计算机科学
COMPUTER SCIENCE

基于主动采样的深度鲁棒神经网络学习

周慧，施皓晨，屠要峰，黄圣君

引用本文

周慧，施皓晨，屠要峰，黄圣君. 基于主动采样的深度鲁棒神经网络学习[J]. 计算机科学, 2022, 49(7): 164-169.

ZHOU Hui, SHI Hao-chen, TU Yao-feng, HUANG Sheng-jun. Robust Deep Neural Network Learning Based on Active Sampling [J]. Computer Science, 2022, 49(7): 164-169.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210600044

基于主动采样的深度鲁棒神经网络学习

周　慧[1,2]　施皓晨[1,2]　屠要峰[1,3]　黄圣君[1,2]
1 南京航空航天大学计算机科学与技术学院　南京 211106
2 模式分析与机器智能工信部重点实验室　南京 211106
3 移动网络和移动多媒体技术国家重点实验室　广东 深圳 518057
(zhouhui@nuaa.edu.cn)

摘　要　随着深度模型在许多实际任务中的广泛应用,提高模型的鲁棒性已经成为了机器学习的重要研究方向。最新的研究表明,通过对训练样本添加噪声扰动进行训练能有效地提升深度模型的鲁棒性。然而,该训练过程往往需要大量已标注样本。在许多实际应用中,准确地标注每一个样本的标记信息往往代价高昂且异常困难。主动学习是降低样本标注代价的主要方法,通过主动选择最有价值的样本进行标注,在提高模型性能的同时,能最大限度地降低查询标记的代价。提出一种基于主动采样的鲁棒神经网络学习框架,该框架能以较低的标注代价显著地提升深度模型的鲁棒性。在该框架中,基于不一致性的主动采样方法通过生成系列扰动样本并采用其预测差异来衡量每个未标注样本对提升模型鲁棒性的潜在效用,同时挑选不一致性值最大的样本用于深度模型的加噪训练。在基准图像分类任务数据集上进行的实验表明,基于不一致性的主动采样策略能以更低的样本标注代价有效地提升深度神经网络模型的鲁棒性。

深度学习;噪声干扰;主动学习;模型鲁棒性;不一致性
关键词:中图法分类号　TP181

Robust Deep Neural Network Learning Based on Active Sampling

ZHOU Hui[1,2],SHI Hao-chen[1,2],TU Yao-feng[1,3] and HUANG Sheng-jun[1,2]
1 College of Computer Science and Technology,Nanjing University of Aeronautics and Astronautics,Nanjing 211106,China
2 MIIT Key Laboratory of Pattern Analysis and Machine Intelligence,Nanjing 211106,China
3 State Key Laboratory of Mobile Network and Mobile Multimedia Technology,Shenzhen,Guangdong 518057,China

Abstract　Recently,deep learning models have been widely used in various real-world tasks. Improving the robustness of deep neural networks has become an important research direction in machine learning field. Recent works show that training the deep model with noise perturbations can significantly improve the model robustness. However,its training requires a large set of precisely labeled examples,which is often expensive and difficult to collect in real-world scenario. Active learning(AL) is a primary approach for reducing the labeling cost,which progressively selects the most useful samples and queries their labels,with the target of training an effective model with less queries. This paper proposes an active sampling based neural network learning framework,which aims to improve the model robustness with low labeling cost. In this framework,the proposed inconsistency sampling strategy is employed to measure the potential utility for improving the model robustness of each unlabeled example with a series of perturbations. Then,those examples with the largest inconsistency will be selected for training the deep model with noise perturbations. Experimental results on the benchmark image classification task data set show that the inconsistency-based active sampling strategy can effectively improve the robustness of the deep neural network model with lower sample labeling cost.

