计算机科学
COMPUTER SCIENCE

基于特征手性的数据无关模型评估方法

苗壮, 季时鹏, 吴波, 付睿智, 崔浩然, 李阳

引用本文

苗壮, 季时鹏, 吴波, 付睿智, 崔浩然, 李阳. 基于特征手性的数据无关模型评估方法[J]. 计算机科学, 2024, 51(7): 337-344.

MIAO Zhuang, JI Shipeng, WU Bo, FU Ruizhi, CUI Haoran, LI Yang. Data-free Model Evaluation Method Based on Feature Chirality [J]. Computer Science, 2024, 51(7): 337-344.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230500179

基于特征手性的数据无关模型评估方法

苗 壮[1] 季时鹏[1] 吴 波[1] 付睿智[2] 崔浩然[1] 李 阳[1]
1 陆军工程大学指挥控制工程学院 南京 210007
2 中国人民解放军 32316 部队 乌鲁木齐 830000
(emiao_beyond@163.com)

摘 要 模型评估是评判卷积神经网络模型性能的重要手段,多用于卷积神经网络模型的设计、对比和应用过程。然而,现有的模型评估方法大多需要使用测试数据运行模型得到相关评估指标,当测试数据因隐私、版权与保密等原因无法获取时难以发挥作用。为了解决此问题,提出了一种数据无关的卷积神经网络模型评估方法,其利用特征手性的相关特性,通过计算卷积核之间的距离来确定模型的评估指标。所提方法利用不同卷积神经网络模型的性能表现与卷积核距离之间的负相关性,验证了在不使用测试数据的情况下,直接利用模型参数评估模型相对性能排名的可行性与有效性。对比实验表明,使用欧氏距离测度来评估 AlexNet,VGGNets,ResNets,EfficientNets 这 4 类包括 17 个卷积神经网络的模型精度时,该模型评估方法的盲评准确性高,能够较好地完成模型评估任务。

关键词:数据无关;模型评估;特征手性;距离测度;卷积核距离
中图分类号 TP391

Data-free Model Evaluation Method Based on Feature Chirality

MIAO Zhuang[1],JI Shipeng[1],WU Bo[1],FU Ruizhi[2],CUI Haoran[1] and LI Yang[1]
1 College of Command and Control Engineering,Army Engineering University of PLA,Nanjing 210007,China
2 Unit 32316 of PLA,Urumqi 830000,China

Abstract Evaluating the performance of convolutional neural network models is crucial,and model evaluation serves as a key component in the process,which is widely used in model design,comparison,and application. However,most existing model evaluation methods rely on running models on test data to obtain evaluation indexes,so these methods are unable to deal with the situation where testing data is difficult to obtain due to privacy,copyright,confidentiality,and other reasons. To address the problem,this paper proposes a novel method to model evaluation that does not require testing data,instead,it is based on feature chirality. The model evaluation obtains the evaluation indexes of the models by calculating the kernel distance of different models. The negative correlation between model performance and kernel distance is then used to analyze model parameters and obtain the relative performance ranking of different models without accessing any testing data. Experimental results show that when using Euclidean distance,the proposed blind evaluation method achieves the highest accuracy across seventeen classic CNNs,including AlexNet,VGGNets,ResNets and EfficientNets. Thus,this method is an effective and viable approach for model evaluation.

Keywords Data-free,Model evaluation,Feature chirality,Distance measure,Convolution kernel distance

1 引言

在过去的几十年里,随着深度学习技术的发展和大规模标注数据集的建立,深度学习技术在众多复杂任务中取得的效果已经超过人类的水平[1]。这使得深度学习方法在图像识别[2]、自然语言处理[3]、语音识别[4]等领域得到了广泛应用。

作为深度学习技术的重要组成部分,卷积神经网络模型一直是研究热点。为了评价不同模型的性能,研究者会使用一些评价指标对模型进行评估[5]。通用的评价指标包括准确率(Accuracy)、精确率(Precision)和召回率(Recall)等。特定任务的评价指标包括:目标检测领域的交并比[6](Intersection over Union,IoU)、检索任务和回归任务的平均精度(Average Precision,AP)、均值平均精度[7](mean Average Precision,mAP)等。针对无标注测试集,也有相关学者[8-10]提出了一些方法先对数据集进行分类再进行模型评估。尽管上述评估方法非常有效,能从多角度对模型的性能进行评估,但是它们都依赖于测试数据集,必须输入数据并运行模型计算得到评估结果。

到稿日期:2023-05-25 返修日期:2023-09-13
基金项目:江苏省自然科学基金(BK20200581)
This work was supported by the Natural Science Foundation of Jiangsu Province,China(BK20200581).
通信作者:李阳(solarleeon@outlook.com)

近年有许多学者发现恶意程序可以隐藏于卷积神经网络模型中，使用测试数据运行模型，可能对模型评估者产生安全威胁。例如，StegoNet[11]提出了最低有效位(Least Significant Bit，LSB)替换、弹性训练、值映射和符号映射4种方法，以嵌入恶意程序；腾讯[12]提出了类LSB隐写法，将恶意程序代码编码到模型参数的浮点数末尾；EvilModel[12]提出通过修改神经元来快速替换参数以嵌入恶意程序的方法。由此可见，恶意程序嵌入神经网络模型的手段多种多样，传统的依赖数据输入模型的评估方法在评估过程中极易受到恶意程序的攻击，存在严重的安全隐患。

此外，由于隐私或版权限制，有些测试数据不可获取。例如，一些组织公布了预训练深度学习模型[13-14]，但因为版权限制无法开放测试数据。同时，医疗和人脸数据因涉及隐私问题，也不能向公众与第三方机构开放。因此，如何在没有测试数据的情况下进行模型评估，成为了一个亟待解决的新课题。

