



基于修正标签分布的乳腺超声图像分类

曹占涛¹, 杨国武¹, 陈 琴², 吴尽昭³, 李晓瑜^{4*}

(1. 电子科技大学计算机科学与工程学院 成都 611731; 2. 电子科技大学医学院 成都 610072;
3. 广西民族大学理学院 南宁 530006; 4. 电子科技大学信息与软件工程学院 成都 610054)

【摘要】针对乳腺超声图像分类中的标签噪音问题, 该文设计了一种协作标签修正网络 (COLC-Net)。该方法基于乳腺超声 BI-RADS 评级噪音分布特点, 为乳腺超声图像定义了软标签, 并设计了双网络协作训练, 以蒸馏优秀知识修正软标签。随着软标签准确性的增加, 可以降低噪音标签负作用, 并增强准确标签知识的学习。与现有最新方法进行比较, 结果证实了该方法具有更好的效果。

关键词 乳腺超声图像; 深度学习; 噪音标签; 弱监督学习

中图分类号 TP391 文献标志码 A doi:10.12178/1001-0548.2020001

Breast Ultrasound Images Classification Based on Correction Label Distribution

CAO Zhan-tao¹, YANG Guo-wu¹, CHEN Qin², WU Jin-zhao³, and LI Xiao-yu^{4*}

(1. School of Computer Science and Engineering, University of Electronic Science and Technology of China Chengdu 611731;
2. School of Medicine, University of Electronic Science and Technology of China Chengdu 610072;
3. College of Science, Guangxi University for Nationalities Nanning 530006;
4. School of Information and Software Engineering, University of Electronic Science and Technology of China Chengdu 610054)

Abstract In order to solve the problem of label noise in breast ultrasound image classification, an efficient method called cooperative label correction network (COLC-Net) is proposed. In this method, based on the noise distribution characteristics of the breast ultrasound BI-RADS (breast imaging-reporting and data system) rating, soft labels are proposed for breast ultrasound images, and two networks are proposed for collaborative training. Excellent knowledge is distilled from the two networks to modify the soft labels. With the increase of the accuracy of soft labels, the negative effects of noise labels can be reduced and the learning of clean labels can be enhanced. In order to verify the effectiveness of the method, extensive comparisons are conducted with existing state-of-the-art methods on the dataset. The results demonstrate the effectiveness of the proposed method.

Key words breast ultrasound image; deep learning; noisy label; weakly supervised learning

乳腺癌已成为当前社会最常见的癌症之一, 自 20 世纪 70 年代以来发病率一直处于上升趋势。在美国, 约 12% 的女性会患乳腺癌疾病, 仅在 2017 年就新增 252 710 个病例^[1], 由于乳腺癌的患病原因仍然未知, 因此尽早发现是降低死亡率的关键^[2]。乳腺癌的常用筛查方法包括影像检查和活组织检查 (活检), 活检花费高, 并且痛苦, 不太适用于早期筛查。在乳腺癌的医学影像诊断方法中, 乳腺超声是无辐射、经济、高效的诊断方法^[3]。乳腺超声成像旨在检测和分类良性或恶性肿瘤, 乳腺超

声的诊断, 基于美国放射学会发布的乳腺影像报告和数据系统 (BI-RADS)^[4]。乳腺超声图像诊断的准确性取决于从业者的经验, 需要训练有素且经验丰富的临床专业医生。此外, 即使训练有素的专家也可能具有较高的观察者间差异。所以基于乳腺超声图像的计算机辅助诊断, 一直是重要研究课题。

深度学习在很多领域都取得了卓越的成果, 例如: 汽车不安全行为识别^[5]、无人机智能跟踪^[6]、人体解析^[7]等。这使我们能够重新思考基于乳腺超声图像的计算机辅助诊断方法。基于深度学习的方法

