

基于神威高性能计算机的大规模有噪声图像训练系统

刘 锐 王 锐 周彧聪 刘 轶

北京航空航天大学计算机学院 北京市海淀区学院路37号 北京 100191

E-mail: lr-ruirui@163.com

摘要: 在深度学习应用中,大量且准确标注的训练数据是保证模型准确性的关键因素之一。在图像识别应用中,为减少人工标注工作量,常常采用从互联网抓取大量图像并自动标注的方法。这带来两方面的问题:首先,引入的错误标注数据将影响整体训练效果,其次,图像数量众多导致训练所需计算量庞大。本文在神威高性能计算机上设计实现了一种基于Caffe的大规模有噪声图像训练系统,该系统采用多进程并行和数据预取方法以充分发挥申威众核处理器和神威高性能计算机的性能优势,进而提升神经网络模型的训练速度,同时采用一种校准训练方法以降低噪声数据给模型训练带来的不利影响。实验数据表明,该系统可在保证训练效果的前提下,大幅减少训练所需时间,同时具有良好的性能可扩展性。

关键词: 图像识别;高性能计算;分布式训练;噪声标签

中图分类号: TP311 **文献标识码:** A

Large-scale noisy image training system based on Sunway high performance computer

LIU Rui, WANG Rui, ZHOU Yucong, LIU Yi

Sino-German Joint Software Institute, Beihang University, 37 Xue Yuan Road, Haidian, Beijing, 100191, China

Abstract: A precise large-scale dataset is crucial for supervising the training of deep neural networks (DNNs). However, manually annotating large-scale dataset is time-consuming and expensive. In practice, there are cheap but imperfect surrogates for automatically collecting and annotating massive images from the web, especially for images recognition tasks. This brings two problems: those annotations from the web are often inaccurate, and the large number of images involves large amount of computation on training. In this paper, we designed a large-scale noisy image training system based on Sunway TaihuLight supercomputer and implemented it using the Caffe framework. The system utilizes parallel processes as well as data prefetching to exploit computing power of the Sunway supercomputer. In addition, the system employs the mutual calibration training method to reduce the impact of noisy labels. Experimental results show that the system can greatly reduce the training time and has good scalability.

Key words: images recognition; high performance computing; distributed training; noisy labels

1 引言

近年来,深度学习技术在计算机视觉、自然语言处

理等领域取得了广泛的应用,例如在图像识别应用中,深度学习模型的识别准确率已超越了人类。另一方面,深度学习是一种计算和数据驱动的技术,神经网络模型的训练依赖大量的标注数据,数据的规模和标注质量在很大程度上决定着模型的识别效果。传统的深度学习训练数据依赖于人工标注,这需要消耗大量的人力 and 时间。针对这一问题,人们提出了多种改进方法,如半监督学习利用少量有标签数据混合无标签数据进行深度模型训练^[1];小样本学习(Few-Shot Learning)希望机器学习模型在学习了一定类别的大量数据后,对于新的类别,

基金项目: 自然科学基金重点项目(61732002)资助

作者简介: 刘锐,女,1989年生,博士研究生,研究方向为计算机系统结构、高性能计算,人工智能;王锐,男,1978年生,博士,副教授,研究方向为计算机系统结构、并行计算;周彧聪,男,1993年生,硕士研究生,研究方向为计算机视觉和机器学习;刘轶,1968年生,博士,博士生导师,教授,研究方向为高性能计算和计算机系统结构。

只需要少量的样本就能快速学习^[2]；自监督学习是无监督学习的一类特定方法，使用数据集本身的信息来构造伪标签^[3]；弱监督学习利用一部分标签带噪声的数据进行学习^[4]。

在图像识别应用中，用于训练的数据常常来源于从互联网抓取的大量图像，为减少人工标注工作量，人们使用查询信息或数据标签对图像进行自动标注，这带来两方面的问题：首先，使用数据标签对图像进行自动标注将不可避免地引入较多的噪声，即错误标注的图像，这些噪声数据将影响训练效果，进而降低推理阶段的图像识别准确率；其次，从互联网抓取的图像数量众多，因而训练所需计算量庞大，在典型的CPU+GPU平台上，训练时间常常达到数天^[5]。例如在^[6]中，使用Nvidia M40 GPU完成90-epoch ImageNet-1k的ResNet-50训练耗时14天。

