

文章编号: 2095-2163(2022)09-0056-09

中图分类号: TP183

文献标志码: A

一种有效神经网络训练优化方法

陈青, 杨晶东, 王晗, 彭坤

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

摘要: 指数移动平均(EMA)算法,通常可用于过滤由小批量梯度下降引起的噪声,提高模型鲁棒性。然而,传统EMA算法在持续训练后期,无法有效优化网络参数,深层神经网络经常出现拟合。因此,本文提出一种以变系数Tanh为衰减函数的动态衰减EMA算法,结合SGD优化器的T-ADEMA+SGD算法,进行神经网络训练。针对MNIST、CIFAR_10、CIFAR_100数据集,采用优化器SGD训练ResNet50模型,并针对胸部X射线图像训练Vision Transformer(ViT)模型,同时采用深度卷积生成对抗网络(DCGAN)进行数据增强和基于 t 分布,随机邻域嵌入(t -SNE)模型用于可视化分析。实验表明,对于CIFAR_100测试集,T-ADEMA+SGD算法的准确率、精度、召回率和 F_1 值分别为74.15%、74.39%、74.15%、74.04%;而对于Kaggle COVID-19三分类图像,相应的评价指标分别为87.94%、91.19%、84.43%、86.87%,与典型算法相比,本文模型可以更好地根据训练时间,动态调整最优参数、降低噪声,具有更好的泛化性能,适用于各种常用数据集。

关键词: 指数移动平均; 深层神经网络; 衰减; 泛化性能

A novel optimization method of neural network training

CHEN Qing, YANG Jingdong, WANG Han, PENG Kun

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

[Abstract] The exponential moving average (EMA) algorithm can often be used to filter noises caused by mini-batch gradient descent for the improvement of model robustness. However, the traditional EMA algorithm cannot optimize the network parameters effectively at the late period of continuous training, and the overfitting often occurs for the deep networks. Therefore, the paper proposes a novel EMA algorithm (T-ADEMA) with dynamic decay which regards Tanh function with variable coefficient as the decay function. In addition, T-ADEMA algorithm is used to train ResNet50 with the optimizer SGD on three different datasets, i.e. MNIST, CIFAR_10, CIFAR_100, and train Vision Transformer (ViT) on chest X-ray images, including data augmentation based on deep convolutional generative adversarial networks (DCGAN) and t -Distributed Stochastic Neighbor Embedding (t -SNE) for visualization. The experiments show that for CIFAR_100 test set, the Accuracy, Precision, Recall and F_1 Score of T-ADEMA algorithm are 74.15%, 74.39%, 74.15%, 74.04% respectively. And for chest X-ray test set, the corresponding evaluation indicators are 87.94%, 91.19%, 84.43%, 86.87% respectively. Moreover, in comparison with the other state-of-art algorithms, T-ADEMA+SGD algorithm can adjust optimal parameters better according to training times dynamically and reduce noises. The proposed method achieves better generalization performance and be suitable for variety of classical datasets.

[Key words] exponential moving average; deep neural networks; decay; generalization performance

0 引言

目前,卷积神经网络(CNN)已广泛应用于图像处理^[1]、语音识别^[2]、自然语言处理^[3]等领域,并取得了巨大成功。典型的CNN由卷积层、池化层和全连接层组成。卷积层和池化层能够自动提取深层特征和降维。然而,训练普通的CNN模型需要大量参数。LéCun等人^[4]提出的LeNets网络中,使用了大约1M个训练参数。Krizhevsky等人^[5]在ImageNet

竞赛中应用了包含60M个训练参数的ResNet网络,并使用了数据增强^[6],其中包括翻转、裁剪、亮度和对比度变换来增加数据量。Yuan等人^[7]提出了基于ResNet50的腮腺肿瘤CT图像分类模型,并加入北京DSG公司的数据脱敏策略,提高了预测性能。然而,脱敏方法只能消除CT图像的特征,不能模糊整个图像的背景。此外,1000次训练和90%的测试集精度表明,模型的收敛速度较慢,并且出现了一定的过度拟合。Sun等人^[8]改进了ResNet50,

基金项目: 国家自然科学基金(81973749)。

作者简介: 陈青(1996-),男,硕士研究生,主要研究方向:人工智能、机器学习与大数据分析等;杨晶东(1973-),男,博士,副教授,主要研究方向:人工智能、机器学习与大数据分析、机器视觉等;王晗(1997-),男,硕士研究生,主要研究方向:人工智能、机器学习与大数据分析等;彭坤(1993-),男,硕士研究生,主要研究方向:人工智能应用研究。

收稿日期: 2022-03-18

其中包括23 602 051个训练参数,通过使用深度卷积生成对抗网络(DCGAN)的数据增强,对苹果品质进行分类。这些操作可以为单个图像添加多个副本,以提高图像利用率和数据多样性,从而提升分类模型的性能。

然而,在同一数据集上进行多次训练后,具有大量参数的迭代会在一定程度上导致过拟合。训练集中的噪声也会对网络性能产生负面影响,导致泛化性能下降。通常使用随机梯度下降(SGD)算法^[9]训练网络参数,并使用 L_2 正则化和指数移动平均(EMA)算法优化网络性能。具有综合自回归^[10]的传统移动平均模型(ARIMA)、指数加权移动平均模型(EWMA)^[11-12]主要适用于时变序列的估计,如股市趋势预测^[13]、风速预测^[14]或旅游兴趣预测^[15]等。Huang等人^[16]提出了正则化和重新初始化指数移动平均(Re^2 EMA),用于更新目标跟踪中的目标模型。 Re^2 EMA施加惩罚以降低新生成目标模型的不可靠性,并使用重新初始化项来缓解退化水平,进一步应用正则化器来限制复杂性,表明了高效性和有效性。

