Classification Method of Variable-Temperature Self-distillation Brain Tumor MRI Images Based on Label-Auto-calibration

XU Chudi, HE Hong*, CHEN Jiayu, CHEN Yucong, LI Zexu

(1. School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Brain tumors are one of the most common diseases in the nervous system, and magnetic resonance imaging (MRI) examination is a common screening method for brain tumors. Although deep learning models have achieved high performance in MRI image recognition, their operational effectiveness is unsatisfactory in scenarios with limited computational resources, and their performance tends to decline after pruning for lightweight. Therefore, this study focuses on improving the brain tumor MRI image classification performance of lightweight models. To address the aforementioned challenges, classification method of variable-temperature self-distillation brain tumor MRI images based on label-auto-calibration is proposed. Starting from the core requirement of enhancing the performance of lightweight models, this study employs a loss function based on label-auto-calibration loss to avoid imparting incorrect knowledge from the self-teacher model during the self-distillation process. Additionally, a temperature-resetting variable-temperature distillation mechanism is introduced. When the validation accuracy does not improve in recent rounds of training, the temperature is reset based on the current training situation to prevent the lightweight network from encountering excessive learning difficulty. Experimental analysis on multiple brain tumor datasets demonstrates that the proposed self-distillation method can effectively enhance the brain tumor recognition performance of lightweight models, and its performance surpasses other typical self-distillation methods. This method requires no changes to the model structure, is easy to implement, and can be used to enhance the performance of lightweight networks.

Highlights:

1. A label-auto-calibration loss is adopted to effectively avoid incorrect knowledge transmission from the teacher model during self-distillation, improving the accuracy and reliability of model learning and addressing the key issue of knowledge transfer bias in self-distillation technology.
2. A temperature-resetting variable-temperature distillation mechanism is introduced to mitigate the excessive learning difficulty of lightweight networks, preventing rapid increase in training difficulty and enhancing the stability of model training.
3. The method requires no change to the original structure of the model, is easy to implement, and has strong applicability. It can be directly applied to brain tumor MRI screening in scenarios with limited computational resources.

Key words: brain tumor; MRI image classification; self-distillation; variable temperature distillation; label-auto-calibration

---

Foundation items: Project of the Ministry of Science and Technology (No.G2021013008); China University Industry-University-Research Innovation Fund (No.2023RY011); Medical and Engineering Interdisciplinary Key Innovation Project of University of Shanghai for Science and Technology (No.1022308502); Huawei AI Computing Power Acceleration Program; Key Project of Teacher Development Research at University of Shanghai for Science and Technology (No.CFTD2025ZD08).

Received: 2025-10-31; **Revised**: 2026-04-07

***Corresponding author**, **E-mail**: hehong@usst.edu.cn.

ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 41,No. 4,Jul. 2026,pp. 1178－1193
DOI:10. 16337/j. 1004-9037. 2026. 04. 018

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于标签自动校准的变温自蒸馏脑肿瘤MRI图像分类方法

徐楚迪，何　宏，陈家毓，陈宇聪，李泽旭

（上海理工大学健康科学与工程学院，上海 200093）

摘　要：脑肿瘤是神经系统中最常见的疾病之一，磁共振成像（Magnetic resonance imaging，MRI）检查是常见的脑肿瘤筛查方法。深度学习模型在MRI图像识别上虽已具备较高性能，但在计算资源受限场景下运行效果不理想，且其性能在剪枝轻量化后易出现下降，为此本研究聚焦于提高轻量化模型的脑肿瘤MRI图像分类性能，针对上述挑战提出一种基于标签自动校准的变温自蒸馏脑肿瘤MRI图像分类方法。研究从轻量化模型性能提升的核心需求出发，使用基于标签自动校准的损失函数，避免自蒸馏过程中自教师模型传授错误知识，同时引入一种温度重启的变温蒸馏机制，当验证准确率在最近多轮训练中未提升时，根据当前训练情况重启温度，防止轻量化网络学习难度过大。在多个脑肿瘤数据集上的实验分析结果表明，所提出的自蒸馏方法可有效提高轻量化模型的脑肿瘤识别性能，且性能优于其他典型的自蒸馏方法，具有无需改变模型结构、易于实现的优点。

关键词：脑肿瘤；MRI图像分类；自蒸馏；变温蒸馏；标签自动校准

中图分类号：R445.2;TP391.4　　**文献标志码：**A

引用格式：徐楚迪，何宏，陈家毓，等. 基于标签自动校准的变温自蒸馏脑肿瘤MRI图像分类方法[J]. 数据采集与处理，2026，41(4)：1178-1193. XU Chudi，HE Hong，CHEN Jiayu，et al. Classification method of variable-temperature self-distillation brain tumor MRI images based on label-auto-calibration[J]. Journal of Data Acquisition and Processing，2026，41(4)：1178-1193.

引　　言

脑肿瘤是神经系统中最常见的疾病之一，近年来，其病例数量急剧攀升，成为全球最具致命性的疾病之一[1]。到目前为止，已经发现了超过一百多种类型的肿瘤[2]，并且肿瘤的形状尺寸存在差异[3]，加重了脑肿瘤图像的分类难度。磁共振成像（Magnetic resonance imaging，MRI）检查被广泛应用于检查脑肿瘤的情况，是脑肿瘤分类中最常用且有效的检查技术[4]。

目前，深度学习中图像分类模型能够自动提取MRI图像特征，高效准确地识别脑肿瘤的类型和位置[5]。随着深度学习的持续发展，出现了很多更深更复杂的模型，在脑肿瘤MRI图像的识别上具有更好的性能，但是更复杂的模型对计算资源和存储容量也有更高的要求。当面临计算资源受限的情况时，需要对模型进行轻量化处理。轻量化的方法包含剪枝和知识蒸馏。模型过大，会导致计算量和存储需求过高，剪枝可以有效去除冗余参数和连接，减小模型规模[6]。但是如果对网络过度剪枝，可能会导致性能上的下降[7]。知识蒸馏可将教师模型的知识传递给学生模型，提升学生模型性能，作为弥补性能的有效手段。知识蒸馏按照蒸馏方法可分为离线蒸馏、在线蒸馏和自蒸馏3种方法[8]。前两种方法

基金项目：国家科学技术部项目（G2021013008）；中国高校产学研创新基金（2023RY011）；上海理工大学医工交叉重点创新项目（1022308502）；华为AI算力加速计划项目；上海理工大学教师发展研究重点项目（CFTD2025ZD08）。
收稿日期：2025-10-31；**修订日期：**2026-04-07

需要训练一个参数量较大的教师模型,仍会消耗大量计算资源及存储容量。并且,虽然教师模型是性能更加优越的模型,但是也可能会出现预测错误的情况,将错误信息传授给学生模型。在蒸馏过程中,通常是使用固定温度进行软化预测概率,这难以适应模型在不同训练阶段的情况。

对此,为了解决上述问题,本文融合了自蒸馏与动态温度技术,提出了一种基于标签自动校准的变温自蒸馏脑肿瘤MRI图像分类方法,以解决计算资源受限情况下脑肿瘤MRI图像准确识别问题。首先针对传统自蒸馏过程中自教师模型可能出现预测错误的情况,提出了基于标签自动校准的损失函数,使用软化的标签用于监督自教师预测错误的样本,避免自教师模型将错误知识传递给学生模型。同时,针对变温蒸馏过程中温度参数单调递减导致训练难度调整过快过难的问题,本文提出了温度重启的变温蒸馏机制,在降低蒸馏温度中,根据验证性能的提升情况进行提高温度,控制训练的难度。使用此自蒸馏方法,有效提高了轻量化模型的性能。

1 相关工作

1.1 自蒸馏

传统的知识蒸馏方法需要一个性能优越的教师模型来引导学生模型的学习。然而,自教师模型的构建与选择往往面临诸多挑战,如训练成本高昂、模型架构适配性等问题。针对这些问题,自蒸馏这一知识蒸馏领域的重要分支在近年来受到了广泛关注。自蒸馏大致可分为基于特征、基于架构分支和基于响应这3种类型。