收稿日期: 2020-01-02; 修回日期: 2020-02-21

基金项目: 国家自然科学基金 (61572109, 61772006); 广西“八桂学者”专项

作者简介: 曹占涛 (1983-), 男, 博士生, 主要从事深度学习、计算机辅助诊断方面的研究。

通信作者: 李晓瑜, E-mail: xiaoyu@uestc.edu.cn

法要想达到较好的效果,需要大量准确标注的训练样本数据。然而,由于受到患者隐私和机构政策之类因素的限制,乳腺超声图像标注数据短缺是研究者面临的挑战之一。乳腺肿瘤诊断方法中只有活检才能最终确诊^[2],并作为超声图像的准确标签。然而电子病历库中很多乳腺超声肿瘤图像没有活检结果^[8-10],而只有基于 BI-RADS 的标签。BI-RADS 评级分为 0-6,不同评级按恶性概率区分:0 级表示评估不完整,需要补充检查;1 级表示无异常发现;2 级表示良性,恶性概率为 0%;3 级表示恶性概率为 0%~2%;4a 级表示恶性概率为 2%~10%;4b 级表示恶性概率为 10%~50%;4c 级表示恶性率概率为 50%~95%,5 级表示恶性率为 95%~100%;6 级表示恶性概率为 100%^[4]。当医生为病人拍摄乳腺超声图像后,依据专业经验,给出每张图像的 BI-RADS 评级,当怀疑有恶性可能时,会推荐病人做活检进行确诊^[11],这就导致很多低评级乳腺超声图像没有活检结果。另外评级高的部分病人可能会选择其他医院进行活检,而医院间的数据并不共享。这些原因导致了从一家医院很难收集大量有活检结论的乳腺超声图像。为了克服有准确标签乳腺肿瘤超声图像短缺的问题,一些先前的工作直接基于乳腺超声 BI-RADS 评级来生成乳腺肿瘤的良恶性标签。这些工作中直接将 BI-RADS 评级的 2-4a(2、3、4a)作为良性肿瘤,BI-RADS 评级的 4b-6(4b、4c、5、6)作为恶性肿瘤^[12-13]。然而 BI-RADS 评级只提供了乳腺肿瘤的恶性可能性,在机器学习领域,这个问题会导致噪音标签问题^[14-15]。近年来,在机器学习领域,对训练带有噪声标签的深度模型为目标的弱监督分类方法进行了广泛的研究^[16]。解决噪声标签的学习方法主要分为 4 类:基于损失的噪音鲁棒性方法^[17-19]、噪音迁移预估方法^[20-23]、样本选择方法^[24-26]、标签修正方法^[27-28]。

本文联合四川省人民医院进行基于 BI-RADS 评级的乳腺超声肿瘤图像分类研究,并收集标注了一个乳腺超声数据集。为了解决基于 BI-RADS 评级的乳腺超声肿瘤分类模型中的噪音标签问题,设计了一个协作标签修正网络(COLC-Net)。COLC-Net 利用深度网络的“记忆能力”^[29],即在训练初期优先记住特征明显或容易识别的样本特征,并结合乳腺超声图像的 BI-RADS 标签噪音率两端低、中间高的分布特点,为乳腺超声图像设计了软标签,在训练过程中进行标签修正,而不是将噪音标签视为固定值。为了通过网络参数和标签联

合训练来优化网络,本文设计了双网络来协作修正标签。在训练过程中,从两个网络中蒸馏优秀知识修正软标签从而提升网络的泛化能力,随着软标签准确性的增加,可以降低噪音标签负作用,并增强准确标签知识的学习。

1 设计方法

1.1 数据集

本文从四川省人民医院历史电子病历库,收集了一个乳腺超声图像数据集。这些乳腺超声图像是被 GE-LOGIQ、Mindray 和 Philips IU-ELITE 医学超声设备采集。原始超声图像包括一些无关信息,例如:医院名字、图像注释和相关参数,这些信息是无用的甚至会引入偏差。由于本研究主要关注肿瘤的良恶性分类,为了排除无关信息及脂肪等其他组织的干扰,对图像进行了手工预处理,即对图像进行裁剪,只裁剪出感兴趣的肿瘤区域,将其他无关信息删除。由于肿瘤区域不规则,本文选择的裁剪形状为长方形,长方形区域包含整个肿瘤区域。值得注意的是,所有图像的裁剪均由 3 位经验丰富的超声医生进行,并进行了交叉确认。