上述EMA算法均使用固定的衰减率来更新网络参数,故而并不适合长时间、多步骤的深层神经网络训练。因此,本文提出了一种基于Tanh函数的动态衰减指数移动平均算法(T-ADEMA)。该算法以变系数Tanh作为衰减函数,动态自适应地调整模型参数,降低噪声。本文应用T-ADEMA算法对经典模型ResNet进行训练和测试,包括对MNIST、CIAFR_10、CIFAR_100数据集和当前流行的非CNN模型Vision Transformer(ViT)进行胸部X射线数据集的训练和测试,通过DCGAN算法进行数据增强。此外,还与传统的EMA算法进行了比较,分析了识别率和泛化性能。

1 动态衰减EMA算法

1.1 EMA算法

一组动态原始序列 $y(t)$ 由数据信号 $f(t)$ 和噪声信号 $e(t)$ 组成。其中, $f(t)$ 是期望的测量值或有效信号, $e(t)$ 是由噪声引起的随机信号。离散数据被视为一组时变数值序列,其表达式如下:

$$y(k) = f(k) + e(k) \quad k = 1, 2, 3, \dots \quad (1)$$

为了抑制 $e(k)$ 并提取 $f(k)$,通常对序列 $y(k)$ 进行平滑或滤波。对序列 $y(k)$ 进行适当的分段,在有数值振荡区间内,对一些合适的单元进行平滑和局部平均,以接近稳定区间,进一步削弱 $e(k)$ 引起的随机振荡。由于只选择了部分小区间

进行局部平均和平滑,因此整个序列近似于 $f(k)$ 。假设有一组离散序列 $x = \{x_1, x_2, x_3, \dots, x_n\}$ 以及一组参数序列 $\alpha = \{\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n\}$, $\alpha \in [0, 1]$,经过加权移动平均优化,序列 x 可由公式(2)得出:

$$\sigma_n := \alpha_n \sigma_{n-1} + (1 - \alpha_n) x_n \quad n > 1, \sigma_1 := x_1 \quad (2)$$

其中, σ_n 是序列 x_n 的指数移动平均值(EMA), α_n 是指数移动平均值的权重。

在EMA的计算中,权重 σ_i 越大,EMA序列越接近序列 x_i ,平均效应越大,反之亦然。此外,由于权重序列项是恒定的,因此可以使用当前移动平均值 x 及其之前的值来计算从 n 到 $n+1$ 的EMA。

1.2 动态衰减EMA算法

在神经网络训练中,每次迭代都会更新权重、偏移量和一些其它训练变量,以获得离散时间序列。假设神经网络模型的训练参数序列 $\omega = \{\omega_1, \omega_2, \dots, \omega_n\}$,则基于EMA算法同步更新的影子参数序列 $v = \{v_1, v_2, \dots, v_n\}$ 可由公式(3)得出:

$$v_n = decay * v_{n-1} + (1 - decay) \omega_n \quad (3)$$

其中, $decay$ 是衰减率,按经验通常取值为0.9或0.99。

按递归方式逐项展开式(3),可以得出公式(4):

$$v_n = \sum_{i=1}^n decay^{(n-i)} * (1 - decay) * \omega_i \quad (4)$$

为了在神经网络训练过程中根据不同的训练阶段更有效地过滤噪声,本文基于公式(5)中ADEMA算法^[17]的衰减系数,提出了基于tanh衰减系数的T-ADEMA算法,见式(6)。衰减系数 $decay$ 应满足以下公式:

$$decay = (1 + n) / (10 + n) \quad (5)$$

$$decay = \alpha * \tanh(n) \quad (6)$$

其中, n 表示训练迭代次数, α 是一个常值参数,控制移动平均值,根据经验将其设置为0.9。可以看出,第 n 个移动平均值 v 只与序列 $\{\omega_{n-c}, \dots, \omega_n\}$ 和衰减系数 $decay$ 有关。在用SGD优化器训练网络参数的同时,选择合适的学习速率 α (本文中的学习速率设置为0.001),可以加快网络收敛速度,在一定时间内学习足够多的有效特征。基于以上原因,本文采用T-ADEMA算法更新训练参数,有效地消除了训练噪声,提高了泛化性能。

2 基于动态EMA优化的分类模型

由于本文实验图像数据集的分辨率低于

ImageNet,因此该体系结构可以应用于 ResNet 网络。常见的 ResNet 网络由 50 层(包括池化层)和 3 个全连接层组成,每个层都有 $ReLU^{[16]}$ 作为激活函数。

本文提出了基于动态 EMA 算法的 T-ADEMA 算法来优化训练参数的权重,并将其应用于图像分类。详细的模型结构和参数如图 1 所示。首先,将大小为 $3 \times 224 \times 224$ 的原始图像输入残差网络^[17]模型。通过 $Conv1$ 和 $Maxpool$ 层后,输出图像的大小为 $64 \times 56 \times 56$ 。 $Conv2_x$ 包含 3 个 $Bottlenecks$,每个 $Bottleneck$ 包含 $Conv$ 模块(3 层虚线)和 $Identity$ 模块(3 层实线)。 $Conv$ 模块首先使用 1×1 卷积降低特

征映射的维数,然后执行 3×3 卷积操作来提取特征,最后使用 1×1 卷积恢复维数。卷积层连接到 BN (批量归一化)和 $ReLU$ 层。 $Identity$ 模块用于深化网络结构。 $Conv2_x$ 、 $Conv3_x$ 、 $Conv4_x$ 、 $Conv5_x$ 的 $Bottlenecks$ 数量分别为 3、4、6 和 3。 $Softmax$ 用于 $Avgpool$ 层和 FC 层之后的分类。 $MNIST$ 和 $CIFAR_{10}$ 有 10 个分类。 $CIFAR_{100}$ 包括 20 类粗分类和 100 类细分类。 $ReLU$ 用于每个全连接层的激活函数。全连接层的 2 层应用了 $Dropout^{[19]}$ 以减少过拟合。输出层使用 $Softmax$ 函数计算反向传播的误差^[20],进而获得分类概率。