针对以上问题，本文提出了一种数据无关(Data-Free)的模型评估方法。受特征手性[15]的启发，本文方法对卷积神经网络模型参数进行分析，得到性能评估指标，从而得出不同模型的相对性能排名。特征手性体现卷积神经网络模型参数间因模型训练而引起的相似度变化，该变化与模型性能有一定的相关性，可以在一定程度上代表模型的性能。有相关学者对模型参数因训练而引起的变化进行了研究，例如Xing等[16]发现，模型在预训练后，参数服从的分布由正态分布变为截断幂律分布，并且模型使用基于截断幂律分布的初始化能使模型具有更高的性能。

利用特征手性的性质，本文方法可以在没有测试数据的情况下直接对模型参数进行分析，得到该模型的相对性能，解决了没有测试数据无法进行评估的问题，避免了用测试数据直接运行模型产生的安全隐患。本文的主要贡献如下：

1)提出了一种数据无关的卷积神经网络模型评估方法。该方法无需使用测试数据，直接通过度量卷积核与翻转卷积核之间的距离，得到模型评估指标，从而进行模型性能评估。

2)针对本文提出的模型评估方法，使用相关系数比较了两大类的4种距离测度，选取了最适合的欧氏距离测度，验证了本文方法满足距离测度的使用条件。

3)本文方法在17个卷积神经网络模型上进行评估，评估结果与实际模型性能排名一致，验证了本文方法的有效性。

2 相关工作

2.1 无标注模型评估方法

模型评估是计算机视觉任务中不可或缺的部分。模型评估通常将带标注测试集输入到待评估模型中，得到模型评估指标[5]，以此判断模型的性能。但由于资源限制或其他原因，模型评估者经常面临没有足够的有标注测试集的情况。

为了在无标注测试数据集上评估模型性能，Chouldechova等[17]利用人脸嵌入相似度评分的参数进行贝叶斯建模，提出了一种系统评估新方法。该方法在没有测试数据身份注释的情况下，也能评估人脸识别系统准确率和种族偏见趋势。Chen等[18]提出了一种样本级无标签模型评估方法，用于预测无标签数据，称为评分预测(Scoring Your Prediction，SYP)。SYP考虑了图像质量因素，即低级的基于图像的特征(如模糊度)，再将基于模型的指标和基于图像的指标相结合来增强样本的代表性。Deng等[8]先将原始图像转换成各种形式来构建元数据集，再采用特征统计来捕捉样本数据集的分布，最终训练回归模型以预测模型性能。Guillory等[9]发现了分类器的预测概率可以用来估计该模型在各个场景下的性能，并在ImageNet上进行了详细的验证。Miao等[10]提出了一种新的基于K-means聚类的特征一致性对齐方法(K-means Clustering Based Feature Consistency Alignment，KCFCA)。该方法可以识别数据中的固定模式与结构，自动给数据分类而不依赖于标注。此外，Miao等[10]还开发了一个动态回归模型来捕捉分布变化和模型精度之间的关系。

虽然这些无标注模型评估方法可以在无标注测试集上评估模型性能，但仍然需要测试集。与之前的方法相比，本文方法不需要测试集，只需模型参数即可进行模型评估。

2.2 数据无关方法

近年来，许多数据无关的模型方法被提出，这些方法被广泛应用于模型压缩[19]、知识迁移[20]、知识蒸馏[21]等领域，其主要思想是不使用额外数据，借助已经训练好的模型辅助完成具体的目标任务。例如，Chen等[22]认为教师模型较高的精度除了来源于表达特征，还来自于它是一个判别分类器，故Chen等重用预训练的教师模型中的判别分类器进行学生推理，从而让学生模型在相同条件下能达到与教师模型相同的性能。Zhang等[23]针对联邦学习中的数据异构问题，利用无数据的知识蒸馏方法来挖掘局部模型中的知识，并将其输入到服务器中的全局模型来缓解模型聚合问题。并且Zhang等开发了定制的标签采样和类级集成，以最大限度地利用知识，从而间接地减小了客户端之间的分布差异。Chen等[19]提出了一种无需数据的深度学习模型压缩技术，将给定的待压缩模型视为判别器，使用生成模型生成的图片代替训练数据集来得到压缩后的模型。Yang等[20]提出了一种模型迁移方法，能将多种已有的预训练模型按照功能距离拆分成子网络再按照需要进行重组，以构建出针对下游任务高效且易用的模型，而无需训练数据进行重新训练。Fang等[21]引入对比学习目标，以鼓励合成样本与已有合成样本具有更大的多样性，并提出了一种从预训练模型中恢复训练数据的方法，从而无需真实数据样本。

尽管数据无关的方法众多，但是目前对用于模型评估的数据无关方法的研究仍然较少。

2.3 手性

手性指任何几何图形或点群在平面镜中的成像在理想情况下不能与自身重合[24]，即具有手性的物体与其镜像不能通过旋转和平移操作使得它们重合。手性是现实世界的一个基础规则，它在各个学科中都有重要研究对象并具有研究价值，如宇称不守恒[25-26]、气旋旋向[27]、有机分子的同分异构体[28]等。

近年来，有学者开始对深度学习领域的手性现象展开研究。Lin等[29]使用ResNet-50[30]网络根据图片是否翻转进行

分类训练,得到的网络具有较高的分类精度。该网络的分类精度表明数据分布存在视觉手性,即该数据集的图像翻转会导致数据集的数据分布变化,从而揭示了手性数据在翻转后出现的语义变化情况。Lin 将视觉手性定义为,任意一张图像 X 的交换残差 $E(X)$非零,则 X 存在视觉手性,即:

$$E(X)=|\boldsymbol{J}(\boldsymbol{T}(X))-\boldsymbol{T}(\boldsymbol{J}(X))|\neq 0 \tag{1}$$

