中图法分类号:TP391.41;TP18　　**文献标识码**:A　　**文章编号**:1006-8961(2026)07-2631-14

论文引用格式:Yang Y Q, Chen Y and Lyu C. 2026. FixMatch++: a method for expanding limited image label data on the basis of semi-supervised learning. Journal of Image and Graphics, 31(7):2631-2644(杨雨晴, 陈谊, 吕程. 2026. FixMatch++:基于半监督学习的有限图像标签数据扩展方法. 中国图象图形学报, 31(7):2631-2644)[DOI:10.11834/jig.250398]

FixMatch++:基于半监督学习的有限图像标签数据扩展方法

杨雨晴,陈谊*,吕程

1. 北京工商大学计算机与人工智能学院食品安全大数据技术北京市重点实验室,北京　100048;
2. 北京工商大学前沿交叉研究院应用统计与数字监管北京市重点实验室,北京　100048

摘　要:**目的**　深度视觉模型对大规模标注数据的高度依赖限制了其实际应用。半监督学习为解决这一问题提供了有效途径,其中的FixMatch框架作为一个具有影响力的基线方法,通过弱—强数据增强策略显著提升了无标签数据的利用效率。然而,该框架在训练初期仍存在特征提取不稳定、图像采集/成像差异致使内容—风格信息易被混淆以及单增强视图伪标签噪声累积等问题,严重制约了模型在实际场景中的性能表现。**方法**　提出了FixMatch++框架,对原始FixMatch进行3个方面的改进:1)设计可学习批量归一化模块修正增强策略带来的通道级统计量偏差,结合双尺度并行卷积模块增强多尺度特征提取能力;2)引入内容/风格分离的双分支表征模块进行特征分离,并采用动态残差门控机制优化特征融合过程;3)提出多级别伪标签融合机制,缓解单增强视图导致伪标签噪声累积,将3种增强视图的分类概率加权,并配合类别阈值筛选策略生成可信标签,有效提升了标签的数量和质量。**结果**　在CIFAR-10(Canadian institute for advanced research 10-class dataset)、CIFAR-100(Canadian institute for advanced research 100-class dataset)和SVHN(street view house numbers)3个公开图像数据集上与7个近期主流半监督学习方法MeanTeacher、UDA(unsupervised domain adaptation)、ReMixMatch、FixMatch、FlexMatch、SoftMatch和UES(uncertainty-aware ensemble structure)进行比较,结果表明:FixMatch++方法在3个数据集上表现出较低的分类错误率,整体优于7种基线方法。在低标签量下表现尤其出色,例如,在CIFAR-10数据集250标签下,FixMatch++的分类错误率为4.56%,比7种基线方法降低了0.35%～27.76%。消融实验与可视化分析进一步验证了各模块的合理性和必要性。**结论**　提出的FixMatch++方法在有限标签数据场景下,能够有效扩展可用训练样本规模并提升图像分类性能。

关键词:图像分类;半监督学习(SSL);弱标签数据;数据增强;伪标签融合;跨域适应

FixMatch++: a method for expanding limited image label data on the basis of semi-supervised learning

Yang Yuqing, Chen Yi*, Lyu Cheng

1. Beijing Key Laboratory of Big Data Technology for Food Safety, School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing　100048, China; 2. Beijing Key Laboratory of Applied Statistics and Digital Regulation, Academy for Interdisciplinary Studies, Beijing Technology and Business University, Beijing　100048, China

收稿日期:2025-08-26;**修回日期**:2025-10-01;**预印本日期**:2025-10-08

*通信作者:陈谊　chenyi@th.btbu.edu.cn

基金项目:国家自然科学基金项目(U23B2009)

Supported by:National Natural Science Foundation of China(U23B2009)