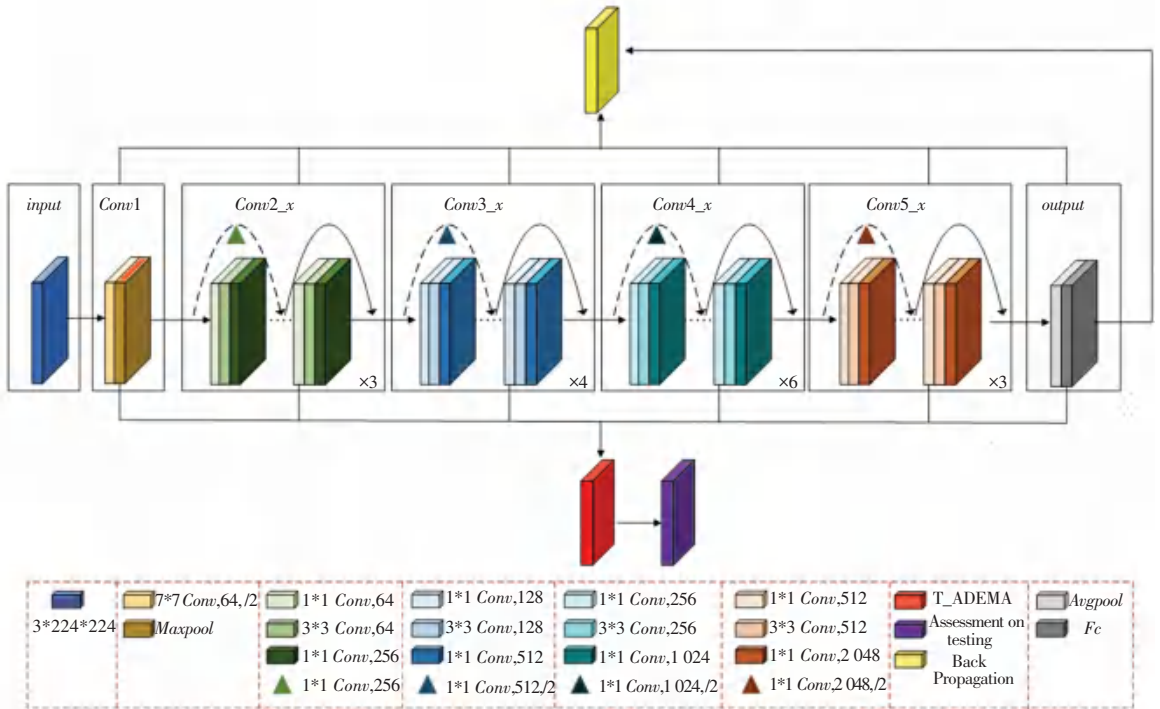


图 1 基于 T-ADEMA 优化的残差网络结构

Fig. 1 The ResNet framework based on T-ADEMA optimization

图 2 展示了基于 T-ADEMA 优化的 Vision Transformer (ViT) 框架,其中原始图像被划分为 16 个切片作为输入特征图像。针对 3 分类的胸部 X 光图像,本文将基于 T-ADEMA 的 ViT 模型与其他主流的 EMA 优化方法进行了比较实验。

3 实验预处理

实验使用的深度学习框架为 Pytorch。实验环境包括 Intel i7-10700 CPU、NVIDIA GeForce RTX 3070 以及 NVIDIA CUDA_CUDNN 加速器、Windows 10 20H2 和 Python3.8 | Pytorch 1.7.0。

3.1 数据集

本文将提出的 T-ADEMA 算法结合 SGD 优化器在 MNIST、CIFAR_10/100 和 COVID-19^[21] 等 4 个

数据集上验证其泛化性能。其中, MNIST 是一个公开手写的数字图像数据集,由 60 000 个训练样本和 10 000 个测试样本组成。样本是 28×28 的二值图像,具体如图 3(a) 所示。CIFAR_10 数据集是应用最广泛的图像识别数据集,由 50 000 个训练样本和 10 000 个测试样本组成,每个样本有 28×28 个像素,数据集拥有 10 个分类,如图 3(b) 所示。类似于 CIFAR_10 数据集, CIFAR_100 数据集拥有 100 个类,包括每个类的 500 个训练图像和 100 个测试图像,如图 3(c) 所示。图 3(d) 展示的 COVID-19 数据集由 3 个类别的胸部 X 光图像组成,即 COVID-19、正常、病毒性肺炎。COVID-19 数据集的图像分辨率为 1024×1024 ,共 2 900 例,这里的 215 例 COVID-19,1 340 例正常,1 345 例病毒性肺炎。