其中,$\boldsymbol{T}$ 为翻转操作,$\boldsymbol{J}$ 为图像处理操作,如去马赛克、JPEG 压缩等。视觉手性是对数据集进行统计分析,从手性角度探究视觉数据的统计量如何随图像的镜像而改变,但并未研究卷积神经网络模型参数中的手性现象。

随后,特征手性[15]研究弥补了上述研究的空白。特征手性体现卷积神经网络模型经过训练后,其卷积核参数间相似度发生的统计变化,在手性角度直观显示模型训练"学习"到的知识对模型参数的影响。当一个卷积神经网络模型各个卷积层的交换残差均非零时,称该模型具有特征手性,即满足:

$$E(M)=\sum_{i=1}^{N}E(L_i)\neq 0 \tag{2}$$

其中,N 为模型 M 的卷积层数,$E(L_i)$为模型 M 的卷积层 i 的交换残差。

$$E(L_i)=|S(L_i)-S(T_r(L_i))| \tag{3}$$

其中,L_i 为模型 M 的第 i 个卷积层中的参数,其形状为$\langle B_i,C_i,H_i,W_i\rangle$,$B_i$ 是卷积核数,C_i 是通道维度,H_i 和 W_i 是空间维度,$S(\cdot)$为卷积核距离计算函数,$T_r(\cdot)$为训练模型操作。未经训练的模型,其卷积核参数仅受到随机初始化方法的影响,各个卷积核之间的距离即为随机参数之间的距离。而经过训练的模型,其卷积核参数之间的距离则会受到训练数据的影响。随着训练数据的输入,卷积核参数的分布会逐渐变成完成指定任务所需模型的参数分布,使模型的卷积核距离与模型的性能呈现一定的相关性[15]。基于上述特征手性及其性质的发现,本文通过计算卷积核距离,设计了一种数据无关的模型评估方法。

3　本文方法

针对卷积神经网络模型缺少测试数据时难以评估模型的问题,本文提出了一种使用待评估模型的参数计算模型评估指标,从而实现模型数据无关评估的方法。卷积核距离计算是得到模型评估指标的关键步骤,其计算方法的具体框架如图 1 所示。具体而言,本文方法首先对模型的卷积层对应的卷积核进行翻转操作,生成相应的翻转卷积核集合。通过这一步骤,可以得到一组与原始卷积核相对应的翻转卷积核。随后计算每个卷积层中原始卷积核集合与构建的翻转卷积核集合之间的距离,即计算集合间的距离,得到卷积核距离。这一距离度量的目的在于评估卷积核的相似性。通过计算得到的卷积核距离数值比较(见 4.2 节),我们能够观察到模型在不同卷积层上的卷积核相似度差异。进一步地,为了得到综合的评估指标,对模型各个卷积层的卷积核距离进行平均,以综合考虑各个卷积层对模型性能的贡献,并得到总体的模型评估指标。本章最后介绍如何选用合适的距离测度用于本文模型的评估方法。

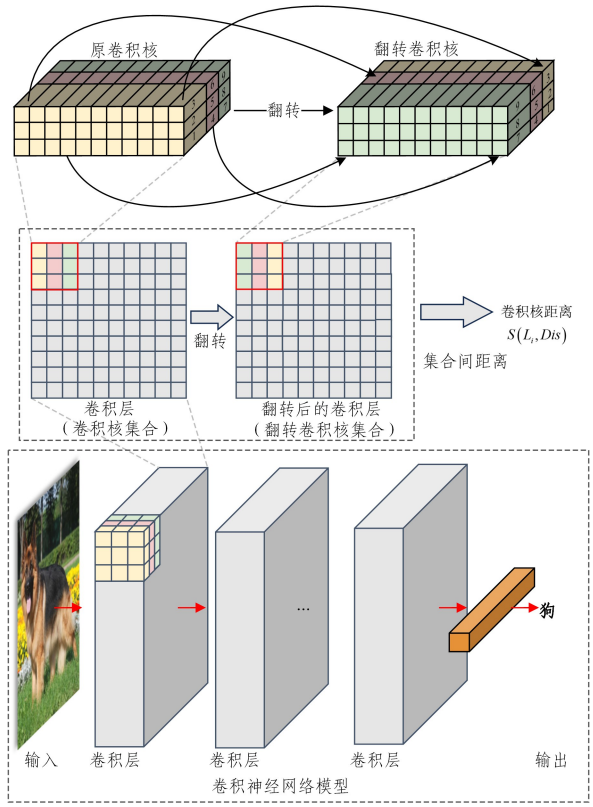


图 1　数据无关模型评估方法的框架图(电子版为彩图)

Fig. 1　Schematic diagram of data-free model evaluation methods

3.1　翻转卷积核集合构建

翻转卷积核是原卷积核切片翻转后的卷积核。相比卷积[31]和 ShuffleNet[32]分别是对特征图和模型结构设计的操作,本文提出的翻转卷积核是一种全新的操作符,它是对卷积核的操作。翻转卷积核通过以下步骤得到:首先将原卷积核按照空间维度切片,再以切片上的中间列卷积核参数为轴,翻转轴左右两侧的卷积核参数,最后按照原本的切片顺序拼接回去。如图 1 最上方所示,对于一个形状为$\langle 11,3,3\rangle$的卷积核,其轴为中间标记为粉色的卷积核参数,翻转轴左右两侧标记为黄色与绿色的卷积核参数,得到新的翻转卷积核,其形状保持$\langle 11,3,3\rangle$不变。但对于每一个$\langle 1,3,3\rangle$切片,都以中间列为轴发生了翻转。