数据集依据是否有活检结论分为两部分:准确标签数据集和噪音标签数据集。准确数据集包括 473 个良性样本和 462 个恶性样本,准确标签数据集具有活检和 BI-RADS 两种标签。对于噪音标签数据集,参照以前的工作^[12-13],使用 BI-RADS 评级来产生良恶性标签,将 BI-RADS 评级为 2-4a(2、3、4a)的样本视为良性肿瘤,将 BI-RADS 评级为 4b-6(4b、4c、5、6)的样本视为恶性肿瘤,噪音标签数据集包括 577 个良性样本和 464 个恶性样本。本文对样本进行随机划分,其中从准确标签数据集中随机取 70 张良性图像和 66 张恶性图像作为验证集,139 张良性图像和 133 张恶性图像作为测试集。准确标签数据集的剩余部分按 BI-RADS 评级产生噪音标签,和噪音数据集一起组成训练集。所有训练集按 BI-RADS 定义的恶性概率平均值,生成软标签初始值。

1.2 方法

定义 $X = \{x_1, x_2, \dots, x_n\}$ 表示具有 n 个样本的训练集,其中 $x_n \in R^{W \times H \times 3}$ 表示乳腺超声图像。定义向量 $l = [1, 1]$, l^T 为 l 的转置,定义 $H = \{h: h \in \{0, 1\}^2, h \times l^T = 1\}$ 为基于 BI-RADS 评级的噪音标签空间,称为硬标签,标签 h_i 可能和真实标签不一致。定义 $S = \{s: s \in [0, 1]^2, s \times l^T = 1\}$ 为基于 BI-RADS 评级的

标签分布空间, 称为软标签, 它可以体现乳腺超声图像类别相关性强弱。每个样本 x_i 对应的软标签 $s_i \in S$ 不同于硬标签 $h_i \in H$, h_i 是一个可以表达真实类别标签的 one-hot 向量, 软标签 s_i 是标签分布, 表示一个样本 x_i 属于某个类别的概率。

传统的监督学习方法基于硬标签进行学习, 其优化问题如下:

$$\min_{\theta} L(\theta|X, H) \quad (1)$$

深度学习训练过程会尽量拟合所有标签, 在基于噪音标签的乳腺肿瘤超声图像良恶性分类问题中, 深度学习网络模型拟合准确标签的同时也会拟合噪音标签。然而, 噪音标签会引入偏差, 对网络模型产生负作用。为了解决噪音问题产生的负作用, 利用 BI-RADS 评级表达的是一张乳腺超声图像属于良恶性概率的特点, 本文为每张超声图像生成噪音标签 H 的同时, 生成一个软标签 S 。通过软标签和硬标签联合训练来优化网络, 本文要解决的优化问题用公式描述如下:

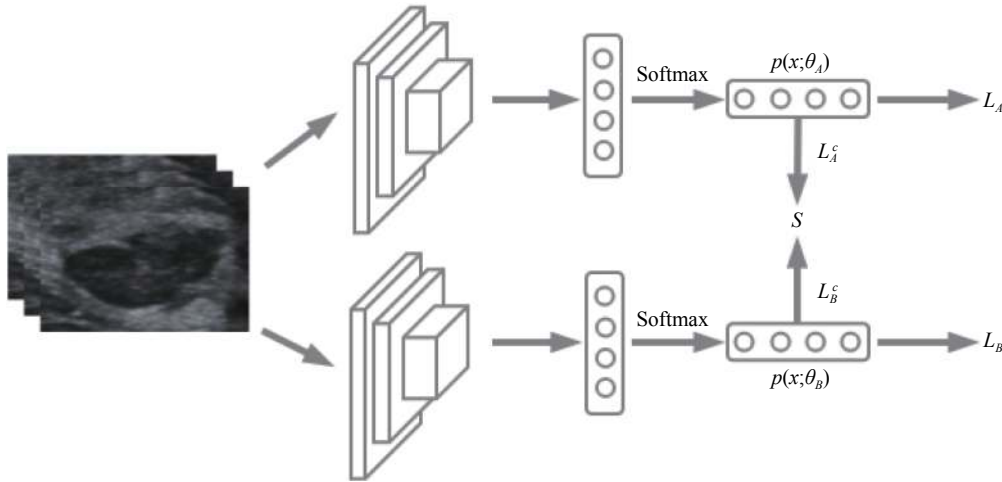


图1 本文提出的 COLC-Net 总体结构

$$\min_{\theta_A} L_A = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^2 h_{ij} \log(p(x_i; \theta_A)) \quad (4)$$

$$\min_{\theta_B} L_B = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^2 h_{ij} \log(p(x_i; \theta_B)) \quad (5)$$