1.1.1 基于特征的自蒸馏策略

基于特征的自蒸馏策略注重于网络中间层特征之间的学习。Zhang等[9]让神经网络的每个浅层块逐步从其相邻的深层块中学习知识,构建自适应通道注意力模块,增强模型对关键特征的关注。Sharma等[10]引入类别亲和图来捕捉数据实例与类别中心之间的关系,对齐自教师模型和学生模型的类别亲和图的微结构,增强学生模型对不同类别实例的区分能力。张睿等[11]结合多级特征知识蒸馏损失函数,提升了模型对图像判别性区域特征的学习能力。唐媛等[12]针对中间层特征设计掩膜分割模块和动态注意力分配策略,挖掘主体与背景知识并引导其协作,优化分类器对图像的关注。Zhang等[13]将网络分割成多个部分,将深层部分的知识提取到浅层。Zhao等[14]通过掩膜生成模块对中间特征图随机掩膜并经过生成器重构,再以自适应加权机制按置信度调监督权重,结合双损失优化对齐中间特征知识。Zhao等[15]通过引入自学习掩码模块保留前一轮训练结果,并对当前轮特征图应用掩码机制来增强特征提取。此类自蒸馏方法都需要将不同的特征层之间进行对齐,具有设计架构复杂、不同尺度特征之间难以对齐的缺点。

1.1.2 基于架构分支的自蒸馏策略

基于架构分支的自蒸馏策略在网络中设置并行分支实现知识传递,以性能更优的分支为教师,指导其他辅助分支学习。倪水平等[16]设计了有深层分类器和浅层分类器的自蒸馏框架,通过辅助分支学习主分支知识,强化主分支性能。Luan等[17]通过构建多分类器分支架构实现分支间协同蒸馏,各分支互为师生,监督输出响应并对齐中间特征。此类自蒸馏方法需要对不同特征层设计多个分类器,但是全连接层的参数量较大,同样具有架构复杂、训练推理开销大的缺点,有悖于轻量化模型节约计算资源。

1.1.3 基于响应的自蒸馏策略

基于响应的自蒸馏策略注重于优化网络的输出。李圣巍等[18]设计了包含学生模型、标签编码器和交叉注意力模块这3个部分的框架,通过标签编码器生成的软损失和硬损失同步更新参数。苏楠等[19]

将自蒸馏技术、EMA算法和迁移学习技术结合，提升了输出层响应的可靠性。Kim等[20]通过逐步调整模型的目标向量，利用模型自身的知识来改善模型的泛化能力。在训练过程中，将模型过去的预测与实际标签结合起来，生成更平滑、更有效的目标向量。Shen等[21]约束每批一半样本与前一迭代一致以重排抽样，同一模型用前一迭代重叠样本生成的软分类概率，对齐当前迭代重叠样本输出。此类自蒸馏方法有架构简单、计算量小的优点，但是对轻量化模型性能的提升效果通常不如前两类自蒸馏方法。

1.2 变温蒸馏

在知识蒸馏中，通常会提前设定好一个固定的温度值，但这难以适应模型在不同训练阶段的情况。近年来，研究者们逐渐认识到温度参数在知识蒸馏中的重要性，并开始探索改变温度的方法。变温蒸馏通过设置一个合适的改变温度参数的方法来引导学生模型由易到难地进行学习。Wei等[22]引入尖锐度指标，量化模型输出分布的平滑度，通过最小化教师和学生模型之间的尖锐度差异，为每个样本动态地生成适合的温度。调整教师和学生模型的温度，使得两者的输出平滑度达到一个适度的共同点。Jafari等[23]按照余弦退火方式通过逐步调整教师模型输出的软标签温度，使学生模型能够逐步、平滑地学习教师模型的知识。汪瑞等[24]通过构建全局和局部温度拟合函数并结合残差卷积温度预测网络，动态生成每个样本的蒸馏温度。Li等[25]通过动态调整温度来控制知识蒸馏过程中任务的难易程度，采用对抗式学习机制调整温度参数，使学生模型逐步从简单任务学习到复杂任务。

2 本文方法

基于标签自动校准的变温自蒸馏方法的训练框架包含3个部分，如图1所示。此方法以轻量化网络为学生模型，可以有效提高轻量化网络的性能，主要包括自教师模型更新模块、基于标签自动校准的损失函数和温度重启的变温蒸馏机制。在训练过程中，自教师模型更新模块在线动态地更新自教师模型，确保自教师模型性能始终优于学生模型。同时使用基于标签自动校准的损失函数，利用标签监督自教师模型预测错误的样本，避免传统蒸馏过程中自教师模型向学生模型传递错误知识。此外，在变温蒸馏过程中，使用温度重启的变温蒸馏机制，根据验证性能的提升情况来调整温度，控制训练难度以适应学生模型当前性能，避免单调降温导致的训练难度提升过快过难情况。

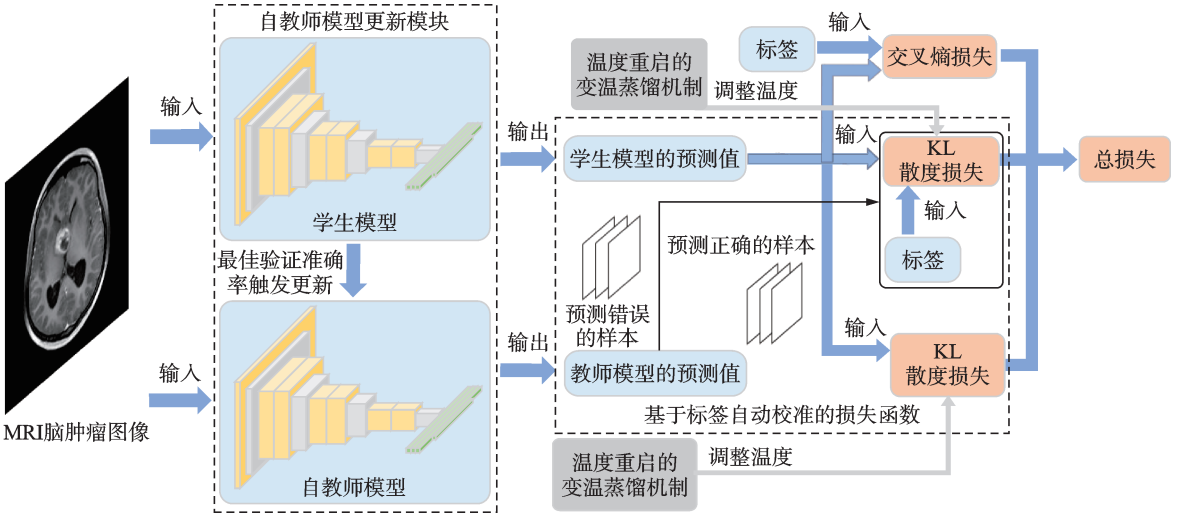

图1 基于标签自动校准的变温自蒸馏方法的训练框架

Fig.1 Training framework for variable temperature self distillation method based on label-auto-calibration

2.1 学生模型与自教师模型

此自蒸馏方法选择轻量化模型为学生模型,自教师模型与学生模型结构相同。本文采用以验证性能驱动的教师模型更新策略实现教师模型的更新[26]。设训练集中的输入图像为$\boldsymbol{X}$,轻量化模型为$M(\theta)$,模型参数为θ,模型输出的预测值为p。在自蒸馏训练过程中,动态地更新自教师模型,以确保自教师模型的泛化性能比学生模型更好。在每一轮epoch训练中,如果学生模型的验证准确率$\mathrm{Acc_{val}}(\mathrm{epoch})$超过历史最佳的验证准确率$\mathrm{BestAcc_{val}}$时,则用学生模型$M(\theta_s)$在第epoch轮次的参数$\theta_s(\mathrm{epoch})$更新为自教师模型$M(\theta_t)$的参数$\theta_t$。此步骤可表示为

$$\mathrm{BestAcc_{val}}=\max(\mathrm{BestAcc_{val}},\mathrm{Acc_{val}}(\mathrm{epoch})) \tag{1}$$

$$\theta_t=\begin{cases}\theta_s(\mathrm{epoch}) & \mathrm{Acc_{val}}(\mathrm{epoch})>\mathrm{BestAcc_{val}}\\ \theta_t & \mathrm{Acc_{val}}(\mathrm{epoch})\leqslant\mathrm{BestAcc_{val}}\end{cases} \tag{2}$$

2.2 基于标签自动校准的损失函数

基于响应的自蒸馏利用网络自身的输出作为监督信号,指导网络训练,从而提升泛化能力,通常使用如下损失函数

$$\mathrm{Loss}=\alpha\cdot\mathrm{CELoss}(p_s,\mathrm{label})+(1-\alpha)\cdot T^2\cdot\mathrm{KLDivLoss}(p_s,p_t,T) \tag{3}$$

式中:CELoss代表交叉熵损失,α为硬损失交叉熵损失的权重,KLDivLoss代表KL散度损失,$1-\alpha$为软损失KL散度损失的权重,p_s为当前学生模型的预测概率分布,label为真实标签,p_t为自教师模型的预测概率分布,T为蒸馏温度。然而,以式(3)作为损失函数时,自教师模型会将自身预测错误样本的概率分布传授给学生模型,不利于学生模型学习到正确的知识。