获得翻转卷积核集合过程的形式化描述如下:对于卷积层 i 的卷积核集合 $K_i=\{K_{i,1},K_{i,2},\cdots,K_{i,j}\}$,按照式(4)翻转 K_i 的每个元素,得到翻转卷积核集合 $K_i'=\{K'_{i,1},K'_{i,2},\cdots,K'_{i,j}\}$,其中 $K_{i,j}$为卷积层 i 的第 j 个卷积核,$K'_{i,j}$为卷积层 i 的第 j 个翻转卷积核,$i=1,2,\cdots,N$,$j=1,2,\cdots,B_i$。

$$K'_{i,j}=Flip(K_{i,j}) \tag{4}$$

其中,$Flip(\cdot)$为对卷积核进行翻转的操作,如图 1 所示。

3.2　模型评估指标计算

现有的模型评估指标[5-7]通常是使用测试数据对模型进行评估得到,如准确率、精确率、召回率、平均精度和均值平均精度等。但是这些模型评估指标都难以在无测试数据的情况下对模型进行评估。针对此问题,一种合理的解决方法就是设计一种数据无关的模型评估方法。因此,

本文提出了一个新的模型评估指标 $S(M,Dis)$，该指标通过比较待评估模型的模型平均卷积核距离大小，来判断模型的相对性能。其中，M 为待评估模型，Dis 为 3.3 节所选出的最适合用于本文所提模型评估方法的距离测度。$S(M,Dis)$ 先翻转卷积核，计算模型 M 各卷积层的卷积核距离 $S(L_i,Dis)$（见式(5)），再用 $S(L_i,Dis)$ 取所有层的平均值计算得到（见式(6)）。

卷积核距离是卷积层中卷积核集合与其翻转卷积核集合之间的距离，用于量化卷积核参数在训练前后发生的距离不对称变化。对于具有特征手性的模型，其每一卷积层的卷积核距离均会在模型训练前后发生较大的差异，而这些差异在不同模型中亦表现出一定的不同，从而体现出不同模型学习的差异。如图 1 所示，模型 M 使用距离测度 Dis 时，卷积核距离 $S(L_i,Dis)$ 的具体计算步骤为：将计算得到的卷积核集合元素 $K_{i,j}$ 与翻转卷积核集合元素 $K'_{i,j'}$ 依次配对求距离并取平均值，如式(5)所示：

$$S(L_i,Dis)=\frac{\sum_{j'=1}^{B_i}\sum_{j=1}^{B_i}|Dis(K_{i,j},K'_{i,j'})|}{(B_i)^2} \tag{5}$$

其中，B_i 为模型 M 的卷积层 i 的卷积核数，$Dis(\cdot)$ 为距离测度，用于计算元素之间的距离。本文使用 3.3 节所述方法选择距离测度 Dis 后，即可利用 $S(L_i,Dis)$ 得到模型评估指标 $S(M,Dis)$。

模型评估指标 $S(M,Dis)$ 是使用距离测度 Dis 计算一个模型的所有卷积层的卷积核距离 $S(L_i,Dis)$ 的均值得到，如式(6)所示：

$$S(M,Dis)=\frac{1}{N}\sum_{i=1}^{N}S(L_i,Dis) \tag{6}$$

3.3 距离测度选择

距离测度是常用的信息测度方法之一[33-34]，主要用于测量不同数据之间的距离，可以极为直观地显示不同数据之间的距离，以供后续分析使用。

本文使用变异系数(Coefficient of Variation)[35]寻找适用于本文提出的模型评估方法的距离测度。变异系数是一种衡量数据中各观测值变异程度的统计量，可以有效避免各距离测度在模型上的标准差和均值差异过大而导致无法比较的问题。

本文所用距离测度的变异系数的计算步骤如下：首先选定距离测度 Dis_b，$b\in\{1,2,\cdots,4\}$，计算各模型的评估指标 $S(M_a,Dis_b)$（M_a 为本文所研究的 17 个模型之一，$a\in\{1,2,\cdots,17\}$），从而得到模型评估指标集合 $Index_{Dis_b}=\{S(M_1,Dis_b),S(M_2,Dis_b),\cdots,S(M_{17},Dis_b)\}$。其次使用模型评估指标集合 $Index_{Dis_b}$，计算变异系数 $Cov(Dis_b)$。变异系数的计算式如下：

$$Cov(Dis_b)=\frac{Var(Index_{Dis_b})}{Mean(Index_{Dis_b})} \tag{7}$$

其中，$Var(Index_{Dis_b})$ 为距离测度标准差，$Mean(Index_{Dis_b})$ 为距离测度均值。最后，更换距离测度并重复上述步骤，得到 4 个距离测度的变异系数。

本文提出的模型评估方法使用的距离测度的变异系数越大，越能有效评估不同模型的相对性能[36]。根据本文的实验结果（见 4.2 节），欧氏距离测度[37]能较好地用于本文提出的模型评估方法。

欧氏距离测度(Euclidean Distance)用来标记多维空间中两个点之间的真实距离。n 维空间中两个点 x 与 y 的欧氏距离测度定义为：

$$d(x,y):=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \tag{8}$$

4 实验结果与分析

本文针对提出的数据无关的卷积神经网络模型评估方法，在 17 个模型（包括 AlexNet[38]、4 种 VGGNet[39]、4 种 ResNet、8 种 EfficientNet[40]）上利用变异系数对 4 种距离测度进行有效性评估。根据变异系数的评估结果，选择最佳的距离测度，并对其评估正确性进行验证。

4.1 实验设置

本文实验基于 PyTorch 在 ImageNet-1K 上预训练的 17 个模型中进行，包括 AlexNet，VGGNet-11，VGGNet-13，VGGNet-16，VGGNet-19，ResNet-34，ResNet-50，ResNet-101，ResNet-152，EfficientNet-B0，EfficientNet-B1，EfficientNet-B2，EfficientNet-B3，EfficientNet-B4，EfficientNet-B5，EfficientNet-B6 和 EfficientNet-B7。这些模型的实际性能数据来源于 PyTorch 官方[1)]。由于翻转空间维度为〈1,1〉的卷积核无变化意义，因此本实验只分析空间维度至少为〈3,3〉的卷积核，同时这也使得模型中各个卷积层的输出通道数在同一个阶段中保持一致，以减少分析干扰。