针对上述问题，本文在神威高性能计算机上设计实现了一种基于Caffe^[7-8]的大规模有噪声图像训练系统，该系统采用多进程并行和数据预取方法以充分发挥申威众核处理器和神威高性能计算机的性能优势，进而提升神经网络模型的训练速度；同时，采用一种面向有噪声数据的校准学习方法^[9]，该方法使用一个大规模有噪声数据集和一个小规模人工标注数据集对两个模型进行训练，以解决噪声数据对模型训练造成的不利影响。实验数据表明，该系统可在保证训练效果的前提下，大幅减少训练所需时间，同时具有良好的可扩展性。

2 大规模有噪声图像的校准训练方法

2.1 校准训练方法

对于深度学习来说，如何获取数据无疑是一个关键的问题，其中获取大量的带精确标签的数据是一个非常耗时和昂贵的过程。一个常用的解决方案是通过搜索引擎获取大量图像，搜索关键字作为图像相应的标签。这种方法可以缩小数据获取的人工成本，但也不可避免的引入了一些噪声标签。文献^[10]表明复杂的模型参数使得深度学习模型不仅拥有学习能力，也具备了“记忆”能力。深度学习模型不仅可以从数据中学习到有用的特征表示，也可以记住噪声标签，而深度学习模型对噪声数据的过拟合会严重影响了模型的性能，因此，近年来基于噪声标签的深度学习图像分类算法成为研究热点^[11]。根据对噪声标签处理方式的不同，有噪声图像数据的训练方法可以分两类：基于噪声建模的方法和非噪声建模的方法。

基于噪声建模的方法旨在提出不同的噪声估计方法，通过对噪声数据的学习获取有用的参数，指导模型的训练^[12]。然而，在实际情况中噪声数据结构通常都比

较复杂，在缺少外部的辅助信息的情况下很难对噪声分布进行准确的估计。非噪声建模方法旨在提出不同的鲁棒学习算法，减轻噪声标注对模型性能的影响^[13]。例如，文献^[14]提出了一种损失函数修正方法，通过估计样本被错分到每个类别的概率构建一个损失函数修正矩阵，从而提高损失函数对噪声的鲁棒性。还有一类基于半监督学习的方法，通过在损失函数中添加与未标记数据相关的项，来鼓励模型举一反三，增加对陌生数据的泛化能力。这类训练方法一般比较复杂，标注降低了，但是训练复杂度等其他方面的代价上来了。

在噪声水平比较小的情况下，以上的噪声训练方法在一定程度上可以减少噪声数据对模型性能的影响，但对训练噪声水平高的数据集来说，由于可利用的干净数据比较少，神经网络模型可以学习到的信息十分有限，存在过拟合噪声数据的风险。根据我们的研究发现，虽然大规模的干净数据比较难获取，但是可以相对容易地通过人工标注的方式获取少量的干净数据的。因此，我们从另一个角度出发，通过引入少量干净数据校准大规模噪声数据训练，同时弥补上述方法在高水平噪声数据情况下的性能不足。

在这种场景下，如何合理利用干净数据校准模型对噪声数据的过拟合成为关键。知识提取方法基于迁移学习的思想，将模型在大数据集上学到的“知识”迁移到有一定相关性的目标任务上，目标任务的数据集一般比较小^[15]。对于噪声数据训练来说，知识提取方法先利用少量干净数据训练一个辅助模型，再将辅助模型生成软目标校准主模型在噪声数据上的学习^[16]。但是这种方法的效果很大程度上依赖干净数据的数量。为此，我们提出了一个相互校准训练方法。

结合知识提取方法，我们的方法同样采用双模型结构，同时训练一个辅助模型 M_a 和主模型 M_p ，校准训练方法的流程如下图1所示，整个训练过程可以分为三个阶段：

首先，我们进行一个模型预训练的阶段。在预训练阶段，我们不对噪声数据进行任何处理，基于干净数据集和噪声数据集，我们采用传统的训练方法在一个固定的大的学习率下分别训练辅助模型 M_a 和主模型 M_p 。根据文献^[17]表明，当训练学习率保持在一个大的值的时候，神经网络模型可以减少拟合噪声数据。因此，我们保持一个较大的学习率值不变，可以使得两个模型尽可能的学习一些干净数据的特征表示。