Keywords　Deep learning,Noise perturbations,Active learning,Model robustness,Inconsistency

1　研究背景及意义

深度神经网络(Deep Neural Network,DNN)已被成功地应用于多种实际任务中,如图像识别[1-2]、自然语言处理[3-4]和目标检测等[5]。然而,在现实应用场景中,模型往往受到噪声干扰,性能严重下降。例如,在自动驾驶任务中,图像视频识别模型通常会受到雾、霜、雪、沙尘暴等天气的干扰,难以准确地识别出路标。目前,涌现出了许多关于如何提高深度模型

到稿日期:2021-06-04　返修日期:2021-10-19
基金项目:科技创新2030—新一代人工智能重大项目(2020AAA0107000);国家自然科学基金(62076128)
This work was supported by the Technological Innovation 2030—"New Generation Artificial Intelligence" Major Project(2020AAA0107000) and National Natural Science Foundation of China(62076128).
通信作者:黄圣君(huangsj@nuaa.edu.cn)

的鲁棒性[6-12]的工作。现有研究主要通过解决两种类型下的输入扰动来提高模型的鲁棒性。第一种类型是对抗扰动，该扰动基于某种距离度量（如 $L\infty$ 距离[13]或 Wasserstein 距离[14]）的约束，针对模型而设计，旨在通过最小的输入变化误导神经网络的输出[15]。第二种类型是常见噪声扰动，该扰动通常在数据收集和预处理阶段偶然产生（如高斯噪声[16]和运动模糊[17]）。本文将着重研究第二种类型扰动，即常见噪声扰动。在许多实际应用中，提高模型在常见噪声扰动下的鲁棒性通常比提高模型在对抗扰动下的鲁棒性更具有现实意义。例如，自动驾驶汽车面对的异常天气条件（如冰冻或沙尘暴），语音识别系统面对的背景音乐或其他嘈杂声，这些噪声都可视为常见噪声扰动。

尽管已有许多研究致力于提高模型的鲁棒性并取得了良好的效果，但训练一个鲁棒的深度模型通常需要大量的已标注样本。最近的研究工作表明，训练样本的数量是决定模型鲁棒性的重要因素，且训练一个能够抵御各种噪声扰动的高精度深度模型需要比传统的分类任务规模更大的数据集[18-20]。在现实应用场景中，获取大规模的已标注样本往往代价高昂且异常困难。因此，如何以较低的标记成本高效地学习鲁棒的深度模型是一个亟需解决的问题，且具有重要的现实意义。

主动学习（Active Learning，AL）是降低标注成本的主要途径，其通过迭代查询的方式，主动挑选最有价值的样本进行标注，并将其加入已标注集中用于训练，进而不断提升模型性能。在该过程中，样本的挑选将直接影响模型的性能，因此，采样策略的设计成为了主动学习能否成功的关键因素。基于不确定性的采样策略[21-23]广泛地应用于主动学习任务中，其选择模型最不确定的样本进行标注。还有一些研究试图将不确定性和代表性进行结合，这类采样策略评估了样本对提高分类器性能的潜在贡献[24-27]。然而，这些策略通常为传统的分类任务设计。

本文提出了一种用于训练鲁棒深度模型的主动学习方法（Active Learning for Robust Deep model，ALRD），试图以最小的标注代价有效地提高模型的鲁棒性。与传统的主动学习方法不同，ALRD 框架使用基于不一致性的采样策略从未标注集中主动挑选样本交由专家标注，然后将样本加入已标注集中进行训练。具体地，在每一轮主动查询中，使用当前分类模型对添加噪声干扰的未标注样本进行预测，并挑选那些加噪前后预测最不一致的样本进行查询。