因此,本文提出了基于标签自动校准的损失函数(Label-auto-calibration loss,LACLoss),以避免自教师模型传递错误知识。当自教师模型的预测类别与真实标签一致时,保留其软标签作为监督信号;当预测类别与真实标签不一致时,使用真实标签替代软标签。使用自教师模型的预测值和真实标签共同实现自蒸馏的软损失,避免自教师模型传授错误概率分布。新的损失函数总体可表示为

$$\mathrm{Loss}=\alpha\cdot\mathrm{CELoss}(p_s,\mathrm{label})+(1-\alpha)\cdot\mathrm{LACLoss}(p_s,p_t,\mathrm{label},T) \tag{4}$$

其中CELoss仍然作为硬损失,LACLoss作为软损失。LACLoss由自教师模型的预测概率和真实标签共同监督学生模型的预测概率,可具体表示为

$$\mathrm{LACLoss}(p_s,p_t,\mathrm{label},T)=\begin{cases}T^2\cdot\mathrm{KLDivLoss}(p_s,p_t,T) & \mathrm{argmax}(p_t)=\mathrm{label}\\ T^2\cdot\mathrm{KLDivLoss}(p_s,\mathrm{label}^{\mathrm{onehot}},T) & \mathrm{argmax}(p_t)\neq\mathrm{label}\end{cases} \tag{5}$$

式中$\mathrm{label}^{\mathrm{onehot}}$为独热编码标签。针对自教师模型预测正确的样本,使用软化的自教师模型的输出来监督学生模型的输出,使用KL散度损失计算软化的学生模型预测值与自教师模型预测值之间的差异。针对自教师模型预测错误的样本,使用真实标签对学生模型的输出进行监督,防止自教师模型传授错误信息。首先将每一批次的独热编码$\mathrm{label}^{\mathrm{onehot}}$映射到该批次样本学生模型响应logit的取值范围$[\mathrm{logit_{min}},\mathrm{logit_{max}}]$内,使独热编码更加符合模型输出分布,适配批次内的数据分布,即

$$\mathrm{label}^{\mathrm{onehot}}=\mathrm{logit_{min}}+(\mathrm{logit_{max}}-\mathrm{logit_{min}})\times\mathrm{label}^{\mathrm{onehot}} \tag{6}$$

式中:$\mathrm{label}^{\mathrm{onehot}}$为独热编码标签,$\mathrm{logit_{min}}$为该批次样本学生模型响应的最小值,$\mathrm{logit_{max}}$为该批次样本学生模型响应的最大值。随后使用蒸馏温度T软化独热编码$\mathrm{label}^{\mathrm{onehot}}$,具体可表示为

$$\mathrm{label}^{\mathrm{onehot}}(\mathrm{class}_j,T)=\mathrm{softmax}(\mathrm{label}_j^{\mathrm{onehot}},T)=\frac{\mathrm{e}^{\mathrm{label}_j^{\mathrm{onehot}}/T}}{\sum_{c=1}^{n}\mathrm{e}^{\mathrm{label}_c^{\mathrm{onehot}}/T}} \tag{7}$$

式中：$\text{label}_j^{\text{onehot}}$为独热编码第$j$个类别的值，$\text{label}^{\text{onehot}}(\text{class}_j, T)$为独热编码第$j$个类别经过蒸馏温度$T$软化后的软标签，softmax为软最大化激活函数，T为蒸馏温度。对真实标签软化时采用与教师模型输出相同的蒸馏温度T，以保证教师软标签与软化真实标签的分布锐度一致。然后使用KL散度损失计算学生模型预测值与软化的真实标签之间的差距，具体可表示为

$$\text{KLDivLoss}(p_s, \text{label}^{\text{onehot}}, T) = \sum_j p_s(\text{class}_j, T) \cdot \log \frac{p_s(\text{class}_j, T)}{\text{label}^{\text{onehot}}(\text{class}_j, T)} \tag{8}$$

式中：class_j表示第j个类别，$p_s(\text{class}_j, T)$为学生模型第j个类别的经过蒸馏温度T软化的预测概率值。通过基于标签自动校准的损失函数，能够精确辨别自教师模型预测的正样本与负样本。对正确预测的样本，保留其软标签所具备的引导价值；对错误预测的样本，则用真实标签的介入来阻断错误知识的传播，既留存了自蒸馏中软标签所蕴含的丰富的类别间关系信息，又依托真实标签达成了知识的纠错与过滤。

2.3 温度重启的变温蒸馏机制

在变温蒸馏中，通常设置一个较大的初始温度值T_{inital}，让学生模型在性能较差的训练初期更易学习到整体平滑的类别概率分布。在训练中，逐渐降低蒸馏的温度，突出不同类别间的概率差异，增加训练难度。但若温度下降过快过低，模型的输出概率分布会变得过于尖锐，概率分布类似于硬标签，失去软化概率分布的功能。模型可能难以适应训练难度的快速增加，影响训练效果。对此，本文提出在变温蒸馏过程中引入了温度重启机制，所提温度重启的变温蒸馏机制如图2所示。

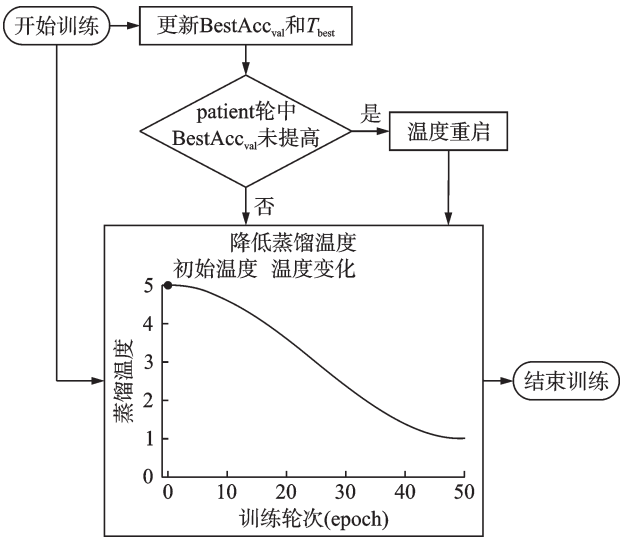

图2 温度重启的变温蒸馏机制

Fig.2 Variable temperature distillation mechanism for temperature restart

本文以Jafari等[23]提出的余弦退火降温方式为基础，在此之上加入温度重启机制，余弦退火降温的温度可表示为

$$T_{\text{next}} = 1 + \frac{T_{\text{initial}} - 1}{2}\left(1 + \cos\left(\pi \frac{\text{epoch}_{\text{current}}}{\text{epoch}_{\text{total}}}\right)\right) \tag{9}$$

式中：T_{initial}为设置的初始最大温度；T_{next}为下一轮训练的蒸馏温度，最低蒸馏温度为1；$\text{epoch}_{\text{current}}$为当前的训练轮次；$\text{epoch}_{\text{total}}$为此降温阶段总的训练轮次。若当前训练轮次的验证准确率$\text{Acc}_{\text{val}}(\text{epoch})$超过历史最佳的验证准确率$\text{BestAcc}_{\text{val}}$，当前的蒸馏温度$T_{\text{current}}$被更新为最佳蒸馏温度$T_{\text{best}}$，具体可表示为

$$T_{\text{best}} = \begin{cases} T_{\text{current}} & \text{Acc}_{\text{val}}(\text{epoch}) > \text{BestAcc}_{\text{val}} \\ T_{\text{best}} & \text{Acc}_{\text{val}}(\text{epoch}) \leqslant \text{BestAcc}_{\text{val}} \end{cases} \tag{10}$$

在降温中，会产生新的最佳验证准确率$\text{BestAcc}_{\text{val}}$和最佳蒸馏温度$T_{\text{best}}$。在训练过程中，若最佳验证准确率$\text{BestAcc}_{\text{val}}$在耐心值patient轮训练中未发生提高，则进行温度重启。将新的余弦退火周期起点记为$\text{epoch}_{\text{new}}$，把最新的最佳蒸馏温度$T_{\text{best}}$作为重启温度。将温度提升至重启温度后，重启后的温度T_{best}成为新一个降温阶段的初始温度T_{inital}，重新开启降温过程，降温周期仍然为$\text{epoch}_{\text{total}}$。若在新的降温周期内，最佳验证准确率$\text{BestAcc}_{\text{val}}$在耐心值patient轮训练过程中未提高，则再进行温度重启操作，进