完成模型的预训练之后，我们再利用校准训练方法进一步提高模型的性能。我们的校准训练方法结合了噪声数据的清洗和知识提取的方法。噪声数据清洗方法的

思想在于经过预训练得到的模型已经具备一定的判别能力，因此模型对于样本预测的置信度在一定程度上可以近似地被看作对该样本进行标注正确的概率，置信度高的样本是干净数据的可能性更大。通过选择高置信度的样本，可以将潜在的干净数据从噪声数据集中筛选出输入模型进行训练，从而不断的提高模型的性能。

然而，这样做会带来一个问题，那就是认知偏差(Confirmation Bias)——如果模型在最初的时候把噪声样本误划入干净样本中，那么在之后的训练过程中它就会对该错误的样本过拟合，使得模型在错误的道路上越走越远。也就是说，单独的模型用这样的方法会变得过度“自信”，因此我们在这里需要用两个模型协同来进行数据清洗。我们根据两个模型的平均置信度选择样本，

而不是完全依赖模型自身的置信度，在某种程度上实现了两个模型之间的校准训练。

在知识提取方法中，一般利用在干净数据上训练得到的辅助模型生成软目标校准噪声数据的标签。然而辅助模型的性能依赖干净数据，当只有少量干净数据的时候，辅助模型也存在较大过拟合干净数据的危险。为了解决这一问题，我们在校准训练过程中，我们通过噪声清洗方法尽可能的选择一些置信度高的样本提供给辅助模型 M_a 训练，这样可以在保持模型干净的同时提高其精度。在噪声数据中采样一个批次样本的过程服从B次实验的二项分布，则：

$$M = \arg \max_m [\text{binocdf}(m, B, 1 - P_{\text{noise}}) < 1\%] \quad (1)$$

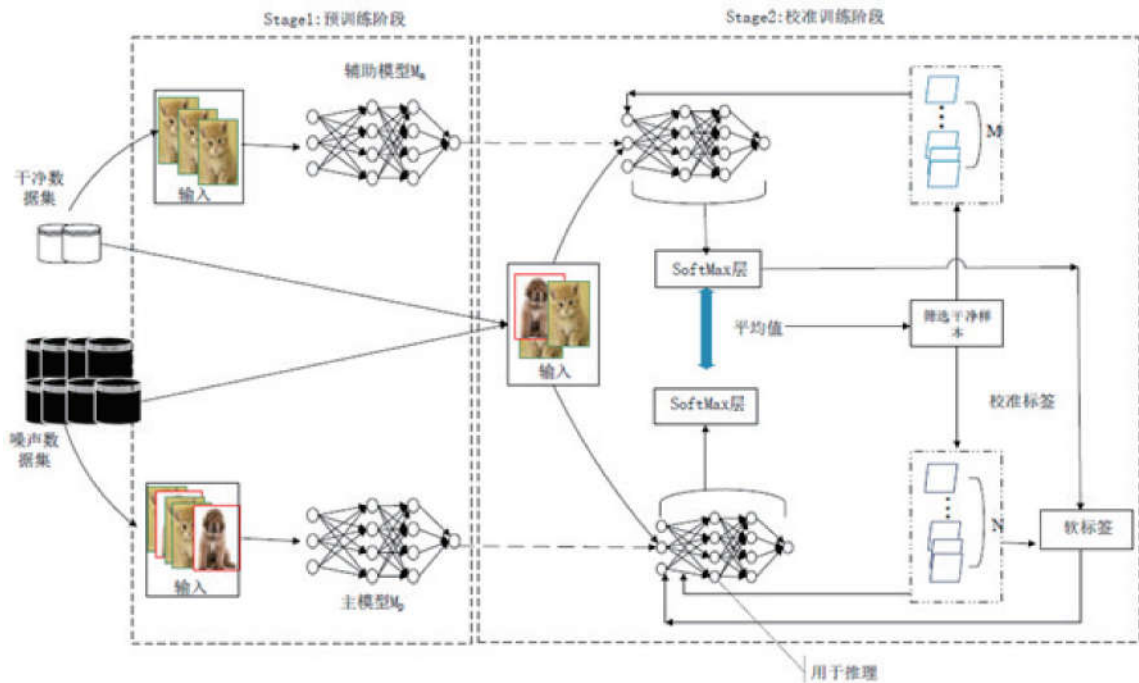


图1 校准训练方法示意图

Fig.1 The Mutual Calibration Training framework (MCT).