Abstract: Objective Deep visual recognition models have achieved remarkable success in diverse domains ranging from everyday object classification to specialized tasks, such as medical image analysis. However, their performance remains heavily dependent on large-scale labeled datasets, the construction of which is costly, time consuming, and often impractical in real-world scenarios. Semi-supervised learning (SSL), which leverages abundant unlabeled data to compensate for limited labeled samples, has therefore emerged as a promising paradigm. Among existing SSL methods, FixMatch is a widely adopted baseline because of its simple yet powerful combination of consistency regularization and pseudo-labeling, driven by weak-strong augmentation. However, FixMatch still faces notable limitations: feature extraction instability during early training, entanglement of content and style caused by acquisition or imaging variations, and noise accumulation from single-view pseudo-labels. These issues reduce the robustness and generalizability of FixMatch in realistic deployment scenarios. **Method** To address these challenges, we propose FixMatch++, an enhanced framework that introduces three tightly integrated improvements. The first improvement is learnable-shift batch normalization (LS-BN) and dual-scale parallel convolution (DSPC). Conventional batch normalization assumes stable channel statistics, but the aggressive augmentations employed in FixMatch frequently distort these distributions, introducing bias. We design LS-BN that adaptively corrects augmentation-induced deviations, and DSPC incorporates parallel convolutional branches with different receptive fields, thus enabling rich multiscale representations and improving robustness to variations in object size and texture. The second improvement is content-style dual-branch representation with dynamic residual gating. Natural images often embed intertwined structural (content) and appearance (style) factors. A single-stream backbone entangles these signals, reducing discriminative power. We disentangle features into two dedicated branches: one focusing on content (shape and semantics) and the other on style (texture, illumination, and acquisition artifacts). A dynamic residual gating mechanism adaptively fuses the two streams and selectively emphasizes informative features under varying conditions, thereby stabilizing representation learning and enhancing generalizability. The third improvement is multilevel pseudo-label fusion. Unlike FixMatch that generates pseudo-labels from a single augmented view, FixMatch++ aggregates predictions from weak, medium, and strong augmentations. Their class probabilities are fused through weighted averaging, ensuring that noisy predictions from any single view are mitigated by corroborating evidence from others. In addition, we introduce a class-wise confidence thresholding strategy, which applies tailored thresholds for different categories to prevent dominant classes from overwhelming minority ones and to ensure balanced pseudo-label adoption. Together, these mechanisms substantially increase the quantity and quality of pseudo-labels, which are critical for effective semi-supervised training. **Result** We conduct extensive experiments on three widely used SSL benchmarks: CIFAR-10, CIFAR-100, and SVHN. We benchmark FixMatch++ against seven state-of-the-art baselines, namely, MeanTeacher, UDA, ReMixMatch, FixMatch, FlexMatch, SoftMatch, and UES. Across all datasets, FixMatch++ consistently achieves the lowest classification error rates among the approaches. Gains are particularly pronounced in low-label regimes, which best reflect real-world annotation scarcity. For instance, under the challenging CIFAR-10 setting with only 250 labeled samples, FixMatch++ reduces the error rate to 4.56%, representing an absolute improvement of 0.35%—27.76% compared with the seven baselines. On CIFAR-100, which involves 100 categories and is highly susceptible to interclass confusion, FixMatch++ demonstrates clear adaptability, outperforming FlexMatch and ReMixMatch. On SVHN, which features digit recognition in complex street-view conditions with substantial imaging variability, our framework surpasses UDA and SoftMatch, validating its robustness to acquisition artifacts. We conduct ablation studies to further validate the contribution of each proposed component. Results show that removing learnable-shift batch normalization considerably increases the error rates, confirming the necessity of correcting augmentation-induced biases. Excluding the content-style disentanglement module leads to unstable training and increased sensitivity to domain shifts, revealing the importance of separating and adaptively fusing features. Eliminating multilevel pseudo-label fusion degrades label quality and amplifies noise, particularly at the late training stages, thereby verifying the effectiveness of the fusion strategy. We also perform visualization analyses, including feature map inspection, class activation mapping, and pseudo-label confidence distribution plots. These visualizations provide intuitive evidence of how FixMatch++ stabilizes feature extraction, disentangles factors of variation, and generates reliable pseudo-labels. Together, the analyses confirm not only the empirical improvements but also the interpretability of the model's internal processes. **Conclusion** In summary, FixMatch++ substantially advances semi-supervised image classification by introducing

improvements at the architectural and algorithmic levels. The learnable-shift batch normalization and dual-scale convolution modules enhance the robustness of feature extraction. The content-style dual-branch representation with dynamic residual gating disentangles and adaptively fuses complementary features, and the multilevel pseudo-label fusion mechanism ensures high-quality supervision from unlabeled data. Comprehensive experiments and analyses demonstrate that these improvements yield consistent and notable performance gains over strong SSL baselines across multiple datasets and labeling regimes. Owing to the modular design of FixMatch++, the proposed components can be seamlessly integrated into other SSL frameworks, extending the applicability and influence of FixMatch++. FixMatch++ enlarges the effective training set under limited-label conditions and strengthens the reliability and interpretability of SSL. It offers a practical solution for real-world scenarios where annotation resources are scarce yet robust performance is essential.