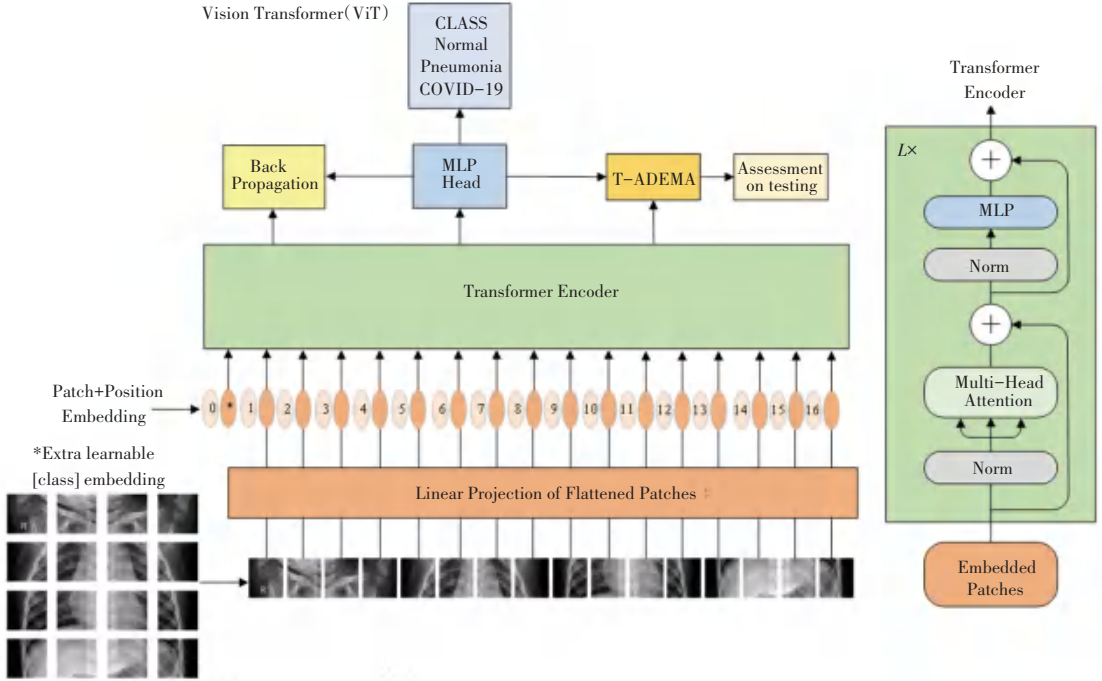


图 2 基于 T-ADEMA 优化的 ViT 网络结构

Fig. 2 The ViT framework based on T-ADEMA optimization

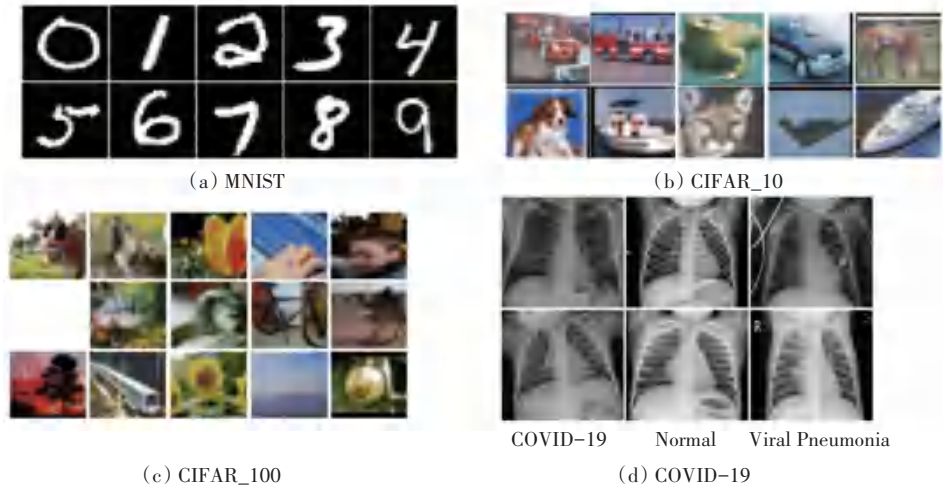


图 3 4 个数据集的训练样本

Fig. 3 The training samples of four datasets

3.2 图像增强

实验中,对 CIFAR_10/100 采用传统图像增强方法,以提高识别率。将训练数据放入模型训练前,随机进行左右、上下翻转,亮度变换或对比度变换;对于 COVID-19 数据集,采用 DCGAN 对数据集进行增强和扩充。图 4 显示了 DCGAN 的批量增强和扩充过程。保留 20% 的 3 类比例原始样本用于测试

集,剩余样本用于 DCGAN 进行数据扩充。DCGAN 图像增强后的样本分布见表 1。由表 1 可知,通过 DCGAN 分别生成 3 类图像样本,达到每个类别有 1 200 个样本。可以看出,在 DCGAN 算法的基础上,经过约 200 轮训练,增强后的图像与原始图像越来越相似,最终的增强图像在 DCGAN 算法的基础上取得了更好的效果。