其中binocdf (n, B, P)函数表示成功概率为P的事件A发生次数不大于m次的概率，则在一个mini-batch的B个采样样本中，有99%以上的概率至少包含M个干净数据。

主模型在学习时候通常选择置信度更低的样本，这样设计的目的是让主模型能够在大规模的训练数据中选择一些难识别的难样本，进一步地利用噪声数据训练模型。虽然主模型选择低置信度的数据，难免会引入少量的噪声数据，但是辅助模型通过生成的软目标将在干净数据上学到的“知识”迁移给主模型，达到修正数据的错标签的效果。因此，我们可以将一个mini-batch中干净

样本数量的期望设置为样本采样的值N：

$$N = \text{Ceil}(B * (1 - P_{\text{noise}})) \quad (2)$$

最后，当两个模型完成校准训练过程后，辅助模型的知识已经通过迁移学习传递给了主模型，所以在推理阶段仅需要用主模型进行预测。

2.2 大规模有噪声图像训练的并行优化

一般来说，深度学习训练任务采用主流的CPU + GPU的异构组织形式。然而，从互联网抓取的图像数量众多，训练所需计算量庞大。例如，作为目前图像识别领域最权威的竞赛之一，Webvision竞赛利用ImageNet

1000个类的文本信息从网站上爬数据, Webvision数据集的数据主要来自Google和Flickr两个网站^[18]。其训练集由240万幅图片构成训练数据, 比ImageNet的两倍还多。传统的计算架构已经不能满足大规模训练数据的需求, 需要采用高性能的集群满足其计算需求。

“神威·太湖之光”是采用自主申威众核处理器构建的高性能计算平台, 双精度峰值性能125PFlops。神威太湖之光高性能计算系统共包含了40960个申威SW26010处理器, 申威SW26010

众核处理器采用异构架构, 每个处理芯片包含4个计算单元(Core Group, 核组), 每个核组包括一个管理单元MPE和一个8*8的运算阵列(Computing Processing Elements clusters, CPE)组成的加速部件; 每个核组拥有独立的8GB内存, 核组之间通过片上网络互联。

神威高性能平台的超强算力以及高效互联的系统结构为深度学习训练任务的并行优化问题提供了一种新的解决方案, 同时具备了解决大规模噪声图像训练问题的能力。因此, 如何设计好灵活高效的训练系统既能提高大规模噪声图像训练的性能, 又能充分发挥神威高性能平台的计算优势, 也成为本文的设计关键。

目前, 针对神经网络模型的并行优化研究可以分为数据并行、模型并行和混合并行三种方式^[20-22]。数据并行方案是指每个计算节点保持相同的模型副本, 通过对训练数据进行切分实现并行训练, 每个节点拥有不同的mini-batch输入。模型并行优化方案是针对深度学习模型的参数规模非常大, 单个计算节点的内存不能全部加载模型参数的情况, 将模型参数进行切分, 每个节点维护部分参数信息。混合并行是一种将数据并行和模型进行混合执行的方案。

针对大规模噪声图像的神经网络训练方法来说, 不同之处在于我们采用了双模型的训练框架, 在训练的过程中不仅涉及到模型内的参数同步更新, 还涉及到双模型间学习样本的选择过程的相互校准。因此, 为了减少分布式训练中的数据传输和同步的开销, 加速所提出方法的训练速度, 我们提出了混合并行的训练方案, 我们从两个角度对校准训练进行了并行优化, 从计算角度来说, 为了充分发挥神威众核架构的计算能力, 我们采用模型并行方案, 将两个模型部署在不同的计算节点上。从高效并行方面来讲, 针对大规模的训练数据, 我们在同一个模型内实行数据并行策略。