行新一轮的降温,此过程可表示为

$$\text{epoch}_{\text{new}} = \text{epoch}_{\text{BestAccval}} + \text{patient} \tag{11}$$

$$T_{\text{next}} = 1 + \frac{T_{\text{best}} - 1}{2}\left(1 + \cos\left(\pi \frac{\text{epoch}_{\text{current}} - \text{epoch}_{\text{new}}}{\text{epoch}_{\text{total}}}\right)\right) \tag{12}$$

通过重启提高温度,可再次软化自教师模型输出的概率和标签的分布,降低当前学生模型的训练难度,以适应当前学生模型的学习情况。

3 实验分析与结果

3.1 数据集与预处理

数据集1来源于https://www.kaggle.com/datasets/hellojahid/brain-tumor-classification-15-classes。此数据集为脑肿瘤数据集,涵盖了15个类别的脑肿瘤图像,所有图像均为.jpeg文件格式,其中训练集3 362张图像,测试集927张图。调整数据集的每张图像形状为256×256,并把像素值从0至255范围归一化到0至1范围。从训练集中按照0.2的比例把验证集划分出来,划分随机种子为42,共有673张图像。训练集采用随机旋转15°和水平翻转进行数据增强,结合加权采样机制以应对类别不平衡。

数据集2来源于https://www.kaggle.com/datasets/fernando2rad/brain-tumor-mri-images-44c。此数据集为脑肿瘤数据集,共有4 480张图像,涵盖了44个类别的脑肿瘤图像,所有图像均为.jpeg文件格式。对数据集的每张图像调整形状为256×256,并把像素值从0至255范围归一化到0至1范围。按照8:1:1的比例划分训练集、验证集和测试集,划分随机种子为42。训练集采用随机旋转15°和水平翻转进行数据增强,结合加权采样机制以应对类别不平衡。

数据集3来源于https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri。此数据集为脑肿瘤数据集,分为测试集394张图像和训练集2 870张图像。其中测试集共有神经胶质瘤100张、脑膜瘤115张、无肿瘤105张、垂体瘤74张;训练集共有神经胶质瘤826张、脑膜瘤822张、无肿瘤395张、垂体瘤827张。将数据集的每张图像调整形状为256×256,并把像素值从0至255范围归一化到0至1范围。从训练集按照0.2的比例把验证集划分出来,划分随机种子为42。训练集采用随机旋转15°和水平翻转进行数据增强,结合加权采样机制以应对类别不平衡。

3.2 实验参数与评价指标

使用Python 3.9版本在GPU加速环境中进行实验,采用Pytorch框架2.0.0 cuda版本搭建神经网络。电脑配置为Windows11系统、64 GB内存、NVIDIA RTX Geforce3090显卡、24 GB显存、CUDA版本为12.8。实验过程中训练周期为100epochs,使用Adam优化器,学习率为0.01,批大小为64,交叉熵损失函数的权重α为0.5。选择在训练过程中验证性能最好的网络在测试集上进行测试,并使用宏平均精度(Precison)、召回率(Recall)、F_1分数(F_1 score)和准确率(Accuracy)指标来评估网络性能。指标计算公式分别为

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{13}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{14}$$

$$F_1\ \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{15}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{16}$$

3.3 消融实验

为了验证所提自蒸馏方法变温机制和损失函数的有效性,本文设计了一系列的自蒸馏消融实验。以经典的MobileNet系列轻量化网络为基准模型,在此基础上加上了基于标签自动校准的损失函数和温度重启的变温蒸馏机制,并单独验证这些改动对模型性能的影响。具体实验分组如表1所示。

表1 消融实验分组与配置

Table 1 Grouping and configuration of ablation experiments

实验编号	实验配置	控制变量
①	基准模型训练	无
②	①+基于标签自动校准的损失函数	使用基于标签自动校准的损失函数为软损失,移除了温度重启的变温蒸馏机制,固定蒸馏温度为5。
③	①+变温蒸馏	移除了温度重启的变温蒸馏机制,按照式(9)进行降低蒸馏温度。使用式(3)作为损失函数,初始蒸馏温度为5,$epoch_{total}$为100。
④	①+温度重启的变温蒸馏机制	包含温度重启的变温蒸馏机制,温度重启的耐心值patient设置为5。使用式(3)的常规自蒸馏损失函数,初始蒸馏温度为5。
⑤	①+所提的自蒸馏方法	包含基于标签自动校准的损失和温度重启的变温蒸馏机制。

消融实验结果如表2(消融实验结果)、图3(消融实验性能雷达图)和图4(消融实验性能雷达图的面积)所示。通过对比实验编号①、②、④的消融实验性能雷达图,可发现所提的基于标签自动校准的损失函数和温度重启的变温蒸馏机制均可以有效提高模型的性能。对比实验编号③和④的性能,可发现温度重启后的变温蒸馏可以帮助模型缓解训练难度,防止难度过高,更易于模型训练。对比实验编号②、④、⑤的性能,将基于标签自动校准的损失和温度重启的变温蒸馏机制结合,可以进一步提高模型的性能。为了更加清晰地观察差异,计算了如图4所示的消融实验性能雷达图的面积,面积越大则代表性能更优。

3.4 对比实验

为了验证所提方法的先进性,本文设计了一系列自蒸馏对比实验,以MobileNet轻量化网络为基准学生模型,对比多种现有的自蒸馏方法,所有对比方法均采用对应原文一致的参数进行训练测试,具体实验分组如表3所示。

对比实验结果如表4(对比实验结果)、图5(对比实验性能雷达图)和图6(对比实验性能雷达图的面积)所示。所提出的自蒸馏方法对应的性能雷达图面积最大,表明所提出的自蒸馏方法更能全面地提升轻量化网络的多项性能指标,相对于其他自蒸馏方法能够更加有效地提升轻量化网络的综合性能。

4 讨 论

从整体上看,本文提出的自蒸馏策略在多个脑肿瘤MRI数据集,均能有效地提升轻量化模型的综合表现能力,且具有不用改变模型结构的优点,因而具有良好的通用性与适应性,能够便捷地应用到多种轻量化模型当中,以提升轻量化模型性能。

表2 消融实验结果

Table 2 Ablation experiment results

%

模型	实验编号	数据集1				数据集2				数据集3			
		精度	召回率	F_1分数	准确率	精度	召回率	F_1分数	准确率	精度	召回率	F_1分数	准确率
MobileNetV1	①	90.675 0	87.930 1	89.038 9	90.938 5	88.139 3	83.635 4	84.231 3	88.839 3	79.156 2	68.563 3	65.337 0	70.304 6
	②	92.600 2	88.032 1	89.741 5	92.448 8	91.524 0	88.169 7	88.795 9	92.857 1	80.197 1	72.418 2	69.469 5	73.604 1
	③	92.088 5	89.055 6	90.270 4	92.448 8	92.419 5	87.521 6	88.656 0	93.080 4	82.358 1	72.582 8	69.997 0	73.350 3
	④	95.233 9	89.212 5	91.599 9	93.096 0	91.895 8	89.099 3	89.321 2	93.526 8	82.153 3	72.876 5	69.831 8	73.857 9
	⑤	94.485 7	89.985 9	91.819 8	93.743 3	94.598 2	92.101 8	92.081 5	94.419 6	80.923 5	74.924 2	71.149 6	75.380 7
MobileNetV2	①	88.444 7	81.525 4	83.650 4	88.349 5	85.731 2	82.695 2	82.699 5	88.169 6	79.231 1	68.310 8	65.747 8	70.558 4
	②	91.980 2	85.528 8	87.937 6	91.046 4	89.691 1	85.067 8	86.407 6	91.741 1	81.024 7	74.544 7	70.141 8	74.873 1
	③	89.459 3	87.178 5	87.829 5	91.370 0	89.796 6	85.894 6	86.615 6	91.964 3	82.139 7	73.475 0	69.599 8	74.365 5
	④	91.625 4	87.238 9	88.860 7	91.585 8	91.617 6	88.343 1	89.117 9	93.080 4	80.047 1	75.648 8	70.530 8	75.380 7
	⑤	91.101 4	88.461 9	89.660 6	91.909 4	91.708 3	89.028 5	89.662 9	93.973 2	83.410 0	75.762 9	73.425 3	75.888 3
MobileNetV3-Small	①	85.678 7	85.068 2	85.059 8	88.457 4	84.384 5	79.284 6	80.715 2	84.598 2	75.940 6	70.532 6	65.184 0	71.319 8
	②	88.418 7	86.174 5	87.116 5	90.399 1	88.814 6	84.529 9	85.120 5	90.401 8	79.700 1	72.534 6	68.854 3	73.350 3
	③	87.060 2	86.842 3	86.352 0	89.751 9	89.167 4	83.522 8	84.865 4	89.285 7	83.012 5	71.187 3	68.739 0	72.588 8
	④	87.810 7	86.802 7	86.975 7	90.507 0	91.473 9	89.117 1	88.911 4	91.517 9	80.981 9	72.360 2	69.675 5	73.857 9
	⑤	87.638 2	86.331 0	86.850 0	91.370 0	92.881 1	89.223 0	90.160 1	93.750 0	82.237 8	74.283 1	70.473 1	74.873 1
MobileNetV3-large	①	85.082 0	81.925 2	82.888 7	86.192 0	79.026 9	77.599 5	76.556 7	82.589 3	77.954 6	68.912 2	65.169 7	70.812 2
	②	89.089 8	85.420 7	86.667 8	90.183 4	89.022 4	85.845 5	86.382 8	90.401 8	79.622 7	72.693 6	69.233 1	73.604 1
	③	87.623 7	87.143 5	87.274 6	90.830 6	89.760 3	83.544 8	84.628 4	89.955 4	77.907 2	72.458 0	67.706 8	72.842 6
	④	89.428 2	86.650 6	87.862 4	91.046 4	88.654 6	88.416 1	87.523 1	91.071 4	82.864 8	74.283 1	70.641 7	74.873 1
	⑤	89.466 4	88.097 4	88.529 0	91.477 9	91.607 8	85.190 9	86.577 3	92.410 7	83.228 1	75.579 2	72.391 6	75.888 3