本文的主要贡献包括以下 3 个方面：

（1）提出基于主动采样的鲁棒深度神经网络学习框架，通过主动地挑选最有价值的样本，以最小的标注代价训练一个鲁棒的深度神经网络；

（2）在该学习框架中，基于鲁棒学习的特性，提出基于不一致性的主动采样方法，挑选加噪前后预测不一致性较大的样本加入训练，以提高模型在噪声扰动下的性能；

（3）在图片分类基准数据集上的实验结果验证了本文提出的基于不一致性的主动采样方法的有效性。

本文第 2 节简要回顾了相关工作；第 3 节详细介绍所提方法；第 4 节展示实验结果；最后总结全文。

2　相关工作

2.1　模型鲁棒性

在深度模型中，鲁棒性意味着当样本受到一定程度的扰动（如旋转、平移、噪声、光照条件变化）时，模型能够容忍扰动而保持原本的输出，即在一定扰动范围内（如人眼能够准确识别），模型的预测结果与未受扰动时的预测结果保持一致。尽管现阶段机器学习模型已经能够获得较高的精度，但最新研究表明，在样本中添加微小的扰动就能使得模型产生误判，而这些扰动通常是肉眼难以察觉的。Dodge 等[28]的研究发现，最先进的图像识别网络极易受到模糊和高斯噪声的影响。文献[29-30]比较了人类和深度模型在识别受损图像时的表现，发现随着扰动的增大，深度模型识别性能的下降速度远超过人类识别性能的下降速度。已有许多方法用于提高深度模型的鲁棒性，其中，加噪训练是最有效的方法之一[7,9,31]。本文主要研究利用主动学习技术降低学习鲁棒深度模型所需的标注代价。

2.2　主动学习

由于具有适用性广和高效利用人类专家的特点，主动学习受到广泛关注并迅速发展，成为了机器学习的重要研究方向之一。大多数主动学习研究侧重于设计采样策略，以确保所选的样本能有效地提高模型性能[21]。近年来，许多学者提出挑选样本的准则[33-38]。其中，一些方法选择模型最不确定的样本[34-36]，而另一些方法则选择能体现数据分布的最具代表性的样本[27,37]。最新的研究将主动学习思想应用于鲁棒深度神经网络训练[38]，该研究在样本均具有标记信息的情况下，通过查询使得样本预测结果变化的最小扰动水平来提高模型的鲁棒性。本文对样本标记进行查询，并提出了一种新的主动采样策略，以最小的标注代价有效地提高深度模型的鲁棒性。

3　基于主动采样的鲁棒深度神经网络学习方法

本文中，$\boldsymbol{x}_i \in \mathbb{R}^d$ 代表第 i 个样本的特征向量，$\boldsymbol{y}_i \in \{1,\cdots,k\}$ 代表该样本对应的真实标记。$L=\{(\boldsymbol{x}_i,\boldsymbol{y}_i)\}_{i=1}^{n_l}$ 代表已标注样本集，其中 n_l 为已标注样本数量；$U=\{(\boldsymbol{x}_i,\boldsymbol{y}_i)\}_{i=1}^{n_u}$ 代表未标注样本集，其中 n_u 为未标注样本数量。$C=\{(\hat{\boldsymbol{x}}_i,\boldsymbol{y}_i)\}_{i=1}^{n}$ 为测试集，其中 $\hat{\boldsymbol{x}}_i \in \mathbb{R}^d$ 为加噪声干扰后的样本。$f_\theta(\boldsymbol{x}_i)$ 为深度神经网络对样本 $\boldsymbol{x}_i$ 的输出，其中 θ 是神经网络参数。

3.1　鲁棒深度神经网络

一般来说，在干净训练集上训练出的模型难以泛化到含有噪声的测试数据集。受到文献[30-32]等的启发，本文采用对训练样本增添噪声扰动的方式进行训练，以提高模型的鲁棒性，具体可通过解决如下优化问题实现：

$$\min_{\theta}\sum_{i=1}^{n_l}[E_{\boldsymbol{\varepsilon}\sim p(\sigma)}[L(f_\theta(\boldsymbol{x}_i+\boldsymbol{\varepsilon}_i),\boldsymbol{y}_i)]] \tag{1}$$

其中，$\boldsymbol{\varepsilon}$ 是根据扰动分布 $P(\sigma)$ 生成的随机噪声，σ 用于控制扰动强度。这里，$P(\sigma)$ 可以是任何常见的数据分布，在本文中，$P(\sigma)$ 是高斯分布，σ 为高斯分布的方差。显然，通过解决上述优化问题，模型在拟合含噪样本的过程中逐渐提高了对噪声