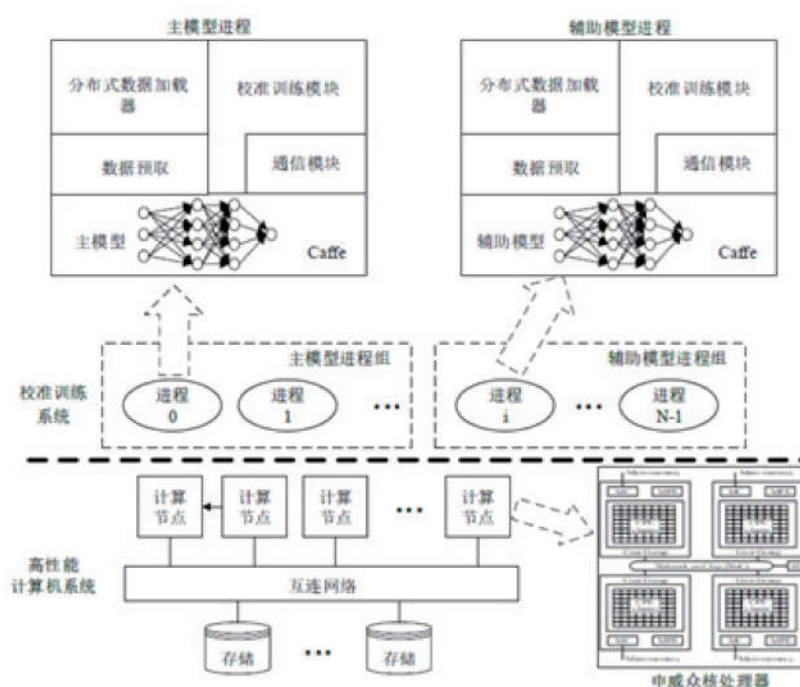


图2 基于神威高性能计算机的大规模有噪声图像训练系统架构图

Fig.2 The architecture of MCT system based on Sunway TaihuLight platform

3 训练系统的组成结构

结合神威高性能计算平台的硬件架构特点, 本文设计和实现了大规模有噪声图像分布式训练系统。系统架构如

图2所示。该系统利用深度学习框架Caffe, 在框架中实现了上节所述校准训练方法的主模型和辅助模型, 以核组为单位执行单个模型训练, 并通过MPI进程实现多节点的并

行训练,同时,为解决训练过程中读取海量图片带来的I/O性能瓶颈问题,系统专门设计了数据预取机制,在模型训练的同时进行下一批图像文件的预取,实现了训练计算与I/O重叠,进一步提升了系统的整体性能。

首先,本文对应用广泛的深度神经网络训练框架Caffe进行改进,利用MPI多进程机制实现Caffe分布式并行训练。每个MPI进程执行不同的模型训练,并负责模型间的同步和数据通信。与传统深度网络模型训练不同,本文设计的校准训练方法采用双模型结构,因此要确定不同计算节点的任务分工,限定参加通信的进程范围。考虑到不同阶段的训练任务不同,本文采用动态任务划分与映射策略。对于预训练阶段来说,两个模型的输入是不同的,同时由于干净数据和噪声数据的规模相差巨大,计算节点间需要进行的负载均衡设计,减少计算节点空闲等待的时间,本文根据干净数据和噪声数据的比例设置负责执行主模型和辅助模型训练的计算节点数量;对于校准训练阶段,两个模型具有相同的输入,为了减少模型间的同步等待时间应该分配同等数量的训练节点。

其次,本文对深度神经网络训练框架Caffe进行扩展实现分布式数据加载器。随着训练数据规模的增大会导致两个问题:加载数据时间过长和内存不足,因此边加载数据边训练,需要设计高效的分布式数据加载器。大规模图像训练不仅是计算密集型应用,也是I/O密集型任务。随着训练节点的增加,图片读取数量也成倍增加,导致文件系统的I/O访问压力增加,必然会引起加载数据时间过长。针对这个问题,本文采用数据预取技术减少数据加载时间,提高模型的训练速度。本文利用双线程进行数据加载,在训练当前mini-batch数据的同时将下一个mini-batch的数据提前读入内存中。同时,对于多个训练节点需要制定全局数据分配策略保证每个计算节点的数据互斥而完备。