4.2 不同距离测度实验结果与分析

本文方法在上述 17 个经典模型上计算 4 种不同距离测度的变异系数。表 1 列出了本文方法使用不同距离测度的模型评估指标及其变异系数。其中各个模型对应于距离测度的数值是本文方法使用该距离测度时评估得到的模型评估指标。可根据表 1 中的模型评估指标大小来判断模型相对性能大小，模型评估指标越小，判定模型性能越好。同时，依据表 1 中各距离测度的模型评估指标计算的变异系数，来判断最适合用于本文模型评估方法的变异系数。根据距离测度的计算原理，可将 4 种距离测度分成两类：基于向量关系的方法和基于绝对距离的方法。基于向量关系的方法包括余弦距离测度和皮尔逊相关系数，基于绝对距离的方法包括欧氏距离测度和切比雪夫距离测度。

由表 1 可知，基于向量关系的方法中，2 种距离测度的变异系数均较小，因此这两种距离测度不适合用于本文方法。基于绝对距离的方法的变异系数总体上均优于基于向量关系的方法，其中切比雪夫距离测度的变异系数最大，欧氏距离测度次之，故基于绝对距离的方法最适合用于本文提出的模型评估方法。

1) https://pytorch.org/vision/stable/models.html

表 1　各距离测度的模型评估指标与变异系数

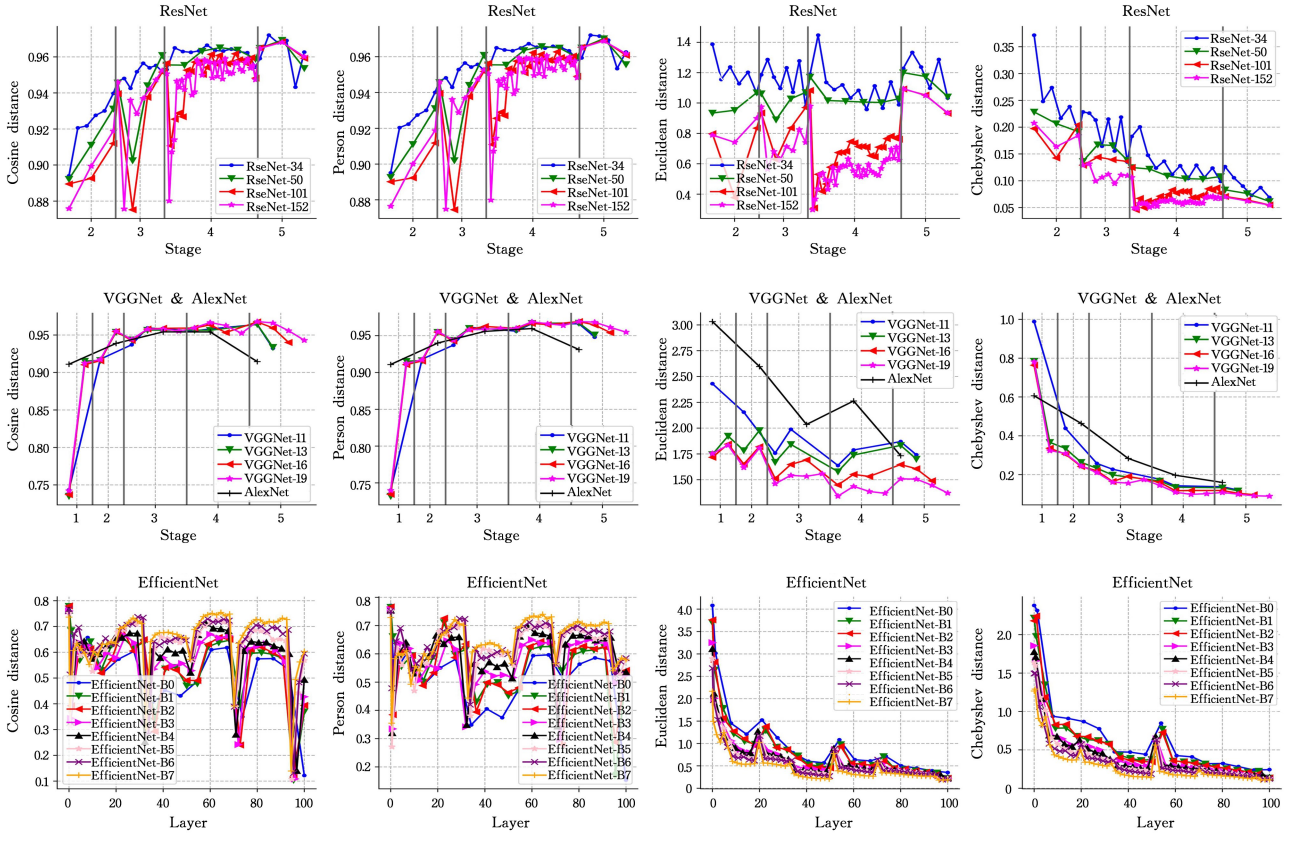
Table 1　Model evaluation indexes and coefficients of variation of distances