的鲁棒性。该训练过程需要大量的已标注样本，然而在实际应用中，精确标注样本的获取通常代价高昂。本文将结合主动学习方法，挑选出对提高模型鲁棒性最有效的样本进行查询，然后用于模型训练，降低样本标注成本。

3.2 基于主动采样的鲁棒深度神经网络学习框架

如图 1 所示，基于主动采样的鲁棒深度神经网络学习框架从未标注集中挑选样本，用于模型的下一轮训练。与传统的主动学习方法不同，在每一轮迭代中，首先，ALRD 使用高斯数据增强的方法对已标注样本集 L 中的每一个样本进行增强，得到含噪标注集 L_{cor}，并根据式(1)训练深度模型；接着，基于模型的输出，ALRD 使用不一致性采样方法估计未标记集 U 中每个样本的不一致性，并挑选不一致性值最大的样本交由专家标注，以加入已标注集 L 中用于下一轮训练。不断地迭代整个过程，直至达到预先设定的总标注成本。

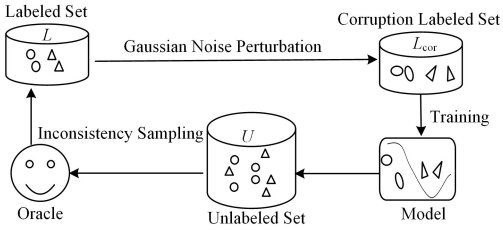


图 1 ALRD 框架

Fig. 1 Framework of ALRD

3.3 不一致性采样

本节将重点研究如何选择对提高模型鲁棒性最有用的样本进行查询。一般来说，鲁棒模型的预测往往具有稳定性，即当对输入样本添加微小扰动时，模型的输出应保持一致。然而，在同样程度的扰动下，模型在不同样本上的预测稳定性也不同。图 2 给出了模型在不同样本上预测稳定性的差异，其中，A，B，C，D 分别代表不同类别，圆形与三角形分别为对样本 X 与样本 Y 加噪后的扰动样本(这些扰动样本的真实标记均与原样本一致)，虚线圆分别为服从高斯分布 $N(X,\sigma^2 I)$ 与 $N(Y,\sigma^2 I)$ 的扰动样本。

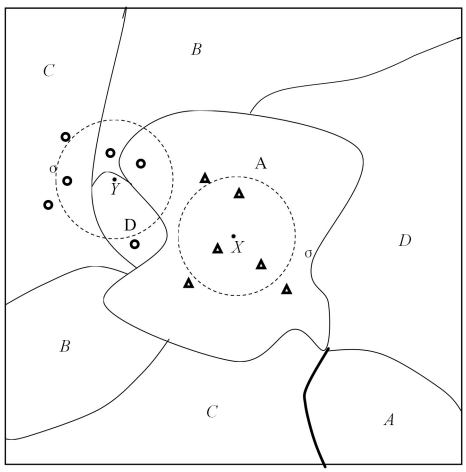


图 2 当前分类器及其决策边界

Fig. 2 Current classifier and its decision regions

从图中可以观察到，模型对 X 周围的扰动样本的预测结果与对 X 的预测结果基本一致，对 Y 周围的扰动样本的预测结果与对 Y 的预测结果存在较大的不一致性。显然，对于 X 这类样本，在遇到噪声扰动时，模型也能保持原来的预测结果，这样的样本对提高模型鲁棒性没有明显帮助；而对于 Y 这类样本，在遇到噪声扰动时，模型的预测结果非常不稳定，将这些样本加入已标注集用以训练模型，能够有效提高模型的鲁棒性。

基于此，本文提出了基于不一致性的主动采样策略，挑选加噪前后预测最不一致的样本进行查询。具体地，如图 3 所示，在每一轮主动查询中，对未标注样本添加随机采样的高斯噪声扰动，并计算模型在这些样本上的预测结果与在干净样本上的预测结果相比类别翻转的概率，并将翻转概率(Flip Probability，FP)用于衡量样本的不一致性。

$$FP(\boldsymbol{x}_i)=\frac{1}{m}\left(\sum_{j=1}^{m}\mathbb{I}(f(\boldsymbol{x}_i^j)\neq f(\boldsymbol{x}_i))\right) \tag{2}$$

