体。CIFAR-10 数据集中的每张图像都是 32 像素 × 32 像素的彩色图片，分为 10 个不同的类别。训练集和测试集分别包含 50 000 张和 10 000 张图像，训练集中每个

类别有 5000 张图像，测试集中每个类别有 1000 张图像。数据集中的图像分辨率较低，增加了分类任务的难度。每张图像都经过手工标注，确保了高质量的标签。

实验详细设置：在实验中，首先对 CIFAR-10 数据集进行标准化处理，使每个像素值在 0 ~ 1 之间。接着，采用数据增强技术对训练数据进行处理，包括随机裁剪、水平翻转和随机旋转，以增加数据的多样性。这些增强操作有助于防止模型过拟合，并提高模型的泛化能力。

在模型训练方面，首先使用迁移学习方法，从预训练的 ResNet 模型中提取特征。具体步骤包括：加载预训练的 ResNet 模型，移除最后一层全连接层，提取中间层的高维特征向量。然后，构建一个新的分类器，将提取的特征作为输入，进行微调训练。训练过程中使用交叉熵损失函数优化模型参数，学习率设置为 0.01，批量大小为 64。

与此同时，训练 GAN 模型以生成高质量的图像数据^[7]，增强训练数据集。GAN 模型的生成器接受一个随机噪声向量 z 作为输入，生成假图像 $G(z)$ ；判别器接受一幅图像作为输入，输出其为真实图像的概率。GAN 的训练过程包括交替优化生成器和判别器的损失函数^[8]，生成器损失函数为 L_G ，判别器损失函数为 L_D 。

通过交替训练生成器和判别器，生成与真实数据分布相近的高质量图像。然后，将这些生成的图像加入到训练集中，以增强数据集的多样性。

实验结果分析：结合迁移学习和 GAN 的方法显著提高了分类模型的性能。如表 1 所示，在 ResNet 模型中，结合 GAN 生成的数据后，top-1 准确率从 86.5% 提高到 89.3%，top-5 准确率从 97.8% 提高到 98.5%。在使用 MobileNetV2 模型时，top-1 准确率从 84.1% 提高到 87.4%，top-5 准确率从 96.3% 提高到 97.6%。

表 1 CIFAR-10 数据集实验结果

模型	方法	Top-1 准确率	Top-5 准确率
ResNet	迁移学习	86.5%	97.8%
	迁移学习 +GAN	89.3%	98.5%
MobileNetV2	迁移学习	84.1%	96.3%
	迁移学习 +GAN	87.4%	97.6%

3.2 ImageNet 数据集实验

为了进一步验证结合迁移学习与对抗生成网络（GAN）的方法在图像分类任务中的有效性，使用 ImageNet 数据集进行实验。ImageNet 数据集是图像分类任务中最为广泛使用的基准数据集之一，包含超过 1400 万张图像，涵盖了 1000 个不同的类别。ImageNet 数据集中的每张图像都属于特定的类别，图像分辨率较高，包含丰富的细节信息。由于数据量庞

大, 本文选择其中一个子集进行实验, 包括 100 个类别的图像, 每个类别包含 1000 张图像, 总计 100 000 张图像。数据集划分为训练集和测试集, 比例为 8:2, 即 80 000 张图像用于训练, 20 000 张图像用于测试。

实验详细设置: 在实验中, 首先对 ImageNet 数据集进行标准化处理, 使每个像素值在 0 ~ 1 之间^[9]。接着, 采用数据增强技术对训练数据进行处理, 包括随机裁剪、水平翻转和颜色抖动, 以增加数据的多样性。这些增强操作有助于防止模型过拟合, 并提高模型的泛化能力。

在模型训练方面, 首先使用迁移学习方法, 从预训练的 ResNet 模型^[10]中提取特征。具体步骤包括: 加载预训练的 ResNet 模型, 移除最后一层全连接层, 提取中间层的高维特征向量。然后, 构建一个新的分类器, 将提取的特征作为输入, 进行微调训练。训练过程中使用交叉熵损失函数优化模型参数, 学习率设置为 0.001, 批量大小为 32。