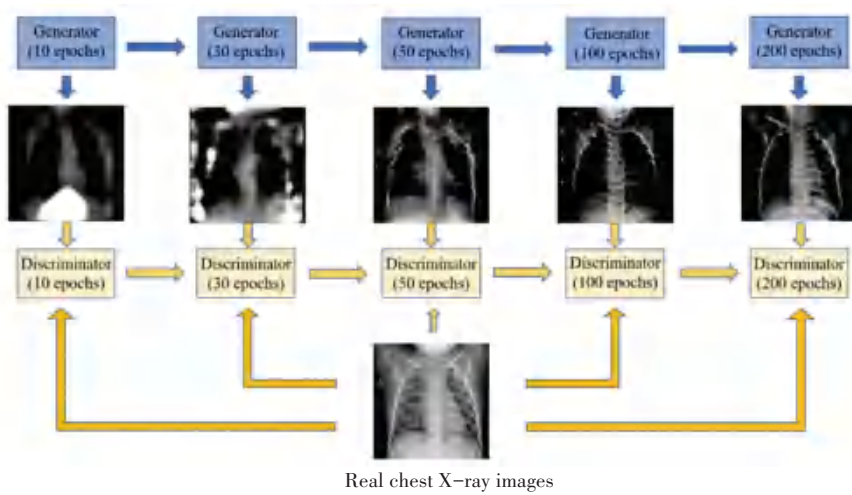


图4 COVID-19数据集的DCGAN图像增强

Fig. 4 The data augmentation via DCGAN algorithm on COVID-19 dataset

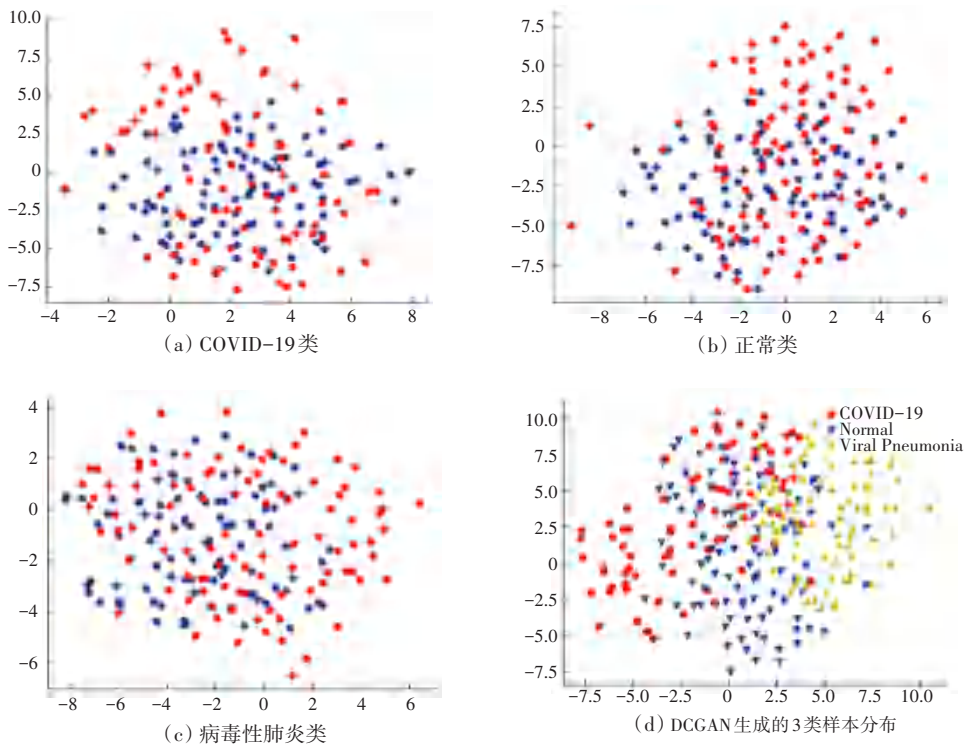
表1 DCGAN 图像增强后的样本分布

Tab. 1 Samples distribution after data augmentation via DCGAN algorithm

胸部X光图像	原始	扩充	训练集	测试集
COVID-19	215	1 028	1 200	43
普通	1 340	128	1 200	268
病毒性肺炎	1 345	124	1 200	269

为了验证基于 DCGAN 算法的数据增强效果,采用 t -分布随机近邻嵌入 t -SNE^[22],用来降低样本分布可视化的维数。 t -SNE 通过将高维空间映射为高斯分布概率,将相似性转化为相邻样本之间的概率。本文从每个类别中随机选取 100 幅真实图

像和 100 幅通过 DCGAN 算法生成的图像,分别通过 t -SNE 进行可视化。原始图像和生成图像通过 t -SNE 可视化分布如图 5 所示。从图 5(a)~图 5(c)可以看出,生成图像(红点)的分布与原始图像(蓝点)的分布近似。可以推断,生成的图像与原始图像具有相似性,可以应用于样本分类。此外,图 5(d)表明,每一类样本超过 1/3 的样本数是可分辨的,并且属于不同的类别。由此可以看出,DCGAN 算法不仅学习了胸部 X 射线图像的全局特征,而且能够分辨出局部细小特征的区别,能有效地识别 3 类样本。

图5 原始图像和生成图像通过 t -SNE 可视化分布Fig. 5 Distribution of real images and its corresponding generated images via t -SNE

4 实验分析

4.1 SGD+T-ADEMA 算法性能分析

为了验证 EMA 算法的有效性,分别在 MNIST、CIFAR_10/100 和 COVID-19 数据集上进行了对比实验,并使用 SGD 优化算法、固定衰减 EMA 算法 (SGD+EMA)、公式 (5) 中的动态衰减 ADEMA 算法、公式 (6) 中的动态衰减 T-ADEMA 算法 (SGD+T-ADEMA) 和主流优化算法 Adam^[23] 等 5 种优化算法,在固定或动态学习率和批量训练的情况下,优化了 ResNet50 和 ViT 分类模型。MNIST 训练集包含 60 000 个样本,批量大小为 30,每轮训练包括 2 000 次迭代。由于模型经过 5 轮训练后接近收敛,因此每 50 次迭代计算一次训练集损失,以记录收敛参数。对于数据集 CIFAR_10/100 和 COVID-19,在训练 100 轮左右后收敛,每次训练一个轮次计算训练集的损失。本文在 5 种不同的 EMA 算法上计算了 5 轮实验的评估指标均值,并分别训练 MNIST、CIFAR_10/100 和 COVID-19 数据集。评价指标包括准确率 (Accuracy)、精度 (Precision)、召回率 (Recall)、 F_1 值。研究推得的计算公式如下:

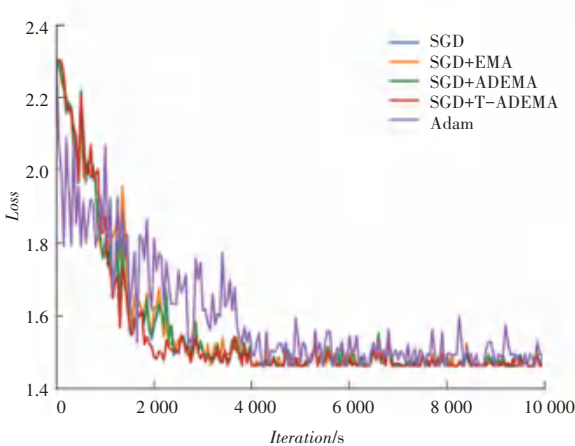
$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN} \quad (8)$$