对每一个样本 $\boldsymbol{x}_i$ 添加 m 次扰动，得到扰动样本集合 $S(\boldsymbol{x}_i)=\{(\boldsymbol{x}_i^{(1)},\boldsymbol{x}_i^{(2)},\cdots,\boldsymbol{x}_i^{(m)})\}$，其中，$\forall 1\leqslant j\leqslant m$，$\boldsymbol{x}_i^{(j)}=\boldsymbol{x}_i+\boldsymbol{\varepsilon}_i^{(j)}$，$\boldsymbol{\varepsilon}_i^{(j)}$ 是一个从高斯分布 $N(0,\sigma^2 I)$ 随机采样的扰动值，$f(\boldsymbol{x}_i)$ 则是模型对样本 $\boldsymbol{x}_i$ 的预测值。

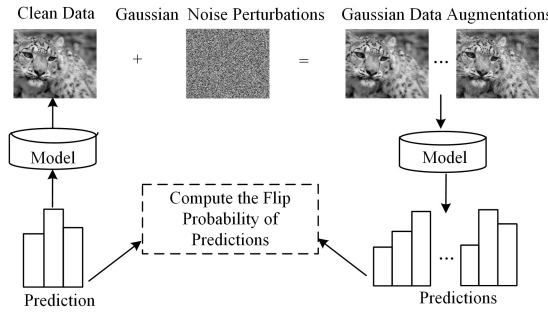


图 3 基于不一致性的采样策略

Fig. 3 Sampling strategy based on inconsistency

4 实验与结果

4.1 实验设置

为了验证基于不一致性的主动采样策略的有效性，本文在两个基准数据集上进行实验。分别在 MNIST[39] 和 TinyImagenet[40] 上进行训练，并在 MNIST-C[41] 和 TinyImagenet-C[17] 上进行测试。MNIST-C 和 TinyImagenet-C 分别是 MNIST 和 TinyImagenet 所对应的含噪数据集，其中，每个噪声数据集包含 15 种噪声，这 15 种噪声分别属于运动模糊、天气扰动以及图像破损 3 个类别，同时，每种噪声扰动的强度包含 5 个等级。对于 TinyImagenet，随机采样 10 个类作为训练集，并将其命名为 TinyImagenet10，对应的噪声测试集则为 TinyImagenet10-C。在实验中，每种方法都在两个数据集上进行 5 次实验，并报告其平均测试准确率。在 MNIST 与 TinyImageNet 上，本研究分别使用 LeNet 与 ResNet-18 作为分类基模型。

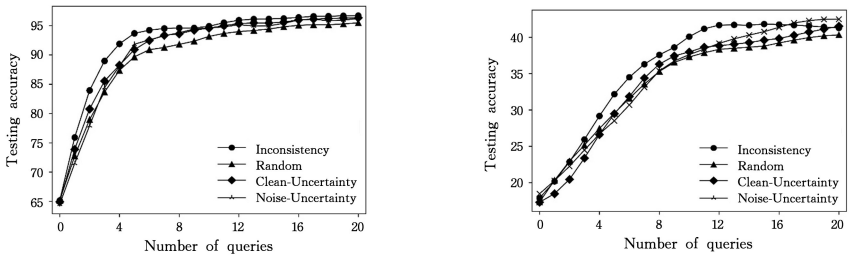
在 MNIST 上，从数据集中随机选择 100 个样本作为初始训练集，每轮迭代挑选 50 个样本进行标注，然后将样本加入训练集；在 TinyImagenet10 上，从数据集中随机挑选 200 个样本作为初始训练集，每轮迭代挑选 200 个样本进行标注，然后将样本加入训练集。在 MNIST 与 TinyImagenet10 上，参数 σ 分别设定为 0.25 与 0.08，具体分析见 4.3 节。在估计