Key words: image classification; semi-supervised learning(SSL); weakly labeled data; data augmentation; pseudo-label fusion; cross-domain adaptation

0 引　言

近年来，深度学习技术在图像分类等计算机视觉任务中取得了突破性进展。然而，现有监督学习方法通常依赖于大规模标签数据集，而高质量标签数据的获取往往面临标签成本高昂、效率低下等现实挑战。与此同时，实际应用场景中存在着海量无标签的数据资源。如何有效利用这些无标签数据提升模型性能，已成为当前机器学习领域亟待解决的关键科学问题。

半监督学习(semi-supervised learning，SSL)通过少量有标签数据引导模型从大量无标签数据中学习潜在信息，是提升数据利用效率的重要手段。近年来，基于一致性正则化与伪标签机制的SSL方法发展迅速。FixMatch作为代表性框架，通过弱增强生成伪标签、强增强执行一致性训练，兼顾训练流程的简洁性与模型性能的优越性，在多个数据集上取得了优异效果。然而，其性能仍受若干问题制约。为了更清晰地理解这些问题，下面从3个与之密切相关的研究方向(多尺度特征提取、特征解耦和伪标签生成)回顾已有工作及其局限。

1)多尺度特征提取。在图像识别任务中，不同尺度信息对模型特征表达能力具有重要影响。传统卷积神经网络多采用固定尺度卷积核，难以充分捕获多尺度目标特征。为此，特征金字塔网络(feature pyramid network，FPN)(Lin等，2017)通过自底向上特征金字塔结构融合多层特征，增强模型的尺度适应性。PyramidNet(Han等，2017)则通过逐层递增通道宽度，丰富多尺度特征表达，但存在计算复杂度较高的问题。此外，后续研究也提出了多种增强多尺度特征的结构，例如HRNet(high-resolution network)(Wang等，2021)通过并行保持高分辨率特征以提升跨尺度建模能力；近期的Transformer架构如PVTv2(pyramid vision Transformer v2)(Wang等，2022)则进一步表明多尺度特征在视觉识别中的重要性。国内学者也关注在复杂条件下的多尺度建模问题，例如提出动态自适应超分辨率框架以提升跨尺度鲁棒性(刘烨 等，2025)，进一步印证了多尺度特征对模型泛化能力的重要性。尽管这些方法在多尺度建模方面效果显著，但往往带来较大的计算开销，难以直接嵌入轻量化的半监督框架中。

2)特征解耦技术。图像往往包含内容与风格等不同类型信息，在跨域或风格变化显著的场景中，特征解耦技术能够将可迁移的内容特征与易变的风格特征分离，是提升泛化性的有效手段。梯度反转层(gradient reversal layer，GRL)通过在反向传播过程中对梯度进行符号翻转实现对抗学习，使提取的特征在不同域之间难以区分(Ganin 和 Lempitsky，2015)。近年的解耦表征综述(Wang等，2024)表明，内容—风格信息分离有助于抵抗成像风格/设备差异。近期提出的DEADiff(disentangled diffusion model)模型(Qi等，2024)结合解耦特征与扩散机制进一步提升了跨域泛化能力。另一类方法如Self-Challenging(Huang等，2020)通过抑制模型对风格特征的依赖，增强模型在跨域环境下的鲁棒性，进一步印证了显式解耦策略的有效性。然而，这些方法大多聚焦于域自适应或生成增强任务，与SSL的一致性正则化和伪标签生成流程结合不足，因而难以在半监督场景下充分发挥作用。