与此同时, 训练 GAN 模型以生成高质量的图像数据, 增强训练数据集。GAN 模型的生成器接受一个随机噪声向量 z 作为输入, 生成假图像 $G(z)$; 判别器接受一幅图像作为输入, 输出其为真实图像的概率。GAN 的训练过程包括交替优化生成器和判别器的损失函数^[11]。通过交替训练生成器和判别器, 生成与真实数据分布相近的高质量图像。然后, 将这些生成的图像加入到训练集中, 以增强数据集的多样性。

实验结果显示, 结合迁移学习和 GAN 的方法显著提高了分类模型的性能。如表 2 所示, 在 ResNet 模型中, 结合 GAN 生成的数据后, top-1 准确率从 75.3% 提高到 78.7%, top-5 准确率从 92.4% 提高到 94.1%。在使用 MobileNetV2 模型时, top-1 准确率从 72.8% 提高到 76.5%, top-5 准确率从 90.2% 提高到 92.8%。

表 2 ImageNet 数据集实验结果

模型	方法	Top-1 准确率	Top-5 准确率
ResNet	迁移学习	75.3%	92.4%
	迁移学习 +GAN	78.7%	94.1%
MobileNetV2	迁移学习	72.8%	90.2%
	迁移学习 +GAN	76.5%	92.8%

4 总结

本文提出了一种结合迁移学习和对抗生成网络 (GAN) 的图像分类方法。通过使用预训练模型提取高层特征, 利用 GAN 生成高质量的图像数据, 增强训练数据的多样性。迁移学习利用在大规模数据集上预训练的模型, 通过微调适应新的分类任务。GAN 通过生成逼真的图像数据扩展训练集, 进一步提升模型的泛化能力。实验结果表明, 该方法在不同数

据集上均显著提高了分类模型的准确性和鲁棒性。本文的方法为解决数据样本不足和提高模型性能提供了一种有效的解决方案, 展示了良好的应用前景。

参考文献:

[1] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016:770-778.

[2] 张晨曦, 姚琼, 秦飞巍, 等. 浅析生成对抗网络及其在计算机视觉中的应用 [J]. 计算机时代, 2023(8):11-15.

[3] 吴健, 贾宏宇. 基于迁移学习的图像分类方法研究 [J]. 河南科技, 2018(31):20-22.

[4] 张松. 基于生成对抗网络的数据增强方法及应用 [D]. 徐州: 中国矿业大学, 2023.

[5] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. (2020). An image is worth 16×16 words: transformers for image recognition at scale[DB/OL].(2020-10-22) [2024-02-22].<https://doi.org/10.48550/arXiv.2010.11929>.

[6] XU J, SUN F M. Unsupervised deep domain adaptation based on weighted adversarial network[J]. IEEE access, 2020, 8:64020-64027.

[7] HUANG G, LIU Z, DER MAATEN L V, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition(CVPR). Piscataway: IEEE, 2017: 700-708.

[8] LI W H, CHEN Z Y, HE G L. A novel weighted adversarial transfer network for partial domain fault diagnosis of machinery[J]. IEEE transactions on industrial informatics, 2021,17(3):1753-1762.

[9] 田腾飞. 基于生成式对抗网络的图像特征表示及应用 [D]. 南京: 东南大学, 2018.

[10] 陈文兵, 管正雄, 陈允杰. 基于条件生成式对抗网络的数据增强方法 [J]. 计算机应用, 2018,38(11):3305-3311.

[11] MIRZA M, OSINDERO S. Conditional generative adversarial nets[DB/OL].(2014-11-06)[2024-03-12]. <https://doi.org/10.48550/arXiv.1411.1784>.

【作者简介】

孙勇 (1998—), 男, 江苏泰州人, 硕士研究生, 研究方向: 大数据审计、计算机视觉、深度学习。

(收稿日期: 2024-07-28)