未标注样本的不一致性过程中,扰动次数越多,不一致性估计值越准确,但往往计算复杂度也增加了。前期的实验表明,当扰动次数 m 设为 10 时,已能够达到较好的效果,出于计算代价考虑,本实验将扰动次数 m 设定为 10。

由于本文提出的是一种新的学习框架,目前还没有方法可以解决该学习任务,因此将提出的基于不一致性的主动采样方法与 3 种基准方法进行比较:1)随机采样(Random),即随机挑选样本;2)原始不确定性采样(Clean-Uncertainty),即直接评估未标注样本的不确定性,挑选不确定性最大的样本;3)加噪不确定性采样(Noise-Uncertainty),即上一种采样方法的扩展,对未标注样本加噪后评估其不确定性,挑选加噪后不确定性最大的样本。在实验中,我们将使用模型预测的信息熵来衡量样本的不确定性。

4.2　实验结果

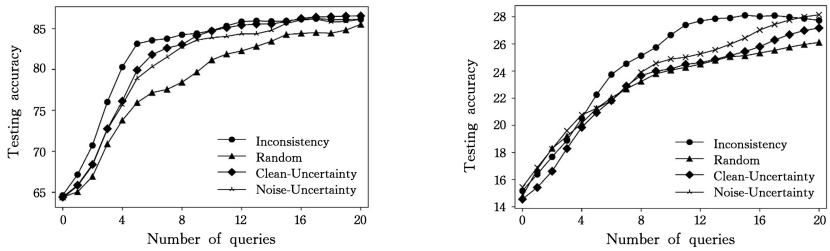
图 4 和图 5 分别给出了随着主动查询轮数的增加,各个对比方法在 MNIST-C 和 TinyImagenet10-C 上的测试准确率。可以看出,本文提出的基于不一致性的主动采样方法明显优于其他两种对比方法。当采样轮数相同(固定 x 轴)时,基于不一致性的主动采样方法几乎在所有情况下都在两个数据集上达到了最高的测试准确率。这是由于不一致性值能准确地反映样本对提高模型鲁棒性的潜在效用,通过挑选效用最高的样本,往往能有效地提升深度模型的鲁棒性。



(a)MNIST-C　(b)TinyImagenet10-C

图 4　对比方法在高斯噪声干扰数据集上的性能比较

Fig. 4　Performance comparison of different methodson Gaussiannoise perturbing datasets



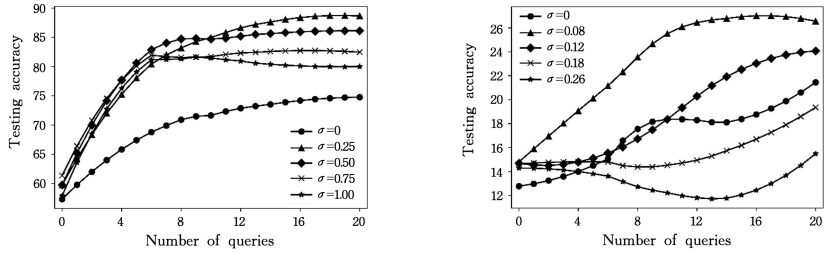
(a)MNIST-C　(b)TinyImagenet10-C

图 5　对比方法在 15 种噪声干扰数据集上的性能比较

Fig. 5　Performance comparison of different methods on 15 types of noise perturbing datasets

4.3　参数 σ 对模型鲁棒性的影响

本小节研究噪声强度 σ 对实验结果的影响。图 6 给出了本文提出的基于不一致性的主动采样方法在不同噪声强度 σ 的情况下在 MNIST-C 和 TinyImagenet-C 上的性能曲线。在 MNIST 上,训练中采用 $\sigma=\{0,0.25,0.5,0.75,1\}$;而在 TinyImagenet 上,采用 $\sigma=\{0,0.08,0.12,0.18,0.26\}$,即测试数据集 TinyImagenet-C 构造高斯噪声数据集时所采用的方差值。由图 6 可知,在噪声强度 σ 不同的情况下,不一致性采样的性能表现差异较大,随着 σ 增大,模型性能呈现出先递增后递减的变化趋势。当 σ 的值设置较小时,由于噪声扰动较微弱,难以有效地提高模型鲁棒性;当 σ 的值设置较大时,由于噪声扰动过于强烈,加噪干扰后的图像语义往往难以辨别,模型通常难以有效地拟合这些样本,导致模型的泛化性能显著下降。进一步地,当 σ 在 0.25 附近时,基于不一致性的主动采样方法在 MNIST-C 上能达到较好性能;当 $\sigma=0.08$ 时,基于不一致性的主动采样方法在 TinyImagenet-C 上达到了最佳性能。