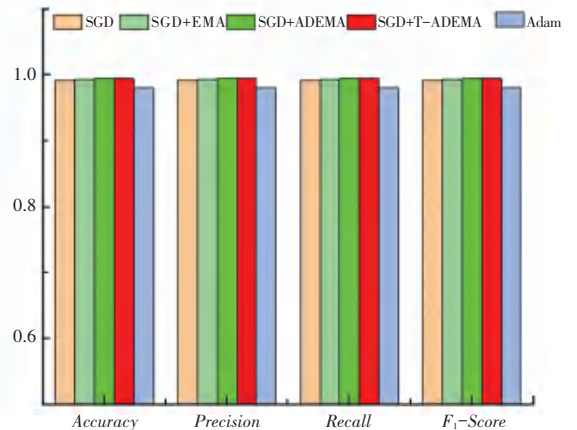
$$F_1\ score = \frac{2Precision \times Recall}{Precision + Recall} \quad (9)$$

其中, TP 、 TN 、 FP 和 FN 分别为真阳性、真阴性、假阳性和假阴性。

在 MNIST 数据集、CIFAR_10 数据集、CIFAR_100 数据集和 COVID-19 数据集上 5 种算法的损失函数和评估指标对比结果如图 6~图 9 所示。图 6 (a)、图 7 (a)、图 8 (a) 和图 9 (a) 分别显示了 5 种优化算法在各数据集上的训练损失函数曲线。实验结果表明,在训练初始阶段,SGD+T-ADEMA 算法只学习到少量图像特征,并不能有效抑制小批量数据中的噪声,其收敛速度较慢,仅相当于 ADEMA 算法在较少训练回合中的平均值,在早期训练迭代中下降最慢。这是因为 SGD+T-ADEMA 算法在滤除噪声方面效果更好,在早期训练阶段需要较少的学习数据,学习速度也相同,因此损失降低相对较慢,且该算法在经过多轮训练后能学习到更有效的特征。SGD+T-ADEMA 算法在训练中期,在 MNIST 和 CIFAR_100 数据集上收敛最快。虽然在 CIFAR_10 数据集的收敛速度略低于 ADEMA,但可以过滤更多的噪声。在训练后期,SGD+T-ADEMA 算法的损失曲线的收敛性明显优于其他 4 种算法。在 COVID-19 数据集上,SGD+T-ADEMA 算法与其他 EMA 算法相比收敛速度更快。经过几轮训练参数的更新,学习了足够多的特征,有效地过滤了噪声。测试集上 5 种算法的评估指标见表 2。



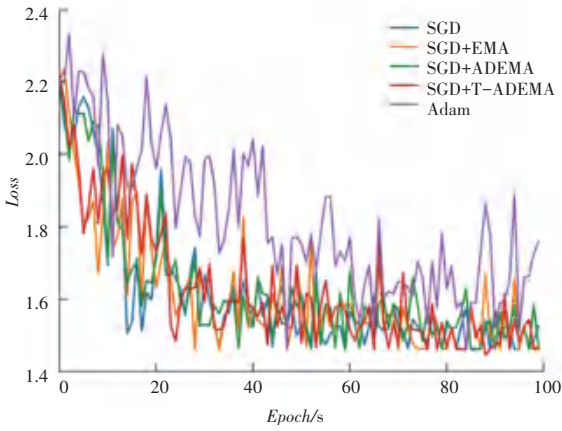
(a) 训练集损失函数曲线



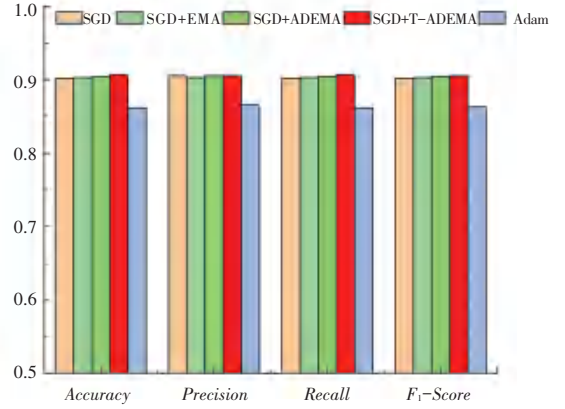
(b) 测试集评估指标

图 6 MNIST 数据集上 5 种算法的损失函数和评估指标对比

Fig. 6 The comparison of training loss and evaluation indicators of five algorithms on MNIST dataset



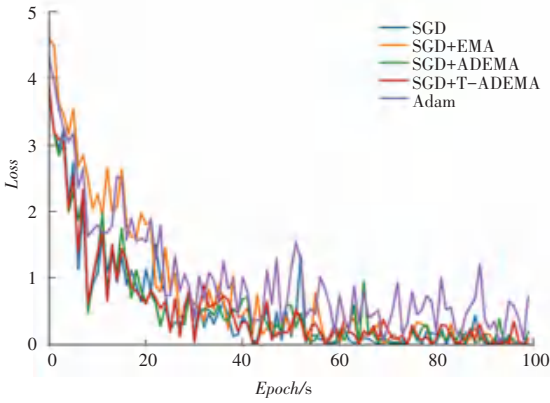
(a) 训练集损失函数曲线



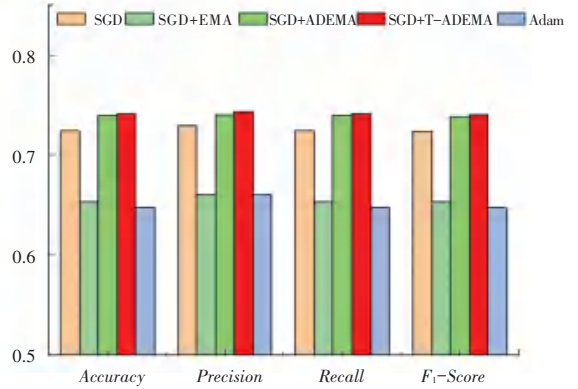
(b) 测试集评估指标

图7 CIFAR_10数据集上5种算法的损失函数和评估指标对比

Fig. 7 The comparison of training loss and evaluation indicators of five algorithms on CIFAR_10 dataset