模型	余弦距离测度	皮尔逊相关系数	欧氏距离测度	切比雪夫距离测度
AlexNet	0.9344	0.9395	2.3319	0.3417
VGG-11	0.9209	0.9242	1.9208	0.3108
VGG-13	0.9230	0.9265	1.7790	0.2729
VGG-16	0.9323	0.9353	1.6274	0.2257
VGG-19	0.9385	0.9411	1.5292	0.1999
ResNet-34	0.9518	0.9529	1.1544	0.1637
ResNet-50	0.9461	0.9464	1.0409	0.1328
ResNet-101	0.9440	0.9446	0.7235	0.0904
ResNet-152	0.9432	0.9442	0.6315	0.0771
EfficientNet-B0	0.5182	0.5123	1.1236	0.7440
EfficientNet-B1	0.5518	0.5620	0.9218	0.6088
EfficientNet-B2	0.5369	0.5474	0.9177	0.6097
EfficientNet-B3	0.5522	0.5641	0.7401	0.4874
EfficientNet-B4	0.5853	0.5953	0.7095	0.4543
EfficientNet-B5	0.6135	0.6185	0.6436	0.4093
EfficientNet-B6	0.6382	0.6385	0.5946	0.3660
EfficientNet-B7	0.6481	0.6430	0.4814	0.3047
均值	0.7693	0.7727	1.1101	0.3411
标准差	0.1808	0.1794	0.5244	0.1868
变异系数	0.2351	0.2322	0.4724	**0.5476**

注：粗体表示最优，下划线表示次优。

为验证上述结论，找到最适用于本文方法的距离测度，并直观可视化各个距离测度在模型上的区别。图 2 给出了不同距离测度对卷积核距离的影响，横坐标表示按照阶段划分的卷积层，纵坐标表示卷积核距离。由于不同模型的卷积层数不一致，为了便于寻找模型参数平均距离变化趋势，本实验将 ResNet，VGGNet 和 AlexNet 的结果按照 ResNet 的 Stage-Block（阶段-块）设计对卷积层进行划分，并在每个阶段上都横向均匀拉伸所有模型的卷积核距离结果以实现对齐效果；将 EfficientNet 的各个模型卷积层按照其卷积层号与该模型的卷积层总数的比例进行均匀拉伸，结果如图 2 所示。例如，ResNet-50 和 ResNet-152 在阶段 4 的空间维度为〈3，3〉的卷积层数是 6 和 36，将这两个模型在阶段 4 的卷积核距离结果在图 2 中水平均匀排列；EfficientNet-B0 和 EfficientNet-B5 的空间维度为〈3，3〉的卷积层数是 17 和 40，将这两个模型的所有卷积核距离结果在图 2 上水平均匀排列，因此横坐标为卷积层数的百分比。

图 2(a)和图 2(b)显示，使用余弦距离测度和皮尔逊相关系数时，模型评估方法在上述 17 个模型上得到的结果基本一致，AlexNet，VGGNet 系列和 ResNet 系列卷积核距离均随着卷积层的增加而呈现上升趋势，EfficientNet 系列模型的卷积核距离随着卷积层数的增加呈现阶段式上升与下降。虽然使用余弦距离测度的结果相比使用皮尔逊相关系数的结果能更明显地区分不同模型，但使用这两种距离测度的模型评估方法得到的卷积核距离折线重合均较为严重，难以对不同模型进行评估。



(a)余弦距离测度　(b)皮尔逊相关系数　(c)欧氏距离测度　(d)切比雪夫距离测度

图 2　不同距离测度对卷积核距离的影响

Fig. 2 Influence of different distance on convolution kernel distance

图 2(c)和图 2(d)显示，使用欧氏距离测度和切比雪夫距离测度时，模型评估方法的结果有一定差异，使用欧氏距离测度的模型评估方法的卷积核距离在各个模型上随着卷积层的增加而呈现下降趋势，同时能最明显地区分不同模型之间

的卷积核距离变化折线；而使用切比雪夫距离测度的结果在最后的卷积层呈现较小的下降趋势。因此，欧氏距离测度最适合用于本文提出的模型评估方法。

选定距离测度后，依据表 1 中的模型评估指标来评估模型的相对性能，以验证本文方法使用欧氏距离测度时的有效性。将表 1 得到的模型评估指标按照其大小进行排序，得到模型相对性能顺序，结果如表 2、表 3 所列。例如，对于 ResNet 系列模型，其模型评估指标越小，其图像识别精度越高。另外，由于 EfficientNet 系列模型使用了深度可分离卷积，使得在部分卷积层上难以体现三维翻转卷积核的作用，因此 EfficientNet 系列模型不能直接与 AlexNet，VGGNet 系列和 ResNet 系列进行比较。

表 2 使用各距离测度评估 AlexNet，VGGNet 系列与 ResNet 系列各模型的相对性能

Table 2 Evaluation of relative performance of AlexNet, VGGNets, and ResNets models using various distances

距离测度	AlexNet	VGGNet-11	VGGNet-13	VGGNet-16	VGGNet-19	ResNet-34	ResNet-50	ResNet-101	ResNet-152
余弦距离测度	4	1	2	3	5	9	8	7	6
皮尔逊相关系数	4	1	2	3	5	9	8	7	6
欧氏距离测度	9	8	7	6	5	4	3	2	1
切比雪夫距离测度	9	8	7	6	5	4	3	2	1
实际性能排名	9	8	7	6	5	4	3	2	1

注：下划线表示评估模型相对性能排名结果错误。

表 3 使用各距离测度评估 EfficientNet 系列模型相对性能

Table 3 Evaluation of relative performance of EfficientNets using various distances

距离测度	EfficientNet-B0	EfficientNet-B1	EfficientNet-B2	EfficientNet-B3	EfficientNet-B4	EfficientNet-B5	EfficientNet-B6	EfficientNet-B7
余弦距离测度	8	6	7	5	4	3	2	1
皮尔逊相关系数	8	6	7	5	4	3	2	1
欧氏距离测度	8	7	6	5	4	3	2	1
切比雪夫距离测度	8	6	7	5	4	3	2	1
实际性能排名	8	7	6	5	4	3	2	1

注：下划线表示评估模型相对性能排名结果错误。

由表 2 可知，基于向量关系的方法评估结果一致，在 VGGNet 系列模型上能逆序地给出正确的模型性能排名，并正序地给出正确的 ResNet 系列模型性能排名，然而并不能正确地评估不同系列模型之间的相对性能排名。例如，使用余弦距离测度时，评估出 AlexNet 的精度排名高于 VGGNet-13，与实际不符；评估出 ResNet-152 的精度排名高于 ResNet-34，与实际相符。而本文方法使用基于绝对距离的方法则能得到正确的评估结果。