对大规模噪声图像校准训练方法来说,整个训练过程不同于标准的深度神经网络操作,不完全是简单的数据并行训练。其特殊之处在于该方法采用校准方法进行样本的选择,即基于两个模型对同一个mini-batch输入预测平均置信度进行采样的操作。同时,考虑到分布式训练,每个计算节点维持不同的mini-batch输入,因此需要对全局的训练样本统计样本分布,根据整体的样本分布选择出来的样本更可信。为此,我们设计了一个分布式校准训练模块实现分布式采样。这个操作涉及到多个计算节点的同步和通信。

分布式校准训练模块可以分为四个步骤:首先,是对计算资源的划分,将全局通信组MPI_COMM_WORLD划分为两个不同的通信组,在两个通信组内分别执行主

模型和辅助模型。第二步,调用分布式数据加载器实现模型的数据并行训练,这样每个进程有一份对于mini-batch输入的预测置信度output,为了进行全局样本分布统计,两个模型通过集合通信Allgather完成所有数据的置信度output同步。第三步,所有进程对output进行排序,根据排序结果进行采样,这里采样结果以与输入数据mini-batch同样大小的0-1掩码形式存储。最后,辅助模型根据采样的样本执行正常的训练过程,同时将学习的知识迁移给主模型,即生成软目标供主模型学习训练。

4 实验

我们基于Caffe^[8]框架所有模型训练,并使用神威高性能系统的MPI并行接口实现噪声图像的分布式校准训练系统。我们进行两组实验以验证大规模噪声图像训练系统的正确性(精度)和性能。

实验使用两种经典的深度学习模型Alexnet、ResNet50,以及两个含有真实噪声数据的图像数据集Webvision^[18]和Clothing1M^[4]。Webvision数据集有1000个分类,这1000分类的数据是通过利用ImageNet的分类标签信息从Google和Flickr两个网站上抓取得到,包括240万幅图片构成训练集,50,000张的验证集和测试集。Clothing1M数据集的图像数据主要从Amazon、eBay以及淘宝网等购物网站抓取得到,共有14个分类,包含一百万张的噪声数据构成的训练集、1.4万张验证集和1万张测试集,以及5万张人工标注过的干净数据。

4.1 性能测试

在分布式训练过程中,影响模型训练的性能瓶颈主要表现在I/O读取和模型间的同步通信开销。针对I/O性能瓶颈,采用在模型训练的同时进行数据预取的方法以隐藏文件I/O延迟。我们测试了AlexNet模型在多个节点上一个mini-batch数据通过数据预取技术优化后的执行时间,如图3所示,可以看到使用多线程预取方法实现了模型训练的计算过程和I/O读取数据操作重叠,达到提升训练性能的效果。

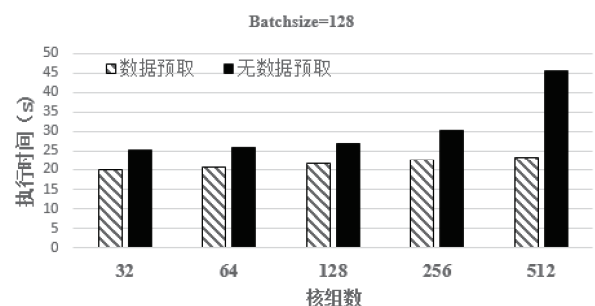


图3 AlexNet模型在不同节点下一个minibatch的执行时间
Fig.3 AlexNet's running time of one minibatch on various CGs

模型间的通信同步开销主要分为参数同步通信和校准训练方法的全局样本的预测置信度分布统计产生的通信。因此，我们在多个节点上进行了通信开销测试，如下图4所示。针对多个计算节点之间的模型参数同步，我们采用高效的RingAllreduce^[23]通信算法实现参数同步，RingAllreduce通信性能受到模型参数大小以及计算节点数的影响，因此我们测试了模型AlexNet和ResNet50在多个节点上的参数同步通信开销，相对来说，ResNet50模型的扩展性更好，这是因为AlexNet的参数规模更大，AlexNet模型拥有233MB的参数量，而Resnet50的参数量仅为103MB，随着节点的增多参数规模大的模型通信增多。对于校准训练方法的全局样本的预测置信度分布统计产生的通信开销主要依赖于样本的数量和计算节点数，因此我们将Resnet5模型的mini-batch设置为32，AlexNet的minibatch值设为128；可以看出全局采样通信开销与样本数据mini-batch成正比，然而全局样本采样的通信时间远小于参数通信的时间。