式中, h_{ij} 是 $h_i \in H$ 的第 j 个元素; θ_A 是 A 网络的学习参数; θ_B 是 B 网络的学习参数。

为了高效利用基于软标签 S 的特征知识, 希望在训练过程中修正软标签, 并让修正后的标签促进网络的整体学习能力。为了让软标签 S 的分布尽量准确, 本文设计的学习策略为: 软标签 S 从两个网

$$\min_{\theta} L(\theta|X, H, S) \quad (2)$$

软标签的初始值来自于 BI-RADS 评级。然而不同的 BI-RADS 评级给出的仅仅是一个概率范围, 且评级边界较模糊, 甚至存在由于医生经验而导致的误差。所以无法仅依据 BI-RADS 评级获得图像准确的软标签值。为了获得尽量可靠的软标签值, 软标签在迭代过程中动态修正。所以本文的目标优化问题通过公式描述如下:

$$\min_{\theta, S} L(\theta, S|X, H) \quad (3)$$

深度神经网络训练数据时, 网络具有一定的随机性, 对同一个训练批次的数据, 两个分类器具有不同的学习能力, 会产生不同的预测值。受此思想的启发, 为了减少噪音标签的负作用, 从而加强训练集中准确标签数据的学习, 本文设计了一个针对标签修正的双网络协作学习框架, 如图1所示。在此框架中, 首先设计了基于硬标签的学习任务, 定义网络学习参数为 θ , 分类器的预测输出为 $p(x, \theta)$, 用公式定义两个网络 A 和 B 的交叉熵损失如下:

络中蒸馏知识, 随着 S 的能力增强, 会约束训练数据集中标签的噪音, 减弱噪音的负作用, 并增强正确标签的学习能力。为了让软标签可以从两个网络中蒸馏出更优的知识, 本文设计了基于软标签的协作损失函数, 用 Kullback-Leibler (KL) 距离表示软标签的分类损失函数:

$$\min_{\theta_A, S} L_A^c = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^2 s_{ij} \log\left(\frac{s_{ij}}{p(x_i; \theta_A)}\right) \quad (6)$$

$$\min_{\theta_B, S} L_B^c = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^2 s_{ij} \log\left(\frac{s_{ij}}{p(x_i; \theta_B)}\right) \quad (7)$$

式中, s_{ij} 表示样本 x_i 对应于类别 j 的软标签分布; L_A^c 表示网络 A 的软标签损失函数; L_B^c 表示网络 B 的软标签损失函数。

软标签修正方法中, 定义大小为 M 的软标签容器。每个迭代基于硬标签分类器的预测输出被放入容器, 当容器满后, 最旧的值会被替换。软标签的更新规则如下:

$$S = \frac{1}{M} \sum_{m=1}^M f(L_A^c, L_B^c)_m \quad (8)$$

式中, $f(L_A^c, L_B^c)$ 表示训练过程中每个 batch 从 L_A^c 和 L_B^c 中取较小者的网络预测值放入软标签容器。

综上所述, 最终目标损失函数定义如下:

$$\min_{\theta_A, \theta_B, S} L = \alpha L_A + \beta L_B + \gamma(L_A^c + L_B^c) \quad (9)$$