由表 3 可知，基于向量关系的方法的评估结果与使用切比雪夫距离测度的结果一致，均未能正确地给出 EfficientNet-B1 与 EfficientNet-B2 之间的模型性能排名，而使用欧氏距离测度得到的评估结果与实际性能排名一致。

本文方法使用切比雪夫距离测度时，有两个错误的评估结果；使用欧氏距离测度时，评估结果与正确结果完全一致；使用基于向量关系的两种方法时，均不能评估不同系列模型的结果，在 EfficientNet 系列上也有两个错误的评估结果。因此，基于向量关系的方法在同系列模型中能够较为准确地评估出模型的性能，但不能给出正确的不同系列模型间的模型性能排名。而基于绝对距离的方法优于基于向量关系的方法，能给出基本正确的结果，其中欧氏距离测度的评估准确性最高。

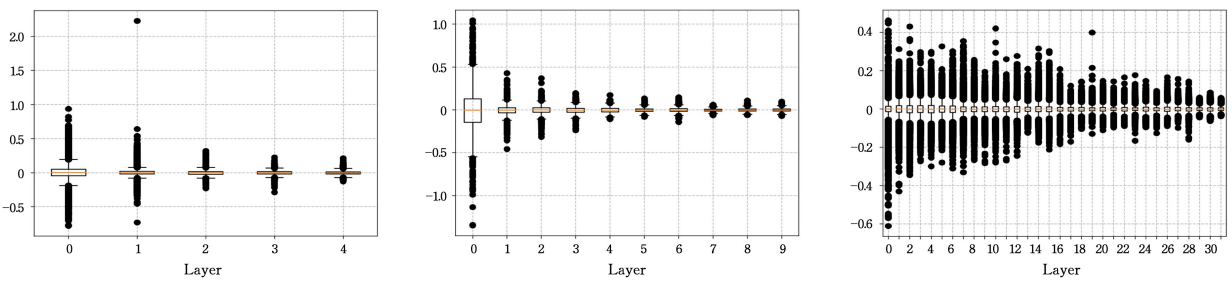
欧氏距离测度比其他距离测度更优的原因可能有如下两点：

1）欧氏距离测度由于其计算方式是通过平方和再开方得到的，因此其使用非常直观，符合距离的朴素理念。同时，欧氏距离测度适用于大多数数据，尤其是本文研究的模型参数值均为实数。

2）欧氏距离会对极端离群点数据分配更大的权值。因此，对于本文研究的模型参数，其数据量较大，数据采样较多，能更好地体现数据间的差异。

4.3 条件验证

本文使用了 4 种距离测度，其中 2 种具有使用条件。为验证本文方法满足距离测度的使用条件，本文设置了补充实验。将本文所用卷积神经网络模型的卷积核参数数据绘制成不同模型卷积层参数差异箱型，能显示模型各个卷积层内参数间的差异与取值范围，如图 3 所示。



(a) AlexNet (b) VGGNet-13 (c) ResNet-34

图 3 不同模型卷积层参数差异箱型图

Fig. 3 Boxplots of convolutional layer parameters differences of different models

图 3 中，不同模型的不同卷积层内卷积核参数差异不大，基本保持在(−1,1)范围内。在使用皮尔逊相关系数测度和欧氏距离测度时，要求数据必须大致符合正态分布且不能有极端异常值，各坐标轴上的数据数值不能差异过大。图 3 显示，本文所用卷积神经网络模型卷积核参数满足上述距离测度。

图 3 中，不同模型的不同卷积层内卷积核参数差异不大的主要原因是，卷积神经网络模型会使用多种方法限制卷积层的输出，从而减小输入到下一个卷积层的特征图内的参数差异，进而限制了学习得到的卷积核参数在维度之间的差异。例如，批标准化(Batch Normalization，BN)限制输入特征图满足均值为 0 且方差为 1 的分布，平均池化限制其领域内特征点的差异等。

结束语　针对缺乏测试数据而无法评估卷积神经网络模型的问题，本文提出了一种新的数据无关的模型性能评估方法。该方法通过比较模型卷积核距离，从特征手性的角度判断模型的相对性能。本文使用距离测度的变异系数分析发现：使用欧氏距离测度时的评估效果最佳，它能最显著地区分不同模型之间的相对性能。本文验证了所提方法在 17 个模型上的评估准确性。

下一步工作将进一步探索深度学习领域中的手性现象，并设计更精细的集合间距离测度，能在零样本的情境下对其他更先进和复杂的卷积神经网络模型进行准确评估，同时提升模型的评估速度。

参 考 文 献

[1] CHENG K Y，WANG N，SHI W X，et al. Research advances in the interpretability of deep learning[J]. Journal of Computer Research and Development，2020，57(6)：1208-1217.

[2] SHAFIQ M，GU Z Q. Deep residual learning for image recognition：a survey[J]. Applied Sciences，2022，12(18)：8972.

[3] OTTER D W，MEDINA J R，KALITA J K. A survey of the usages of deep learning for natural language processing[J]. IEEE Transactions on Neural Networks andLearning Systems，2020，32(2)：604-624.

[4] MALIK M，MALIK M K，MEHMOOD K，et al. Automatic speech recognition：a survey[J]. Multimedia Tools and Applications，2021，80：9411-9457.

[5] HU K，WENG C H，ZHANG Y W，et al. An overview of underwater vision enhancement：from traditional methods to recent deep learning[J]. Journal of Marine Science and Engineering，2022，10(2)：241.

[6] GU X J，ZHU J F，REN S G，et al. Multi-scale U network realizes segmentation and recognition of tomato leaf disease[J]. Computer Science，2021，48(S2)：360-366.