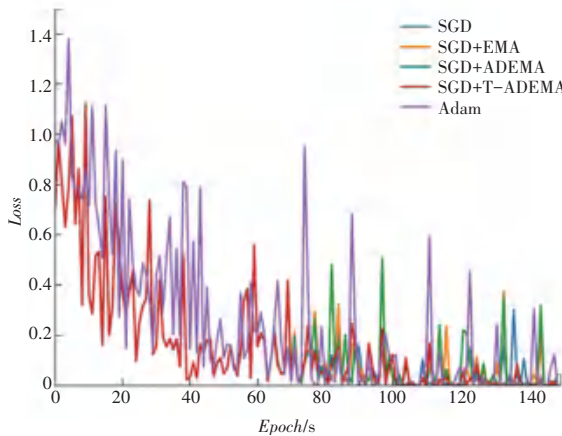
(a) 训练集损失函数曲线



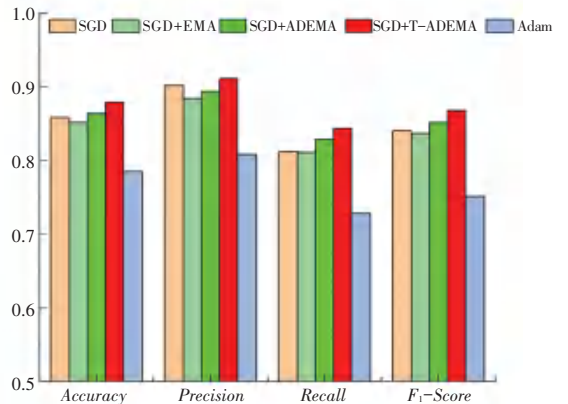
(b) 测试集评估指标

图8 CIFAR_100数据集上5种算法的损失函数和评估指标对比

Fig. 8 The comparison of training loss and evaluation indicators of five algorithms on CIFAR_100 dataset



(a) 训练集损失函数曲线



(b) 测试集评估指标

图9 COVID-19数据集上5种算法的损失函数和评估指标对比

Fig. 9 The comparison of training loss and evaluation indicators of five algorithms on COVID-19 dataset

由表2以及图6(b)、图7(b)、图8(b)、图9(b)可见,在上述4种数据集上的实验结果表明,SGD+T-ADEMA模型的各评估指标均优于传统EMA算法和主流Adam优化器,数据集训练越困难,T-

ADEMA算法就越有优势。此外,考虑到GPU内存(8GB)的限制,本实验选择的最大批量为32,不同训练批量可能会提高模型的预测性能。

表2 4种数据集上评估指标对比

Tab. 2 Comparison of evaluation indicators on the four datasets

算法	Accuracy			Precision			Recall			F ₁ Score		
	MNIST	CIFAR_10/ COVID-19		MNIST	CIFAR_10/ COVID-19		MNIST	CIFAR_10/ COVID-19		MNIST	CIFAR_10/ COVID-19	
		100	19		100	19		100	19		100	19
SGD	99.22	90.21/ 72.43	85.88	99.23	90.54/ 73.00	90.29	99.21	90.21/ 72.43	81.28	99.22	90.13/ 72.40	84.12
SGD+EMA	99.39	90.34/ 65.38	85.29	99.39	90.33/ 66.04	88.47	99.38	99.34/ 65.38	81.21	99.38	90.32/ 65.39	83.71
SGD+ADEMA	99.40	90.47/ 74.00	86.47	99.40	90.54/ 74.09	89.40	99.39	90.47/ 74.00	82.88	99.40	90.44/ 73.88	85.22
SGD+T-ADEMA	99.41	90.67/ 74.15	87.94	99.41	90.64/ 74.39	91.19	99.40	90.67/ 74.15	84.43	99.41	90.64/ 74.04	86.87
Adam	98.01	86.16/ 64.75	78.53	98.01	86.50/ 66.05	80.92	98.00	86.16/ 64.75	72.90	98.00	86.25/ 64.73	75.22

综上所述,动态衰减EMA算法比固定衰减的传统EMA算法更适合模型优化。其中,衰减系数决定了更新速度,衰减系数越大,网络收敛越稳定。当网络开始训练时,使用较小的衰减来确保初始学习的准确性,随着迭代次数的增加,衰减系数逐渐增大,可以有效地滤除噪声引起的无效学习,提高训练精度,使网络具有更好的收敛性。

4.2 T-ADEMA+SGD 算法实时性分析

为了分析本文模型的实时性,对5种优化算法在

4个数据集上的训练时间进行了比较,其结果见表3。由表3中数据可知,与其它算法相比,由于T-ADEMA算法使用Tanh函数计算移动平均,因此基于T-ADEMA+SGD网络每批训练时间增加约0.02s,比其它算法需要更高的计算成本。然而,SGD+T-ADEMA算法比主流的Adam算法需要更少的训练时间。虽然T-ADEMA算法的训练时间略高于传统算法,但可以有效地加快模型收敛速度,并在准确率、精度、召回率和F₁值等评价指标上取得较高的精度。

表3 4类数据集训练时间对比

Tab. 3 Comparison of training time on the four datasets

数据集	算法				
	SGD	SGD+EMA	SGD+ADEMA	SGD+T-ADEMA	Adam
MNIST	0.042±0.02	0.045±0.02	0.047±0.01	0.051±0.01	0.062±0.02
CIFAR_10	0.170±0.02	0.183±0.02	0.192±0.03	0.201±0.02	0.225±0.01
CIFAR_100	0.171±0.01	0.186±0.01	0.193±0.02	0.219±0.03	0.247±0.02
COVID-19	0.350±0.04	0.380±0.04	0.410±0.03	0.420±0.04	0.560±0.05