3)一致性正则化与伪标签生成。一致性正则化与伪标签机制是SSL的两条主线。Π-Model通过对

同一无标签图像施加不同扰动，使两次预测保持一致（Laine 和 Aila，2017）。Temporal Ensembling 对图像预测做指数滑动平均作为目标（Laine 和 Aila，2017）。Mean Teacher 对模型权重做指数滑动平均得到教师模型，从而稳定训练目标（Tarvainen 和 Valpola，2017）。VAT（virtual adversarial training）在对抗方向寻找最敏感扰动以加强一致性（Miyato 等，2019）。MixMatch 将标签平滑、混合增强与分布对齐结合到一个统一框架（Berthelot 等，2019）。UDA（unsupervised domain adaptation）强调“弱增强/强增强”的一致性（Xie 等，2020）。ReMixMatch 在此基础上引入分布对齐与锚点增强（Berthelot 等，2020）。FixMatch 以“弱增强产生高置信伪标签 + 强增强一致性”的极简范式取得了强劲效果（Sohn 等，2020）。随后，FlexMatch 提出类别自适应阈值以缓解类不平衡问题（Zhang 等，2021），SoftMatch 通过软伪标签与置信度加权降低硬阈值引入的噪声放大（Chen 等，2023）。近年来，也有学者探索新的伪标签生成机制。FreeMatch 结合预测分布熵提出自适应阈值策略以动态调整伪标签采纳条件（Wang 等，2023）；SimPLE（similarity-preserving label enhancement）利用伪标签间的相似性约束缓解噪声干扰（Hu 等，2021）；Meta Pseudo Labels 通过教师—学生协同训练优化伪标签生成（Pham 等，2021）。类似地，国内学者提出结合类激活图回放与最小熵采样的多标签类增量学习方法（周怡凡 等，2025），在减少伪标签错误采纳方面展现出成效。已有实证研究还表明，不同一致性正则化策略在稳定性与鲁棒性上的表现存在差异（Fan 等，2023），说明现有方法在训练早期仍容易受到不可靠伪标签和增强策略波动的影响。且上述方法普遍采用统一或相对粗粒度的阈值调控，缺乏对类间差异性与训练阶段动态性的细粒度建模，这往往导致伪标签质量不足或类间采纳失衡。

综合来看，虽然多尺度特征建模、特征解耦与伪标签生成方面已有大量工作，但 FixMatch 框架在实际应用中仍存在三大瓶颈：1）特征提取不稳定。强增强策略带来通道级统计量偏差，导致标准归一化的统计量发生漂移，容易在训练早期引发优化振荡与表征退化；同时，裁剪/缩放等增强策略使弱/中/强增强视图的可见尺度不同，限制了模型对多尺度结构的有效建模能力。2）内容—风格信息易混淆。复杂场景下由采集条件与成像风格差异引起的域偏移，使得内容信息与风格因素相互混淆，不仅削弱了模型的判别一致性，也限制了其跨域泛化能力。3）伪标签噪声积累。基于单视图（通常来自弱增强图像）生成的伪标签在训练早期阶段不够稳定，错误标签易干扰强增强分支中一致性学习的进行，从而影响模型的整体收敛质量，已有工作（如 DivideMix（Li 等，2020））在噪声标签学习场景中提出利用混合分布建模来区分干净与噪声样本，并结合半监督训练以缓解噪声干扰，其思想也为 SSL 中伪标签质量控制提供了启发。然而，如何在半监督框架下设计更高效的伪标签过滤与融合机制仍是亟待解决的问题。

针对上述问题，本文在 FixMatch 框架基础上提出面向有限标签场景的扩展方法 FixMatch++。主要贡献如下：1）提出可学习批量归一化模块（learnable-shift batchnorm，LS-BN），缓解强增强策略带来的通道级统计量偏差；同时引入并行双尺度卷积结构（dual-scale parallel convolution，DSPC）利用 3 × 3/5 × 5 的卷积核，增强多尺度表征能力，从而稳定训练初期优化过程并提升细粒度判别性能。2）设计内容/风格分离的双分支表征模块（content-style dual-branch representation，CS-DBR），将图像特征分解为内容与风格两个独立分支，结合梯度反转层（gradient reversal layer，GRL）抑制风格信息对分类决策的干扰；进一步引入动态残差门控（dynamic residual gating，DRG）机制，实现内容与风格特征的自适应融合，通过动态调节融合权重优化特征解耦与融合过程，提升跨域一致性与模型鲁棒性。3）提出多级别伪标签融合机制，对弱/中/强 3 种级别的增强视图的分类概率加权，得到各分类融合概率，根据类别阈值进行可信标签采纳，降低类不均衡带来的偏差，提升模型收敛质量与稳定性。