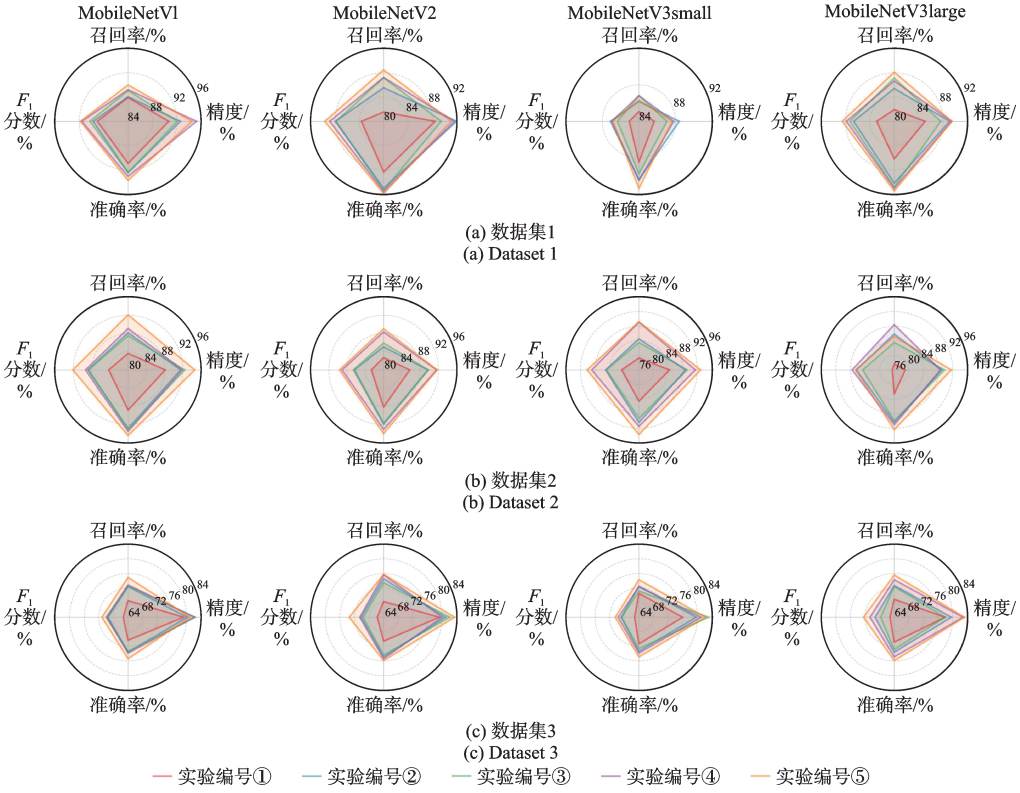

图 3 消融实验性能雷达图

Fig.3 Radar chart of ablation experiment performance

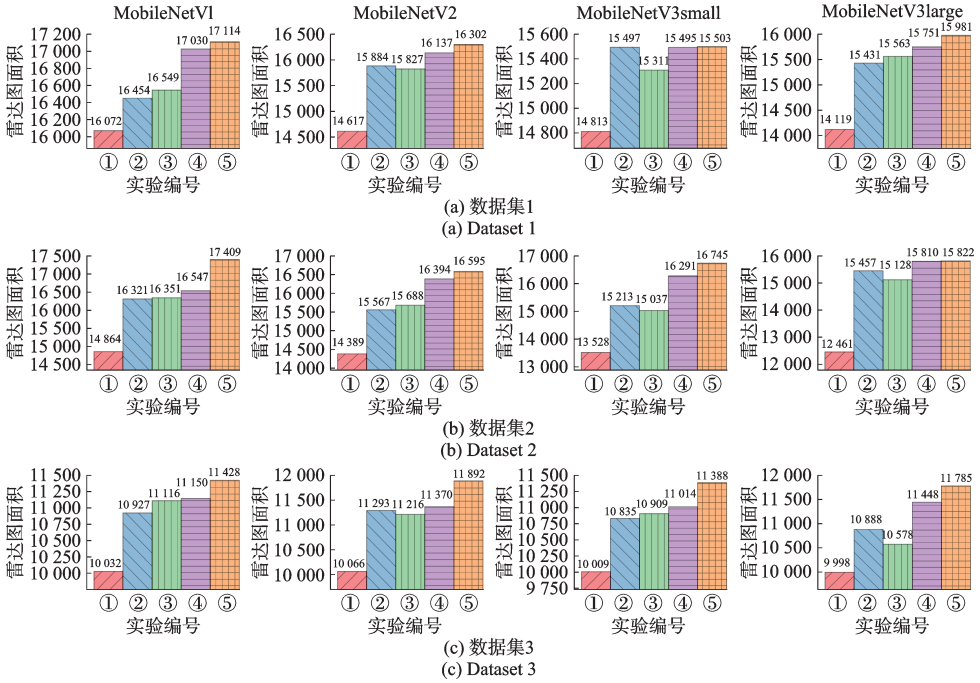

图 4 消融实验性能雷达图的面积

Fig.4 Area of radar chart for ablation experiment performance

表3 对比实验分组与配置

Table 3 Comparative experimental grouping and configuration

实验编号	实验配置	具体方法
①	所提自蒸馏方法	结合所提温度重启的变温蒸馏机制和基于标签自动校准的损失函数,对网络进行自蒸馏。
②	SDL自蒸馏框架	把基于响应的SDL自蒸馏框架与基准模型结合[18]。
③	EMA自蒸馏	使用预训练好的轻量化网络为基准模型,将基于响应的EMA与轻量化网络结合进行自蒸馏[19]。
④	ECASKD	使用轻量化网络为基准模型,使用基于特征的ECASKD方法进行自蒸馏[11]。
⑤	MSD	使用轻量化网络为基准模型,使用基于架构分支的MSD方法进行自蒸馏[17]。
⑥	Epoch-Wise-SD	使用轻量化网络为基准模型,使用基于特征的Epoch-Wise-SD方法进行自蒸馏[15]。
⑦	GSD	使用轻量化网络为基准模型,使用基于特征的GSD方法进行自蒸馏[9]。

结合表2对比4个轻量化模型的消融实验结果,可发现性能提升存在一定差异。MobileNetV1、MobileNetV2基础特征提取能力基础扎实,在引入所提的自蒸馏方法后整体稳定提升3%至5%,性能平滑收敛;而MobileNetV3small、MobileNetV3large基础模型性能较差,引入所提的自蒸馏方法后,有效提高模型的性能,实现5%至9%提升。除模型差异外,3个数据集在类别数量、样本数量及病灶类型上也存在差异。数据集1包含15个肿瘤类别,类别数量较多但单类样本分布相对均衡,模型在该数据集上通过自蒸馏获得3%至5%的性能提升。数据集2包含44个细粒度肿瘤类别,类别数量最多、分类难度最高,且部分罕见肿瘤样本量有限,自蒸馏方法对特征学习与泛化能力的强化效果更为明显,性能提升达到6%至9%。数据集3仅包含神经胶质瘤、脑膜瘤、垂体瘤与无肿瘤4个类别,但神经胶质瘤与脑膜瘤的影像特征高度相似,易产生混淆,提升了5%左右的性能。