式中, α , β 和 γ 是可调节超参数, 用来平衡不同目标损失函数的权重。模型通过梯度下降算法来优化最终目标函数 L 。

2 实验对比

2.1 实验细节

本文设计的所有实验是基于深度学习框架 Pytorch, 实验环境的 GPU 为 NVIDIA GTX 1080, 操作系统为 Ubuntu14.04。由于数据集较小, 为了实验的公平性, 所有对比实验中均采用 AlexNet 作为基础网络, 并采用 AlexNet 基于 ImageNet 的预训练模型作为网络的初始值。实验过程中, 训练阶段采用的优化器为 Adam optimizer。为了适应 Alexnet 网络的输入, 图片大小重置为 227×227 , 并随机裁剪为 224×224 。超参数均按经验值设置, 其中分类器初始学习率为 0.005, 网络其他参数的基础学习率为 0.000 1, 衰减周期为 4。设置 Momentum 值为 0.9, 权重衰减值为 0.001, 批尺寸的值 16。最终目标损失函数的超参数 $\alpha=0.5$ 、 $\beta=0.5$ 、 $\gamma=2$ 。在标签更新过程中, 软标签容器大小为 2, 前 2 个迭代使用初始软标签值, 从第 3 个迭代开始使用修正后的软标签。在对比实验中, 为了获得收敛结果, 所有试验均进行了 20 个迭代。

2.2 结果和讨论

为了验证 COLC-Net 算法的有效性, 在收集的乳腺超声数据集上设计了充分的实验, 并和最新的优秀方法进行了对比, 对比方法的详情描述如下:

AlexNet: 本实验使用的基础网络模型。

Co-teaching: 文献 [25] 设计的有优异成绩的处理噪声标签的弱监督学习方法, 此方法使用“最小

损失”标准筛选的样本来交叉训练两个网络。

Decouple: 文献 [26] 设计的一个非常简单的处理噪声标签的方法, 此方法使用两个网络来进行训练, 每个迭代只选择两个网络预测不一致的样本参与网络参数更新。

JOF: 文献 [28] 设计的标签修正方, 此方法通过最小化网络预测值和软标签值之间的 KL 距离来修正软标签值。

PENCL: 文献 [30] 设计的标签修正方法, 此方法是对 JOF 的改进, 将软标签的更新融入反向传播和梯度下降迭代过程中。

COLC-Net: 本文设计的协作标签分布修正方法。

为了实验比较的公平性, 所有的比较方法均使用 AlexNet 作为基础模型。实验中恶性肿瘤是正类。表 1 为本文设计的 COLC-Net 和现有的最新优秀方法在准确率、精度、召回率和 F1-Score 上的对比。从实验结果的第一行可以看到由于数据有噪声, 网络在学习过程中会拟合错误标签, AlexNet 的表现并不能令人满意。第二行和第三行采用基于深度学习的样本选择方案来解决标签噪声问题, 但是, 与 MINIST^[31] 等标准基准数据集不同, 本文处理的乳腺超声图像类别之间的差异非常细微。因此, 这两种比较的基线方法将遇到样本选择偏差的问题, 样本选择偏差将导致误差在深度模型的训练过程中累积。第四行和第五行, 采用基于深度学习的软标签修正方案来解决标签噪声问题, 然而从结果来看, 这两种方法的性能相比于 AlexNet 并没有明显提升。这是由于这两种方法的共同点是必须采用大学习率来训练, 因为在大学学习率下, 网络不容易过拟合噪声标签。然而对于乳腺超声图像, 图像类别区别度低, 使用大学习率很容易陷入局部最优, 影响其学习性能, 所以 JOF 和 PENCL 并不适合处理此类问题。最后一行可以看到, 本文方法的效果明显优于其他方法。

表 1 本文算法与相关算法结果比较

模型	准确率/%	精度/%	召回率/%	F1-Score
AlexNet	72.4	65.6	91.7	0.765
Co-teaching	74.6	69.8	85.0	0.766
Decouple	74.3	67.6	91.0	0.776
JOF	73.5	67.8	87.2	0.763
PENCL	71.7	65.1	91.0	0.759
COLC-Net	77.2	70.5	91.7	0.797