5 结束语

本文提出了基于动态衰减的T-ADEMA+SGD算法更新模型训练参数,在一定程度上提高了泛化性能,加快了模型训练收敛速度。为了验证算法的有效性,将基于T-ADEMA+SGD优化算法应用于ResNet50和ViT分类模型,用来优化各模型训练参

数。实验表明,基于T-ADEMA+SGD模型均能较好地提升各种样本库(如MNIST、CIFAR_10/100和COVID-19)的分类精度和泛化性能,对神经网络的训练和优化具有较好的启发意义。

参考文献

[1] AZADNIA R, KHEIRALIPOUR K. Recognition of leaves of different medicinal plant species using a robust image processing

- algorithm and artificial neural networks classifier[J]. Journal of Applied Research on Medicinal and Aromatic Plants, 2021, 25: 100327.
- [2] SOULI S, AMAMI R, YAHIA S B. A robust pathological voices recognition system based on DCNN and scattering transform[J]. Applied Acoustics, 2021, 177: 107854.
- [3] LINDA H, ALESSIA P, GIANCARLO P. Automated classification of cancer morphology from Italian pathology reports using Natural Language Processing techniques: A rule-based approach[J]. Journal of Biomedical Informatics, 2021, 116: 103712.
- [4] LÉCUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]// International Conference on Neural Information Processing Systems. Nevada, USA: Curran Associates Inc., 2012: 1097-1105.
- [6] TAKASE T, KARAKIDA R, ASOH H. Self-paced data augmentation for training neural networks[J]. Neurocomputing, 2021, 442: 96-306.
- [7] YUAN Jiantin, FAN Yangyang, LV Xiaoyi, et al. Research on the practical classification and privacy protection of CT images of parotid tumors based on ResNet50 model [C]//Journal of Physics: Conference Series, 4th International Conference on Artificial Intelligence, Automation and Control Technologies (AIACT 2020). Hangzhou, China: IOP Publishing, 2020, 1576(1): 012040.
- [8] SUN L, LIANG K, SONG Y, et al. An improved CNN-based apple appearance quality classification method with small samples [J]. IEEE Access, 2021, 9: 68054-68065.
- [9] WANG Ting, KNAP J. Stochastic gradient descent for semilinear elliptic equations with uncertainties[J]. Journal of Computational Physics, 2021, 426: 109945.
- [10] FAN Dongyan, SUN Hai, YAO Jun, et al. Well production forecasting based on ARIMA-LSTM model considering manual operations[J]. Energy, 2021, 220: 119708.
- [11] NGUYEN H D, TRAN K P, TRAN K D. The effect of measurement errors on the performance of the Exponentially Weighted Moving Average control charts for the Ratio of Two Normally Distributed Variables [J]. European Journal of Operational Research, 2021, 293(1): 203-218.
- [12] CHAN K, MUKHERJEE A, CHONG Z, et al. Distribution-free double exponentially and homogeneously weighted moving average leakage schemes with an application in monitoring exit rate [J]. Computers & Industrial Engineering, 2021, 161: 107370.
- [13] ZHANG Yan, SHAN Rui, WANG Huanpeng, et al. A new wavelet - neural network - ARIMA shares index combination forecast model [C]//International Conference on Automatic Control and Artificial Intelligence. Xiamen: IET, 2013: 199-201.
- [14] LI Chunxiang, CHI Ennan, LI Zhengnong, et al. Fast forecasting method for fluctuating wind velocity based on Extreme Learning machine[J]. Journal of Shanghai Jiao Tong University, 2016, 50(11): 1719-1723.
- [15] KADIR S N, TAHIR N M, YASSIN I M, et al. Mala-ysian tourism interest forecasting using Nonlinear Auto-Regressive Moving Average (NARMA) model [C]// Wireless Technology and Applications. Kota Kinabalu, Malaysia: IEEE, 2014: 193-198.
- [16] HUANG Jianglei, ZHOU Wengang. EMA: Regularized and reinitialized exponential moving average for target model update in object tracking [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(01): 8457-8464.
- [17] ZHENG Zeyu, GU Siyu. Tensorflow combat Google deep learning framework [M]. Beijing: Publishing House of Electronics Industry, 2017.
- [18] BOOB D, DEY S S, LAN G. Complexity of training ReLU neural network [J]. Discrete Optimization, 2022, 44 (part I): 100620.
- [19] ZHANG Kai, TANG Baoping, DENG Lei, et al. A hybrid attention improved ResNet based fault diagnosis method of wind turbines gearbox [J]. Measurement, 2021, 179: 109491.
- [20] CHEN Yuanyuan, YI Zhang. Adaptive sparse dropout: Learning the certainty and uncertainty in deep neural networks [J]. Neurocomputing, 2021, 450: 354-361.
- [21] HUANG Yanjiang, CHEN Kaibin, ZHANG Xianmin, et al. Motion estimation of elbow joint from sEMG using continuous wavelet transform and back propagation neural networks [J]. Biomedical Signal Processing and Control, 2021, 68: 102657.
- [22] OUCHICHA C, AMMOR O, MEKNASS M. CVDNet: A novel deep learning architecture for detection of coronavirus (COVID-19) from chest x-ray images [J]. Chaos, Solitons & Fractals, 2020, 140: 110245.
- [23] WU Qiufeng, CHEN Yiping, MENG Jun. DCGAN-based data augmentation for tomato leaf disease identification [J]. IEEE Access, 2020, 8: 98716-98728.
- [24] KHAN M A, JANI S P, SENTHIKUMAR A, et al. Machining parameter optimization using Adam - Gene Algorithm while turning lightweight composite using ceramic cutting tools [J]. International Journal of Lightweight Materials and Manufacture, 2021, 4(2): 262-267.