但是此自蒸馏方法也存在一定的局限性。首先,此方法在模型训练过程中,由于需要同时维护学生模型和自教师模型的参数更新,相比传统的单模型训练,也会增加一定的计算资源消耗和训练时间成本。现实中的MRI多以三维图像进行存储与分析,难以适配三维MRI图像数据和临床需求。其次,此方法只针对脑肿瘤MRI图像这一单模态进行自蒸馏,未结合脑肿瘤相关的临床医学先验知识。从图7的结果可发现,使用所提方法在4个模型中对神经胶质瘤类别均存在明显的分类短板,以MobileNetV3large模型为例,真实标签为神经胶质瘤的样本中仅22例被正确分类,31例被误判为脑膜瘤,46例被误判为无肿瘤,正确分类率不足25%。这一现象表明,神经胶质瘤与脑膜瘤在影像纹理、灰度分布等视觉特征上高度相似,导致难以捕捉脑肿瘤的独特特征。所提方法在部分脑肿瘤类别的识别性能上存在不足。从表4的对比实验结果可见,尽管所提自蒸馏方法的综合能力优于其他自蒸馏方法,但在数据集2针对MobileNetv3large的对比实验中,其召回率低于实验编号②④⑥,这表明该方法对部分小众且难以区分的肿瘤类别的识别能力仍有欠缺。

未来工作将重点围绕这几个问题展开改进:首先,针对计算资源消耗和训练时间成本增加的问题,探索更高效的模型参数更新策略,在保证学生模型和自教师模型学习效果的前提下,减少不必要的参数更新次数或降低更新频率,使其有望应用于三维卷积神经网络处理三维MRI图像数据的分类任务;其次,为了增强对脑肿瘤MRI分析的医学逻辑性和解释性,可通过构建脑肿瘤类型与特点的临床知识库,通过知识图谱方法将知识融入到自蒸馏过程当中,引导模型在自蒸馏过程中关注更具医学意义的

表 4 对比实验结果

Table 4 Comparative experimental results

%

模型	实验编号	数据集 1				数据集 2				数据集 3			
		精度	召回率	F1分数	准确率	精度	召回率	F1分数	准确率	精度	召回率	F1分数	准确率
MobileNetV1	①	**94.485 7**	**89.985 9**	**91.819 8**	**93.743 3**	**94.598 2**	**92.101 8**	**92.081 5**	**94.419 6**	**80.923 5**	**74.924 2**	**71.149 6**	**75.380 7**
	②	93.058 3	89.052 3	90.622 8	92.522 3	91.463 9	87.035 8	87.351 6	92.187 5	74.787 8	74.292 2	69.282 0	73.437 5
	③	90.758 5	87.866 7	89.015 1	91.046 4	87.818 6	85.641 7	85.706 9	90.848 2	80.394 1	69.859 4	67.098 4	71.319 8
	④	93.498 0	88.323 4	90.353 1	91.629 5	95.394 8	91.174 1	92.404 1	93.526 8	79.743 8	75.067 4	71.431 9	74.218 8
	⑤	91.346 1	86.971 6	88.546 5	91.693 6	88.975 7	84.554 5	85.506 5	90.625 0	73.568 2	72.351 7	67.054 7	71.354 2
	⑥	92.943 7	89.468 7	90.852 7	92.017 3	88.920 3	86.800 8	87.205 2	90.862 9	80.973 4	74.558 5	69.676 2	74.873 1
	⑦	90.232 9	86.209 5	87.988 4	89.644 0	88.618 9	84.020 1	84.696 9	90.609 1	78.943 5	71.533 5	68.341 7	72.081 2
MobileNetV2	①	**91.101 4**	**88.461 9**	**89.660 6**	**91.909 4**	**91.708 3**	**89.028 5**	**89.662 9**	**93.973 2**	**83.410 0**	**75.762 9**	**73.425 3**	**75.888 3**
	②	90.185 4	86.717 2	87.890 2	91.294 6	92.764 1	87.774 6	89.297 2	92.857 1	78.022 1	73.020 4	68.225 2	73.177 1
	③	89.831 5	84.114 6	85.931 2	89.536 1	90.712 2	85.178 3	86.135 5	91.071 4	80.478 6	72.529 5	68.886 8	73.350 3
	④	90.091 7	86.705 5	87.846 9	90.625 0	89.943 5	88.077 5	88.014 6	91.964 3	80.781 1	72.759 5	68.714 0	73.437 5
	⑤	93.254 9	85.984 6	88.586 4	91.585 8	92.245 0	87.911 3	88.645 2	92.633 9	80.277 9	72.754 5	70.127 1	72.916 7
	⑥	89.420 4	86.986 6	87.809 9	90.614 9	86.082 0	82.642 2	83.467 0	90.609 1	78.457 6	73.939 2	69.851 0	74.365 5
	⑦	89.718 1	87.257 2	87.913 8	90.399 1	87.089 9	86.448 2	86.070 6	90.101 5	79.319 8	74.795 3	69.164 0	74.619 3
MobileNetV3-small	①	**87.638 2**	**86.331 0**	**86.850 0**	**91.370 0**	**92.881 1**	**89.223 0**	**90.160 1**	**93.750 0**	**82.237 8**	**74.283 1**	**70.473 1**	**74.873 1**
	②	84.055 4	84.478 1	83.985 6	86.272 3	89.889 6	86.687 6	87.050 9	91.741 1	74.846 7	72.969 5	69.733 4	72.656 2
	③	87.063 1	83.259 6	84.485 6	88.133 8	86.006 3	82.312 7	82.992 0	89.062 5	81.283 5	72.148 1	69.095 4	72.842 6
	④	88.754 7	84.700 9	86.368 3	88.058 0	88.805 2	85.730 9	85.622 5	91.071 4	79.011 2	73.399 1	68.046 6	73.437 5
	⑤	79.465 8	77.930 2	78.402 5	84.897 5	83.951 8	80.642 1	81.253 4	88.392 9	7 823.31	6 910.19	6 514.83	7 040.82
	⑥	85.660 0	83.383 7	83.910 4	87.378 6	88.962 5	86.742 6	87.081 1	90.355 3	81.201 6	72.395 6	70.010 1	73.604 1
	⑦	84.850 2	81.399 4	82.714 7	86.623 5	89.037 6	84.583 8	85.453 1	88.324 9	74.053 7	74.068 5	69.401 9	74.365 5
MobileNetV3-large	①	**89.466 4**	**88.097 4**	**88.529 0**	**91.477 9**	**91.607 8**	**85.190 9**	**86.577 3**	**92.410 7**	**83.228 1**	**75.579 2**	**72.391 6**	**75.888 3**
	②	85.934 3	82.790 2	83.843 2	86.830 4	87.979 6	85.845 9	86.116 4	89.955 4	75.476 2	74.638 7	69.377 7	73.437 5
	③	83.590 2	81.269 3	81.786 5	86.084 1	85.468 8	83.580 3	82.922 3	88.839 3	76.913 0	69.861 4	66.425 5	71.319 8
	④	86.427 6	83.909 2	84.524 6	88.504 5	90.481 6	85.470 5	86.769 3	91.517 9	81.207 2	73.079 2	71.170 2	73.958 3
	⑤	84.502 2	82.477 4	82.974 9	86.299 9	86.136 8	83.588 3	83.427 3	89.955 4	72.517 9	73.420 1	69.402 2	72.395 8
	⑥	88.689 7	86.333 0	87.108 2	89.536 1	88.418 8	85.927 1	86.241 8	92.132 0	82.363 2	71.863 0	69.136 2	73.096 4
	⑦	87.958 5	85.957 7	86.728 4	89.751 9	88.078 6	84.364 7	85.089 5	91.878 2	78.506 3	72.199 5	68.251 7	72.842 6