(a)MNIST-C　(b)TinyImagenet10-C

图 6　参数 σ 对不一致采样性能的影响

Fig. 6　Performance curve of inconsistency sampling with different σ

结束语 本文将主动学习技术运用于鲁棒深度模型的学习中，通过主动选择最有价值的样本，以最小的标注代价有效地提高深度神经网络的鲁棒性。所提基于不一致性的主动采样方法，通过使得模型对样本加噪声扰动前后的预测保持一致来有效提高模型的鲁棒性。在两个基准图片识别数据集上进行的实验表明所提方法能以较低的标注代价有效地提高模型在噪声干扰下的性能。将不一致性采样用于解决对抗噪声下的鲁棒深度神经网络的学习问题，是下一步的研究方向。

参 考 文 献

[1] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[2] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification [C]// Proceedings of the IEEE International Conference on Computer vision. 2015: 1026-1034.

[3] XIONG W, DROPPO J, HUANG X, et al. Achieving human parity in conversational speech recognition [J]. arXiv: 1610.05256, 2016.

[4] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[J]. arXiv: 1409.3215, 2014.

[5] LIU L, OUYANGW, WANG X, et al. Deep learning for generic object detection: A survey[J]. International Journal of Computer Vision, 2020, 128(2): 261-318.

[6] GEIRHOS R, RUBISH P, MICHAELIS C, et al. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness[J]. arXiv: 1811.12231, 2018.

[7] RUSAK E, SCHOTT L, ZIMMERMANN R S, et al. A simple way to make neural networks robust against diverse image corruptions [C]// European Conference on Computer Vision. Cham: Springer, 2020: 53-69.

[8] HENDRYCKS D, MAZEIKA M, KADAVATH S, et al. Using self-supervised learning can improve model robustness and uncertainty[J]. arXiv: 1906.12340, 2019.

[9] HENDRYCKS D, MU N, CUBUK E D, et al. Augmix: A simple data processing method to improve robustness and uncertainty [J]. arXiv: 1912.02781, 2019.

[10] MAO C, ZHONG Z, YANG J, et al. Metric learning for adversarial robustness[J]. arXiv: 1909.00900, 2019.

[11] ZHANG R. Making convolutional networks shift-invriance again [J]. Proceedings of Machine Learning Research, 2019, 97: 7324-7334.

[12] TRAMER F, CARLINI N, BRENDEL W, et al. On adaptive attacks to adversarial example defenses[J]. arXiv: 2002.08347, 2020.

[13] MADRY A, MAKELOV A, SCHMIDT L, et al. Towards deep learning models resistant to adversarial attacks[J]. arXiv: 1706.06083, 2017.

[14] WONG E, SCHMIDT F, KOLTER Z. Wasserstein adversarial examples via projected sinkhorn iterations [C]// Internatioal Conference on Machine Learning. PMLR, 2019: 6808-6817.

[15] SZEGEDY C, ZAREMBA W, SUTSKEVER I, et al. Intriguing properties of neural networks[J]. arXiv: 1312.6199, 2013.

[16] CHAPELLE O, WESTON J, BOTTOU L, et al. Vicinal risk minimization[C]// Conference and Workshop on Neural Information Processing Systems (NIPS). 2000.

[17] HENDRYCKS D, DIETTERICH T. Benchmarking neural network robustness to common corruptions and perturbations[J]. arXiv: 1903.12261, 2019.

[18] XIE Q, LUONG M T, HOVY E, et al. Self-training with noisy student improves imagenet classification [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2020: 10687-10698.