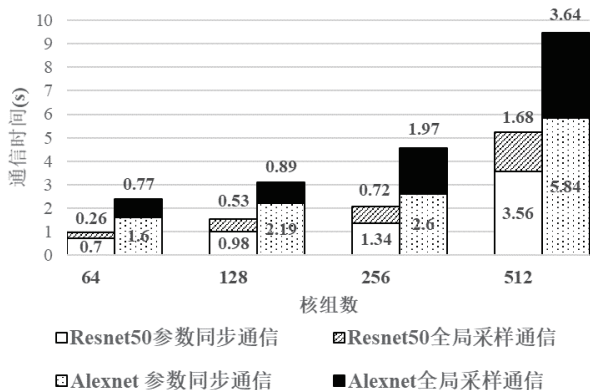


图4 AlexNet和Resnet50模型在不同节点上的通信开销
Fig. 4 Communication overhead of AlexNet and Resnet50 on various CGS

最后，我们对大规模图像噪声训练系统在神威高性能平台的多个计算节点上的整体训练性能进行评估，并与传统训练架构Intel Xeon CPU+GPU Titan X模式进行了对比。采用的GPU型号为NVIDIA Titan X，单精度浮点运算峰值性能为11 TFLOPs，而SW26010的单精度浮点峰值性能为3.03TFLOPs，性能是GPU的0.27X。在Clothing1M数据集上对扩展性能较差的AlexNet模型进行了评估，评估实验的mini-batch都设为128，共执行90 epoch训练，结果如下表1所示。

由于在训练过程中内存空间占用大于8GB，申威SW26010处理器需要使用内存大共享模型，因此我们的训练以核组为并行单位。实验结果显示，在对大规模数据进行训练时候，我们的训练系统可以很好的在多个节

点上进行扩展，大幅减少训练所需时间，同时具有良好的可扩展性。使用512个节点训练速度是单个Titan X GPU训练速度的9.6X。从峰值性能来说，单个GPU计算性能大约是一个核组的16X，然而我们的系统在扩展到32个节点时候才取得跟GPU对应的性能，这是由于多个节点上的系统训练性能还受到通信开销的影响。从图5也可以看出我们的训练系统的扩展性能接近线性。

表1 训练系统性能与GPU对比

Table 1 Performance of training system vs. GPU

平台	节点数(核组)	训练时间(小时)
GPU (Titan X)	1	150
神威	32	157.5
神威	64	81.1
神威	128	46.3
神威	256	26.3
神威	512	15.5

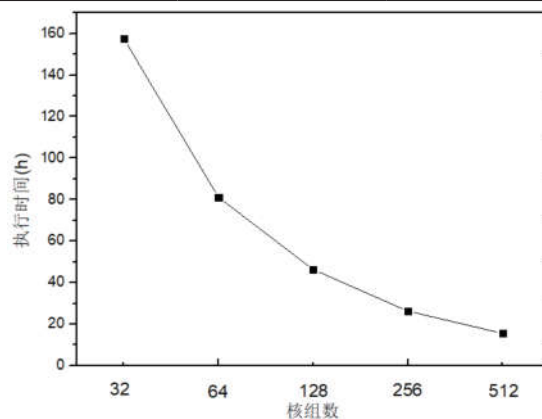


图5 系统在神威·太湖之光上的可扩展性测试

Fig.5 Scalability of the training system

4.2 正确性验证

正确性测试主要测试了ResNet50模型在真实噪声数据集Webvision和Cloting1M上使用校准训练方法的训练精度，同时也验证了在分布式校准训练系统在多个节点上的训练精度是否和单节点的训练精度一致。实验结果如表2所示，可以看到在两个数据集上，我们的校准训练精度都有所提升。在Clothing1M数据集上，跟Loss Correction方法相比，我们的校准训练方法提升了1.02个点，跟LossCorrection方法比，我们的方法不需要对噪声分布进行建模，可以更好的适应真实场景的噪声数据。在Webvision数据集上，我们的方法比CleanNet方法也提升了1.1个点。此外，我们在512个节点上验证了我们的分布式训练系统的正确性，跟单个GPU节点比，训练精度仅下降了0.1%和0.2%，属于正常波动范围，这证明了我们的训练系统保证了训练效果，具有很强的可靠性。