3 结束语

为了解决乳腺肿瘤超声图像标签不足的问题,

本文针对依据 BI-RADS 评级生成的良性/恶性标签会引起标签噪音的问题进行研究,提出一种处理噪音标签的方法 COLC-Net。由于乳腺超声图像 BI-RADS 评级表示的是肿瘤恶性可能性,天然适合改造为标签分布方式,进行噪音标签问题优化。COLC-Net 利用深度网络的“记忆能力”,即在训练初期优先记住特征明显或容易识别的样本特征,并结合乳腺超声标签噪音率两端低、中间高的分布特点,为乳腺超声的 BI-RADS 评级设计了软标签。另外不同的分类器可以生成不同的决策边界,并具有不同的学习能力。为了对乳腺超声的软标签进行修正, COLC-Net 采用网络协作框架来过滤标签噪音。训练中,每个批次对不同网络的预测值进行选择,采用类似软件工程中常用的“结对编程”方法,从两个网络中蒸馏优秀知识,并动态修正软标签来阻止噪音标签的拟合,强化准确标签的学习。本文方法主要针对乳腺超声图像的 BI-RADS 标签设计,在其他类似医学数据集也有潜在价值。然而已开源相关性较高的乳腺 X 光数据集^[32],只给出了良恶性标签,未给出 BI-RADS 标签。其他开源医学图像,如结肠癌 CT 影像^[33],和 BI-RADS 评级诊断方法明显不同。所以无法采用这些数据集对本文方法进行验证。基于此,在本文收集的数据集上对不同方法进行了对比实验,实验证实了本文方法在处理具有 BI-RADS 标签的乳腺超声图像分类问题上更加高效。

参 考 文 献

- [1] DESANTIS C, MA J, SAUER A G, et al. Breast cancer statistics, 2017, racial disparity in mortality by state[J]. *CA-Cancer J Clin*, 2017, 67(6): 439-448.
- [2] APANTAKU L M, FINCH M D. Breast cancer diagnosis and screening[J]. *Am Fam Physician*, 2000, 62(3): 596-602.
- [3] COSTANTINI M, BELLI P, LOMBARDI R, et al. Characterization of solid breast masses: Use of the sonographic breast imaging reporting and data system lexicon[J]. *J Ultrasound Med*, 2006, 25(5): 649-659.
- [4] LEE J. Practical and illustrated summary of updated BI-RADS for ultrasonography[J]. *Ultrasonography*, 2017, 36(1): 71-81.
- [5] 田文洪, 曾柯铭, 莫中勤, 等. 基于卷积神经网络的驾驶员不安全行为识别[J]. *电子科技大学学报*, 2019, 48(3): 381-387.
TIAN Wen-hong, ZENG Ke-ming, MO Zhong-qin, et al. Recognition of unsafe driving behaviors based on convolutional neural network[J]. *Journal of University of Electronic Science and Technology of China*, 2019, 48(3): 381-387.
- [6] 史豪斌, 徐梦. 基于强化学习的旋翼无人机智能追踪方法[J]. *电子科技大学学报*, 2019, 48(4): 553-559.
SHI Hao-bin, XU Meng. An intelligent tracking method of rotor uav based on reinforcement learning[J]. *Journal of University of Electronic Science and Technology of China*, 2019, 48(4): 553-559.
- [7] 邵杰, 黄茜, 曹坤涛. 基于深度学习的人体解析研究综述[J]. *电子科技大学学报*, 2019, 48(5): 644-654.
SHAO Jie, HUANG Xi, CAO Kun-tao. A review on deep learning techniques applied to human parsing[J]. *Journal of University of Electronic Science and Technology of China*, 2019, 48(5): 644-654.
- [8] BYRA M, GALPERIN M, OJEDA-FOURNIER H, et al. Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion[J]. *Med Phys*, 2019, 46(2): 746-755.
- [9] HUANG Q, ZHANG F, LI X. Machine learning in ultrasound computer-aided diagnostic systems: A survey[J]. *Biomed Res Int*, 2018(5137904): 1-10.
- [10] SUN W, TSENG T L, ZHANG J, et al. Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data[J]. *Comput Med Imaging Graph*, 2017, 57(1): 4-9.
- [11] UNIYAL N, ESKANDARI H, ABOLMAESUMI, et al. Ultrasound RF time series for classification of breast lesions[J]. *IEEE Trans Med Imaging*, 2014, 34(2): 652-661.
- [12] AKSELROD-BALLIN A, KARLINSKY L, ALPERT S, et al. A region based convolutional network for tumor detection and classification in breast mammography[C]//International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis. Athens, Greece: Springer, 2016: 197-205.
- [13] CAO Z, DUAN L, YANG G, et al. Breast tumor detection in ultrasound images using deep learning[C]//International Workshop on Patch-based Techniques in Medical Imaging. Quebec City, Canada: Springer, 2017: 121-128.
- [14] LI Y, YANG J, SONG Y, et al. Learning from noisy labels with distillation[C]//IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 1928-1936.
- [15] MA X, WANG Y, HOULE M E, et al. Dimensionality-driven learning with noisy labels[C]//The 35th International Conference on Machine Learning. Stockholm, Sweden: IMLS, 2018: 3361-3370.
- [16] RODRIGUES F, PEREIRA F C. Deep learning from crowds[C]//The Thirty-Second AAAI Conference on Artificial Intelligence. New Orleans, Louisiana, USA: AAAI, 2018: 1611-1618.
- [17] BROOKS J P. Support vector machines with the ramp loss and the hard margin loss[J]. *Oper Res*, 2011, 59(2): 467-479.
- [18] ROOYEN B V, MENON A K, WILLIAMSON R C. Learning with symmetric label noise: The importance of being unhinged[C]//The Annual Conference on Neural Information Processing Systems. Montréal, Canada: NIPS, 2015: 10-18.
- [19] MASNADI-SHIRAZI H, VASCONCELOS N. On the