[7] GU G，XU X D. Real-time vehicle detection and information recognition technology based on YOLOv3 improved algorithm[J]. Computer Engineering and Applications，2020，56(22)：173-184.

[8] DENG W J，ZHENG L. Are labels always necessary for classifier accuracy evaluation? [C]// 2021 IEEE Conference on Computer Vision and Pattern Recognition. 2021：15069-15078.

[9] GUILLORY D，SHANKAR V，EBRAHIMI S，et al. Predicting with confidence on unseen distributions[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021：1134-1144.

[10] MIAO S Y，ZHENG L，LIU J J. K-means clustering based feature consistency alignment for label-free model evaluation[J]. arXiv：2304. 09758，2023.

[11] LIU T，LIU Z H，LIU Q，et al. StegoNet：Turn deep neural network into a stegomalware[C]// Annual Computer Security Applications Conference. 2020：928-938.

[12] WANG Z，LIU C G，CUI X. EvilModel：hiding malware inside of neural network models[C]// 2021 IEEE Symposium on Computers and Communications. IEEE，2021：1-7.

[13] ZHANG X G，QIN H T，DING Y F，et al. Diversifying sample generation for accurate data-free quantization[C]// 2021 IEEE Conference on Computer Vision and Pattern Recognition. 2021：15658-15667.

[14] CAI Y H，YAO Z W，DONG Z，et al. Zeroq：A novel zero shot quantization framework[C]// 2020 IEEE Conference on Computer Vision and Pattern Recognition. 2020：13169-13178.

[15] JI S P，LI Y，FU R Z，et al. Feature chirality in deep learning models[C]// 2023 International Conference on Intelligent Perception and Computer Vision. 2023：19-21.

[16] XING T T，SUN R C，SHAO F J，et al. Research on weight initialization method in deep learning[J]. Computer Engineering，2022，48(7)：104-113.

[17] CHOULDECHOVA A，DENG S，WANG Y，et al. Unsupervised and semi-supervised bias benchmarking in face recognition[C]// European Conference on Computer Vision. 2022：289-306.

[18] CHEN Y，ZHANG S，SONG R. Scoring your prediction on unseen data[C]// 2023 IEEE Conference on Computer Vision and Pattern Recognition. 2023：3278-3287.

[19] CHEN H T，WANG Y H，XU C，et al. Data-free learning of student networks[C]// 2019 IEEE Conference on Computer Vision and Pattern Recognition. 2019：3514-3522.

[20] YANG X Y，ZHOU D Q，LIU S H，et al. Deep model reassembly[J]. Advances in Neural Information Processing Systems，2022，35：25739-25753.

[21] FANG G F，SONG J，WANG X C，et al. Contrastive model inversion for data-free knowledge distillation[J]. arXiv：2105. 08584，2021.

[22] CHEN D，MEI J P，ZHANG H，et al. Knowledge distillation with the reused teacher classifier[C]// 2022 IEEE Conference on Computer Vision and Pattern Recognition. 2022：11933-11942.

[23] ZHANG L，SHEN L，DING L，et al. Fine-tuning global model via data-free knowledge distillation for non-IID federated learning[C]// 2022 IEEE Conference on Computer Vision and Pattern Recognition. 2022：10174-10183.

[24] KELVIN W T B. The molecular tactics of a crystal[M]. Clarendon Press，1894.

[25] LEE T D，YANG C N. Question of parity conservation in weak interactions[J]. Physical Review，1956，104(1)：254.

[26] WU C S，AMBLER E，HAYWARD R W，et al. Experimental test of parity conservation in beta decay[J]. Physical Review，1957，105(4)：1413.

[27] KUMAR K R, SINGH B B, KUMAR K N. Intriguing aspects of Asian Summer monsoon anticyclone ozone variability from microwave limb sounder measurements [J]. Atmospheric Research, 2021, 253: 105479.

[28] NI B, CÖLFEN H. Chirality communications between inorganic and organic compounds[J]. Smart Mat, 2021, 2(1): 17-32.

[29] LIN Z Q, SUN J, DAVIS A, et al. Visual chirality [C]// 2020 IEEE Conference on Computer Vision and Pattern Recognition. 2020: 12295-12303.

[30] 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. 2016: 770-778.

[31] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[32] ZHANG X Y, ZHOU X Y, LIN M X, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices [C]// 2018 IEEE Conference on Computer Vision and Pattern Recognition. 2018: 6848-6856.

[33] XU Z S, CHEN J. An overview of distance and similarity measures of intuitionistic fuzzy sets[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2008, 16(4): 529-555.

[34] LIAO H C, XU Z S, ZENG X J. Distance and similarity measures for hesitant fuzzy linguistic term sets and their application in multi-criteria decision making[J]. Information Sciences, 2014, 271: 125-142.

[35] MURTAZA G, ALI H, BADSHAH N. A robust local model for segmentation based on coefficient of variation[J]. International Journal of Information and Communication Technology, 2011, 5: 30-39.

[36] MORA M, TAUBER C, BATATIA H. Robust level set for heart cavities detection in ultrasound images[C]// Computers in Cardiology. IEEE, 2005: 235-238.

[37] WU H, CAO Y, WEI H, et al. Face recognition based on Haar like and Euclidean distance[J]. Journal of Physics: Conference Series, 2021, 1813(1): 012036.

[38] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[39] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409. 1556, 2014.

[40] TAN M X, LE Q. EfficientNet: Rethinking model scaling for convolutional neural networks[C]// International Conference on Machine Learning. PMLR, 2019: 6105-6114.

MIAO Zhuang, born in 1976, Ph.D, professor, Ph.D supervisor. His main research interests include artificial intelligence, pattern recognition and computer vision.

LI Yang, born in 1984, Ph.D, associate professor, is a senior member of CCF (No. D24215). His main research interests include computer vision, deep learning and image processing.

（责任编辑：喻藜）