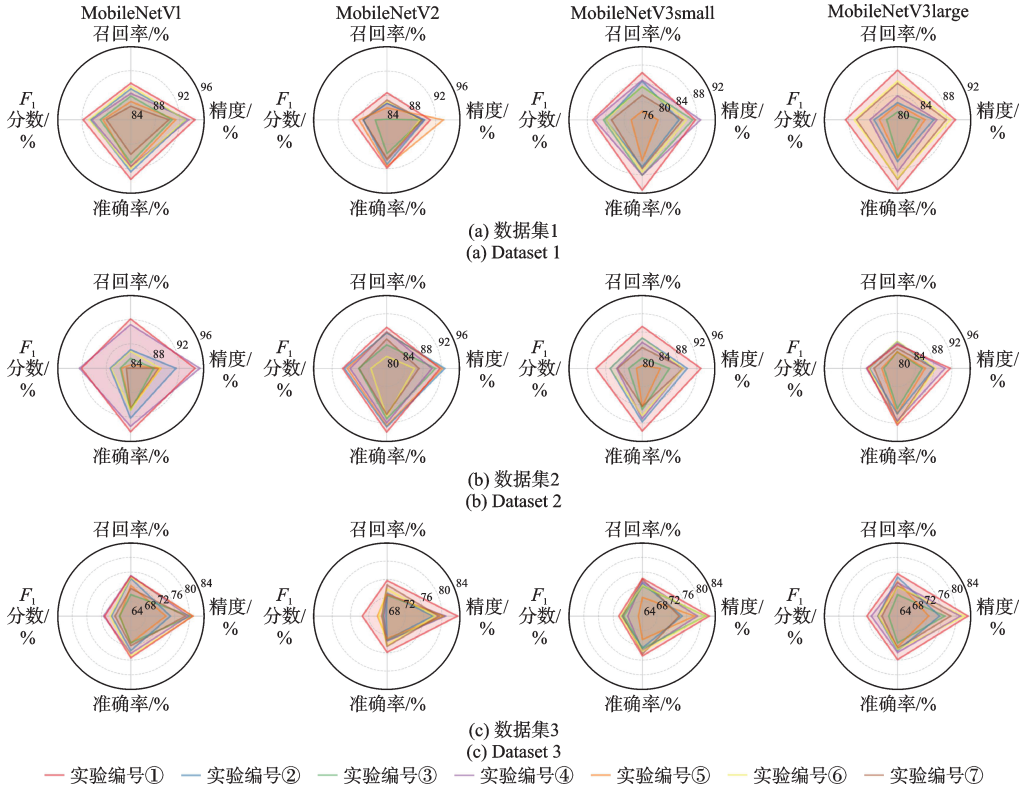

图5 对比实验性能雷达图

Fig.5 Radar charts of comparison experiment performance

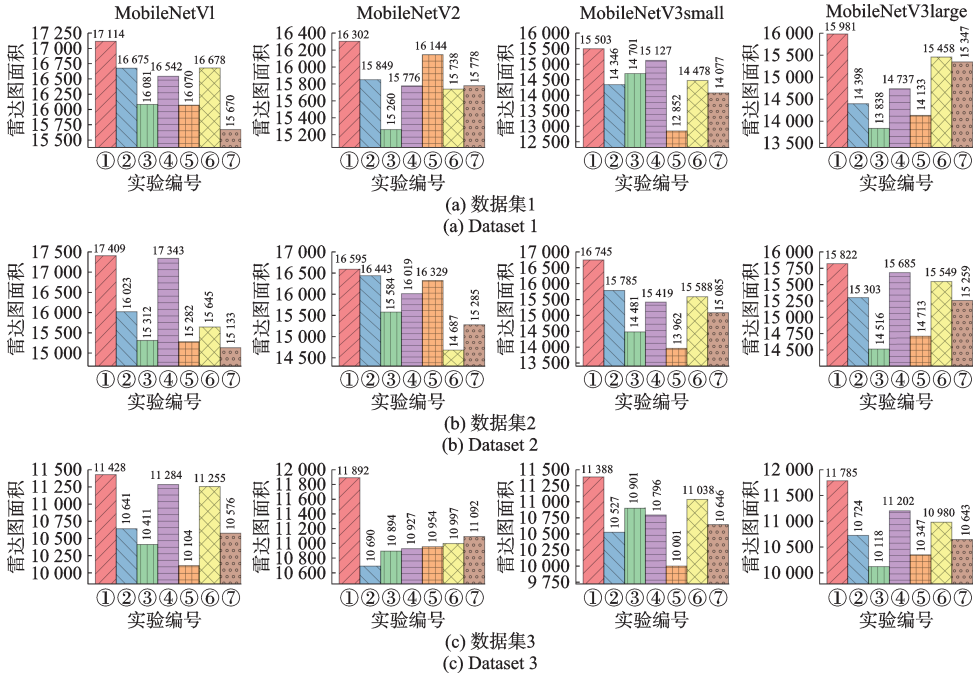

图6 对比实验性能雷达图的面积

Fig.6 Radar chart area of comparative experimental performance

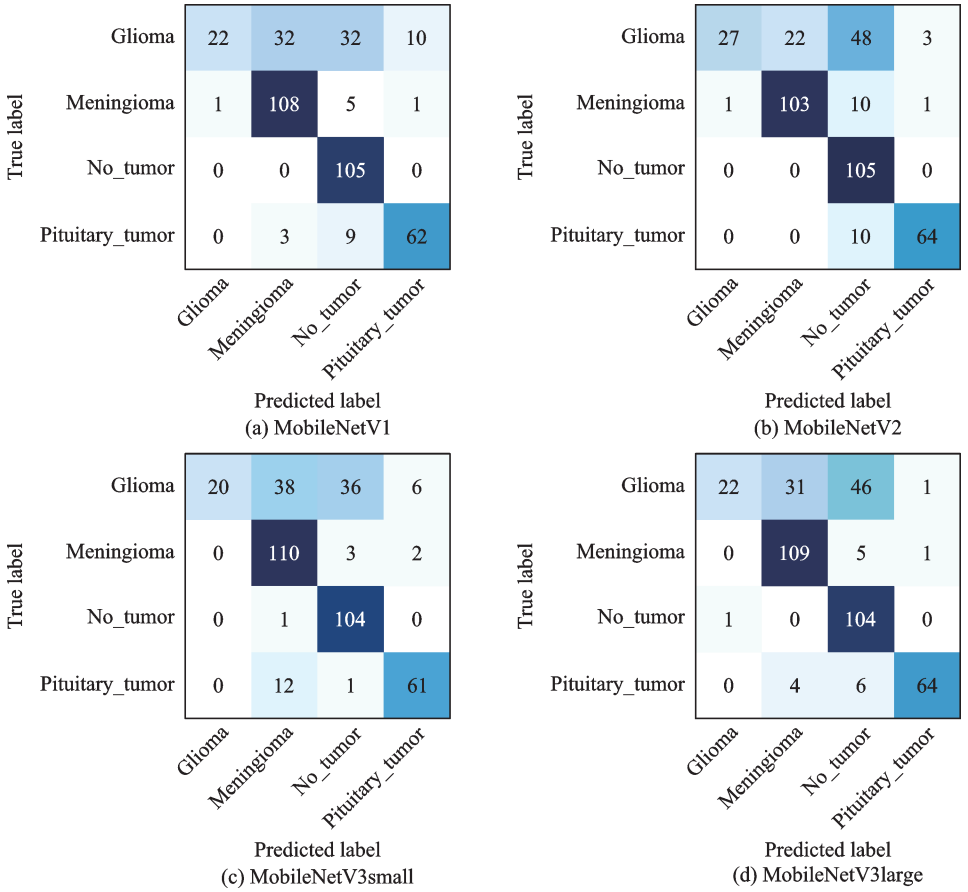

图 7 所提方法在数据集 3 上的混淆矩阵

Fig.7 Confusion matrix of the proposed method on Dataset 3

区域与特征;最后,可考虑引入动态权重调整机制与标签校准机制相结合的方式进行优化,根据不同类别样本数量的比例动态调整其在损失函数中的权重,对小样本类别赋予更高的权重,提高在类别不平衡的数据集分类性能。

5 结束语

本文提出基于标签自动校准的变温自蒸馏脑肿瘤MRI图像分类方法,以提高轻量化模型在脑肿瘤数据集上的识别性能。轻量化模型计算速度高效,但识别性能通常较低。本文提出基于标签自动校准的损失函数,防止自教师模型向学生模型传授错误信息,避免学生模型学习到错误知识。同时引入温度重启的变温蒸馏机制,若验证准确率在多轮训练中未提升,根据当前训练情况重启温度,防止轻量化网络学习难度过大。实验结果显示,该自蒸馏方法在多个脑肿瘤数据集上有效提高了轻量化模型的脑肿瘤识别性能,且优于其他典型自蒸馏方法。此方法无需改变模型结构、易于实现,可用于提升轻量化网络性能。

参考文献:

[1] SIEGEL R L, MILLER K D, JEMAL A. Cancer statistics, 2016[J]. CA: A Cancer Journal for Clinicians, 2016, 66(1): 7-30.

[2] BIRATU E S, SCHWENKER F, AYANO Y M, et al. A survey of brain tumor segmentation and classification algorithms[J].

Journal of Imaging, 2021, 7(9): 179.

[3] NAZIR M, SHAKIL S, KHURSHID K. Role of deep learning in brain tumor detection and classification (2015 to 2020): A review[J]. Computerized Medical Imaging and Graphics, 2021, 91: 101940.