- design of loss functions for classification: Theory, robustness to outliers, and SavageBoost[C]//The Annual Conference on Neural Information Processing Systems. Vancouver, British Columbia, Canada: NIPS, 2008: 1049-1056.
- [20] GOLDBERGER J, BEN-REUVEN E. Training deep neural-networks using a noise adaptation layer[C]//International Conference on Learning Representations. Toulon, France: ICLR, 2017.
- [21] PATRINI G, ROZZA A, MENON A K, et al. Making deep neural networks robust to label noise: {a} loss correction approach[C]//IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 2233-2241.
- [22] HAN B, YAO J, NIU G, et al. Masking: A new perspective of noisy supervision[C]//The Annual Conference on Neural Information Processing Systems. Montréal, Canada: NIPS, 2018: 5841-5851.
- [23] HENDRYCKS D, MAZEIKA M, WILSON D, et al. Using trusted data to train deep networks on labels corrupted by severe noise[C]//The Annual Conference on Neural Information Processing Systems. Montréal, Canada: NIPS, 2018: 10477-10486.
- [24] JIANG L, ZHOU Z, LEUNG T, et al. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels[C]//The 35th International Conference on Machine Learning. Stockholm, Sweden: IMLS. 2018: 2309-2318.
- [25] HAN B, YAO Q, YU X, et al. Co-teaching: Robust training of deep neural networks with extremely noisy labels[C]//The Annual Conference on Neural Information Processing Systems. Montréal, Canada: NIPS, 2018: 8536-8546.
- [26] MALACH E, SHALEV-SHWARTZ S. Decoupling ‘when to update’ from ‘how to update’[C]//The 31th Annual Conference on Neural Information Processing Systems. Long Beach, California, USA: NIPS, 2017: 960-970.
- [27] GAO B B, XING C, XIE C W, et al. Deep label distribution learning with label ambiguity[J]. *IEEE Trans Image Process*, 2017, 26(6): 2825-2838.
- [28] TANAKA D, IKAMI D, YAMASAKI T, et al. Joint optimization framework for learning with noisy labels[C]//The IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 5552-5560.
- [29] ARPIT D, JASTRZEBSKI S, BALLAS N, et al. A closer look at memorization in deep networks[C]//The 34th International Conference on Machine Learning. Sydney, Australia: IMLS, 2017: 233-242.
- [30] KUN Y, JIANXIN W. Probabilistic end-to-end noise correction for learning with noisy labels[C]//The IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 7017-7025.
- [31] DENG L. The MNIST database of handwritten digit images for machine learning research[J]. *IEEE Signal Process Mag*, 2012, 29(6): 141-142.
- [32] SUCKLING J, PARKER J, DANCE D, et al. The mammographic image analysis society digital mammogram database[J]. *Int Congr Ser*, 1994, 1069: 375-378.
- [33] JOHNSON C, CHEN M, TOLEDANO A, et al. Accuracy of CT colonography for detection of large adenomas and cancers[J]. *N Engl J Med*, 2008, 359(12): 1207-1217.

编辑 税红