[19] MAHAJAN D, GIRSHICK R, RAMANATHAN V, et al. Exploring the limits of weakly supervised pretraining [C]// Proceedings of the European Conference on Computer Vision (ECCV). IEEE, 2018: 181-196.

[20] UESATO J, ALAYRAC J B, HUANG P S, et al. Are labels required for improving adversarial robustness? [J]. arXiv: 1905.13725, 2019.

[21] LEWIS D D, GALE W A. A sequential algorithm for training text classifiers[C]// SIGIR'94. London: Springer, 1994: 3-12.

[22] TONG S, KOLLER D. Support vector machine active learning with applications to text classification [J]. Journal of machine learning research, 2001, 23: 45-66.

[23] YAN Y, HUANG S J. Cost-Effective Active Learning for Hierarchical Multi-Label Classification [C] // IJCAI. 2018: 2962-2968.

[24] HUANG S J, GAO N, CHEN S. Multi-instance multi-label active learning[C]// IJCAI. 2017: 1886-1892.

[25] HUANG S J, ZHOU Z H. Active query driven by uncertainty and diversity for incremental multi-label learning [C]// 2013 IEEE 13th International Conference on Data Mining. IEEE, 2013: 1079-1084.

[26] HUANG S J, RONG J, ZHOU Z H. Active learning by querying informative and representative examples [C]// Conference and Workshop on Neural Information Processing Systems (NIPS). 2014: 1936-1949.

[27] WANG Z, YE J P. Querying discriminative and representative samples for batch mode active learning[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2015, 9(3): 1-23.

[28] DODGE S, KARAM L. Understanding how image quality affects deep neural networks [C]// 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2016: 1-6.

[29] DODGE S, KARAM L. A study and comparison of human and deep learning recognition performance under visual distortions [C]// 2017 26th International Conference on Computer Communication and Networks (ICCCN). IEEE, 2017: 1-7.

[30] GEIRHOS R,TEMME C R M,RAUBER J,et al. Generalisation in humans and deep neural networks[J]. arXiv:1808.08750,2018.

[31] CARLINI N,ATHALYE A,PAPERNOT N,et al. On evaluating adversarial robustness[J]. arXiv:1902.06705,2019.

[32] CHEN P,ZHAO J C,YU X S. Ensemble Method of K-Nearest Neighbor Enhancement Fuzzy Minimax neuralNetworks with Centroid[J]. Journal of Chongqing University of Technology (Natural Science),2021,35(9):116-129.

[33] FU Y,ZHU X,AND LI B. A survey on instance selection for active learning[J]. Knowledge and Information Systems,2013,35(2):249-283.

[34] SEUNG HS,OPPER M,SOMPOLINSKY H. Query by committee[C]// Proceedings of the Fifth Annual Workshop on Computational Learning Theory. 1992:287-294.

[35] YOU X,WANG R,TAO D. Diverse expected gradient active learning for relative attributes[J]. IEEE Transactions on Image Processing,2014,23(7):3203-3217.

[36] ROY N,MCCALLUM A. Toward optimal active learning through sampling estimation of error reduction[C]// International Conferenceon Machine Learning(ICML). 2001:441-448.

[37] GEMAN S,BIENENSTOCK E,DOURSAT R. Neural networks and the bias/variance dilemma[J]. NeuralComputation,1992,4(1):1-58.

[38] NING KP,TAO L,CHEN S,et al. Improving Model Robustness by Adaptively Correcting Perturbation Levels with Active Queries[J]. arXiv:2103.14824.2021.

[39] 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.

[40] YAO L,MILLER J. Tiny imagenet classification with convolutional neural networks[J]. CS 231N,2015,2(5):8.

[41] MU N,GILMER J. Mnist-c:A robustness benchmark for computer vision[J]. arXiv:1906.02337,2019.

ZHOU Hui, born in 1997, master. Her main research interests include machine learning and so on.

HUANG Sheng-jun, born in 1987, professor. His main research interests include machine learning and data mining.

(责任编辑:柯颖)