[4] SAJID S, HUSSAIN S, SARWAR A. Brain tumor detection and segmentation in MR images using deep learning[J]. Arabian Journal for Science and Engineering, 2019, 44(11): 9249-9261.

[5] CHAN H P, SAMALA R K, HADJIISKI L M, et al. Deep learning in medical image analysis[J]. Advances in Experimental Medicine and Biology, 2020, 1213: 3-21.

[6] 王琳, 宋权润, 耿世超, 等. 神经网络滤波器剪枝技术研究综述[J]. 计算机工程与应用, 2026, 62(2): 1-25.
WANG Lin, SONG Quanrun, GENG Shichao, et al. Research review on neural network filter pruning techniques[J]. Computer Engineering and Applications, 2026, 62(2): 1-25.

[7] LIEBENWEIN L, BAYKAL C, CARTER B, et al. Lost in pruning: The effects of pruning neural networks beyond test accuracy[EB/OL]. (2021-03-04). https://doi.org/10.48550/arXiv.2103.03014.

[8] GOU J, YU B, MAYBANK J S, et al. Knowledge distillation: A survey[J]. International Journal of Computer Vision, 2021, 129(6): 1-31.

[9] ZHANG X, ZHU J, WANG D, et al. A gradual self distillation network with adaptive channel attention for facial expression recognition[J]. Applied Soft Computing, 2024, 161: 111762.

[10] SHARMA S, KUMAR A, MONPARA J, et al. StAlK: Structural alignment based self knowledge distillation for medical image classification[J]. Knowledge-Based Systems, 2024, 304: 112503.

[11] 张睿, 陈瑶, 王家宝, 等. 细粒度图像分类的自知识蒸馏学习[J]. 中国图象图形学报, 2024, 29(12): 3756-3769.
ZHANG Rui, CHEN Yao, WANG Jiabao, et al. Self-knowledge distillation learning for fine-grained image classification[J]. Journal of Image and Graphics, 2024, 29(12): 3756-3769.

[12] 唐媛, 陈莹. 基于动态混合注意力的自知识蒸馏[J]. 控制与决策, 2024, 39(12): 4099-4108.
TANG Yuan, CHEN Ying. Self-knowledge distillation based on dynamic hybrid attention[J]. Control and Decision, 2024, 39(12): 4099-4108.

[13] ZHANG L, SONG J, GAO A, et al. Be your own teacher: Improve the performance of convolutional neural networks via self distillation[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). [S.l.]: IEEE, 2019.

[14] ZHAO H, TIAN S, WANG J, et al. Boosting accuracy of student models via masked adaptive self-distillation[J]. Neurocomputing, 2025, 637: 129988.

[15] ZHAO G, SUN Z, LIU Y. Image recognition based on self-distillation and spatial attention mechanism[J]. International Journal of Modern Physics C: Computational Physics & Physical Computation, 2025, 36(10): 2542001.

[16] 倪水平, 马新良. 一种自蒸馏的轻量化图像分类网络方案[J]. 北京邮电大学学报, 2023, 46(6): 66-71.
NI Shuiping, MA Xinliang. A lightweight image classification network scheme based on self-distillation[J]. Journal of Beijing University of Posts and Telecommunications, 2023, 46(6): 66-71.

[17] LUAN Y, ZHAO H, YANG Z, et al. MSD: Multi-self-distillation learning via multi-classifiers within deep neural networks [EB/OL]. (2019-12-02). https://doi.org/10.48550/arXiv.1911.09418.

[18] 李圣巍, 李海, 马克, 等. 基于自蒸馏的图像分类框架[J]. 信息技术, 2025(4): 16-21.
LI Shengwei, LI Hai, MA Ke, et al. A self-distillation-based image classification framework[J]. Information Technology, 2025(4): 16-21.

[19] 苏楠, 宋政道, 曹晴, 等. 基于改进自蒸馏EMA-DeiT与迁移学习的番茄叶片病害识别方法[J]. 中国农机化学报, 2025, 46(10): 192-202, 209.
SU Nan, SONG Zhengdao, CAO Qing, et al. A tomato leaf disease recognition method based on improved self-distilled EMA-DeiT and transfer learning[J]. Journal of Chinese Agricultural Mechanization, 2025, 46(10): 192-202, 209.

[20] KIM K, JI B M, YOON D, et al. Self-knowledge distillation with progressive refinement of targets[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, 2021.

[21] SHEN Y, XU L, YANG Y, et al. Self-distillation from the last mini-batch for consistency regularization[C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022.

[22] WEI Y, BAI Y. Dynamic temperature knowledge distillation[EB/OL]. (2024-04-19). https://doi.org/10.48550/

arXiv.2404.12711.

[23] JAFARI A, REZAGHOLIZADEH M, SHARMA P, et al. Annealing knowledge distillation[C]//Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics. [S.l.]: Association for Computational Linguistics, 2021: 2493-2504.

[24] 汪瑞,吴向阳. 一种基于信息熵动态调节温度知识蒸馏的图像分类方法: CN202411435827.5[P]. 2025-09-12.
WANG Rui, WU Xiangyang. An image classification method based on information entropy for dynamic temperature knowledge distillation: CN202411435827.5[P]. 2025-09-12.

[25] LI Z, LI X, YANG L, et al. Curriculum temperature for knowledge distillation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(2): 1504-1512.

[26] LI Z, LI X, YANG L, et al. Student helping teacher: Teacher evolution via self-knowledge distillation[J]. arXiv preprint arXiv: 2110.00329, 2021.

[27] HINTON E G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[EB/OL]. (2015-03-09). https://doi.org/10.48550/arXiv.1503.02531.

[28] 司兆峰, 齐洪钢. 知识蒸馏方法研究与应用综述[J]. 中国图象图形学报, 2023, 28(9): 2817-2832.
SI Zhaofeng, QI Honggang. A review on research and applications of knowledge distillation methods[J]. Journal of Image and Graphics, 2023, 28(9): 2817-2832.

[29] 陈幸喆, 邱桃荣, 胡效康, 等. 基于自注意力和自蒸馏的甲状腺结节恶性C-TIRADS危险分层检测方法[J/OL]. 计算机辅助设计与图形学学报,1-15[2025-06-27]. http://kns.cnki.net/kcms/detail/11.2925.TP.20250206.1613.023.html.
CHEN Xingzhe, QIU Taorong, HU Xiaokang, et al. A thyroid nodule malignancy C-TIRADS risk stratification detection method based on self-attention and self-distillation[J/OL]. Journal of Computer-Aided Design & Graphics, 1-15[2025-06-27]. http://kns.cnki.net/kcms/detail/11.2925.TP.20250206.1613.023.html.

[30] HUANG W, YE M, SHI Z, et al. Self-knowledge distillation with dimensional history knowledge[J/OL]. Science China (Information Sciences),1-15[2025-06-29]. http://kns.cnki.net/kcms/detail/11.5847.TP.20250331.0909.002.html.

[31] 张恒, 张赛, 孙佳伟, 等. 深度学习脑肿瘤MRI图像分类研究进展[J]. 磁共振成像, 2023, 14(1): 166-171, 193.
ZHANG Heng, ZHANG Sai, SUN Jiawei, et al. Research progress in deep learning-based MRI image classification for brain tumors[J]. Magnetic Resonance Imaging, 2023, 14(1): 166-171, 193.

[32] 徐新智, 何宏. 基于局部通道编码的轻量化人体姿态估计算法[J]. 数据采集与处理, 2025, 40(6): 1625-1636.
XU Xinzhi, HE Hong. A lightweight human pose estimation algorithm based on local channel encoding[J]. Journal of Data Acquisition and Processing, 2025, 40(6): 1625-1636.

[33] 薛保, 周俊杰, 邵伟. 基于可变形注意力和多尺度多实例学习的全切片病理图像分类方法[J]. 数据采集与处理, 2026, 41(1): 231-243.
XUE Bao, ZHOU Junjie, SHAO Wei. A whole-slice pathological image classification method based on deformable attention and multi-scale multi-instance learning[J]. Journal of Data Acquisition and Processing, 2026, 41(1): 231-243.

作者简介:

徐楚迪(2001-),男,硕士研究生,研究方向:医学人工智能、医学图像处理,E-mail:xjxcd2001@163.com。

何宏(1973-),**通信作者**,女,教授,研究方向:医学人工智能、人机交互系统、医疗大数据分析,E-mail:hehong@usst.edu.cn。

陈家毓(1997-),男,博士研究生,研究方向:医学人工智能、嵌入式AI。

陈宇聪(2001-),男,硕士研究生,研究方向:机器视觉、目标检测、嵌入式AI。

李泽旭(2000-),男,硕士研究生,研究方向:脑电动作想象、元宇宙、虚实交互。

(编辑:王静)