5 总结

针对深度学习应用中数据标注成本高, 以及数据规模大导致训练所需计算量庞大的问题, 本文在神威高性能计算机上设计实现了一种基于Caffe的大规模有噪声图像训练系统。该系统用Web噪声数据代替人工标注, 采用一种相互校准的训练方法降低噪声数据对模型性能的影响, 同时为了大幅减少训练所需时间, 本文采用多进程并行和数据预取方法实现分布式噪声图像训练系统, 充分发挥了申威众核处理器和神威高性能计算机的性能优势, 进而提升神经网络模型的训练速度。最后, 通过实验验证该系统可在保证训练效果的前提下, 大幅减少训练所需时间, 同时具有良好的性能可扩展性。

表2 训练系统的正确性验证

Table 2 Correctness verification of training system

数据集	方法	实验平台	精度
Clothing1M	Noise only	GPU	68.94
Clothing1M	Loss Correction[14]	GPU	80.18
Clothing1M	校准训练	GPU	81.20
Clothing1M	校准训练	神威 512个核组	81.08
Webvision	Noise only	GPU	68.53
Webvision	CleanNet[24]	GPU	70.3
Webvision	校准训练	GPU	71.29
Webvision	校准训练	神威 512个核组	71.4

References:

- [1] Li J, Socher R, Hoi S C H. DivideMix: Learning with Noisy Labels as Semi-supervised Learning[C]//International Conference on Learning Representations. 2019.
- [2] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning[C]//Advances in neural information processing systems, 2016, pages 3630–3638.
- [3] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]//International conference on machine learning. PMLR, 2020: 1597-1607.
- [4] Xiao T, Xia T, Yang Y, et al. Learning from massive noisy labeled data for image classification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 2691-2699.
- [5] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems, 2012, 25.
- [6] Murphy J. Deep learning benchmarks of NVIDIA tesla P100 PCIe, tesla K80, and tesla M40 GPUs[J]. 2017.
- [7] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding [C] // Proceedings of the 22nd ACM international conference on Multimedia, 2014: 675-678.
- [8] Li L, Fang J, Fu H, et al. swcaffe: A parallel framework for accelerating deep learning applications on sunway taihulight[C]//IEEE International Conference on Cluster Computing (CLUSTER). IEEE, 2018: 413-422.
- [9] Liu R, Liu Y, Wang R, et al. Mutual calibration training: Training deep neural networks with noisy labels using dual-models[J]. Computer Vision and Image Understanding, 2021, 212: 103277.
- [10] C. Zhang, S. Bengio, M. Hardt, B. Recht, O. Vinyals, Understanding deep learning requires rethinking generalization, arXiv preprint arXiv:1611.03530 (2016).
- [11] Algan G, Ulusoy I. Image classification with deep learning in the presence of noisy labels: A survey[J]. Knowledge-Based Systems, 2021, 215: 106771.
- [12] Yi K, Wu J. Probabilistic end-to-end noise correction for learning with noisy labels[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 7017-7025.
- [13] Jindal I, Nokleby M, Chen X. Learning deep networks from noisy labels with dropout regularization[C]//2016 IEEE 16th International Conference on Data Mining (ICDM). IEEE, 2016: 967-972.
- [14] Patrini G, Rozza A, Krishna Menon A, et al. Making deep neural networks robust to label noise: A loss correction approach[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1944-1952.
- [15] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network (2015)[J]. arXiv preprint arXiv:1503.02531, 2015, 2.
- [16] Li Y, Yang J, Song Y, et al. Learning from noisy labels with distillation[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 1910-1918.
- [17] Tanaka D, Ikami D, Yamasaki T, et al. Joint optimization framework for learning with noisy labels[C]//Proceedings of the IEEE conference on computer vision and