所提出的 3 项改进并非常见模块的简单堆叠，而是针对 FixMatch 在增强策略后的图像分布偏移、图像采集/成像导致的内容—风格混淆和单级别增强视图伪标签噪声 3 方面结构性问题提出了系统性修复机制。首先，LS-BN 稳定了不同级别增强视图的统计量分布，DSPC 缩小了不同增强级别下的尺度差异；其次，CS-DBR 结合 GRL 实现了内容特征与风格特征的分离，使其风格特征仅通过通道校准与门控机制参与融合；最后，再辅以 3 种级别增强视图分类概率加权融合策略，共同提升伪标签质量与收敛稳定性。在 CIFAR-10（Canadian institute for

advanced research 10-class dataset）、CIFAR-100（Canadian institute for advanced research 100-class dataset）与SVHN（street view house numbers）3个经典图像分类数据集上的实验表明，FixMatch++在不同标签数量条件下的分类准确率均优于7个基线方法，尤其在低标签率场景下表现突出，验证了方法的有效性。消融实验和可视分析进一步证明了各模块的合理性及其贡献。

1 本文方法

1.1 FixMatch++方法整体框架

FixMatch++方法基于改进的ResNet（residual network）网络结构，引入多级别伪标签融合机制，实现对有限图像标签数据的扩展。FixMatch++方法的整体框架由图像分类概率预测和图像分类标签生成两部分组成，其工作原理如图1所示。

首先，采用弱增强、中增强和强增强方法对输入图像数据集中的每幅图像进行多级别图像增强，生成多个增强视图。然后，将各级别增强视图分别输入到改进的ResNet神经网络中进行特征提取，得到每个增强视图的图像特征，再经softmax函数得到各级别图像增强视图的分类概率。进而提取出每个增强视图分类概率中的最大分类概率 P_{wmax}、P_{mmax} 和 P_{smax}，将其进行归一化后作为各级别增强视图权重值。然后，对各级别增强视图分类概率进行加权融合，得到每个图像的各分类融合概率 P_{avg}，并从中选取最大概率 P_{max} 以及 P_{max} 对应的类别作为该图像的伪标签。最后，通过与该类别阈值 τ_c 比较，判断伪标签是否有效。若 $P_{\mathrm{max}} \geqslant \tau_c$，则将其作为可信标签加入图像标签扩充数据集；否则为不可信伪标签，将其舍弃。

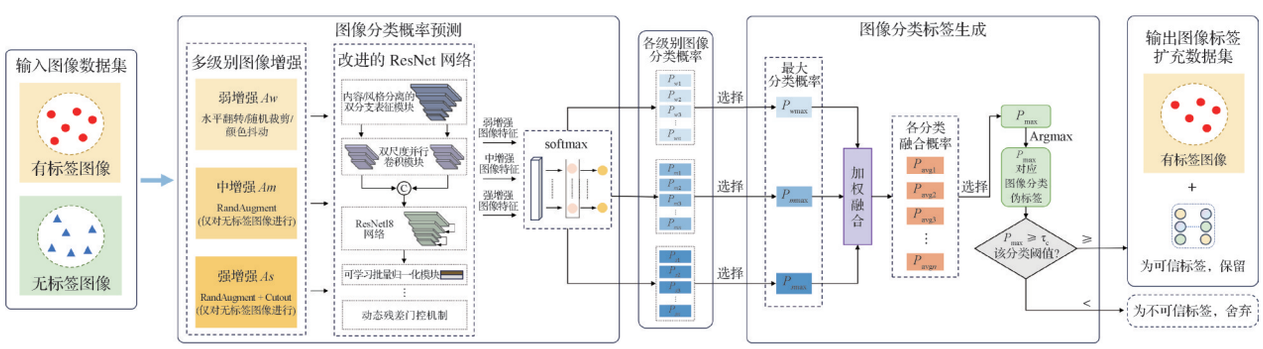


图1 FixMatch++整体框架

Fig. 1 The framework of the FixMatch++ method

整个框架的设计理念在于通过多级别伪标签融合机制提高了伪标签的质量，从而增强模型的性能；并在原有ResNet18基础上进行网络结构改进，引入了多个关键模块，有效增强了模型的学习建模能力和鲁棒性。

1.2 改进的ResNet网络

改进的ResNet网络是图像分类概率预测模块的核心部分，它引入了可学习批量归一化模块（LS-BN）与双尺度并行卷积模块（DSPC），缓解强增强导致的归一化统计漂移并强化了多尺度表征，最终共同增强了ResNet网络的多尺度特征提取能力。引入了内容/风格分离的双分支表征模块（CS-DBR）和动态残差门控机制（DRG）解耦内容与风格特征，并实现了动态调节内容与风格特征的融合权重，进一步提升跨域一致性与鲁棒性。

1.2.1 LS-BN和DSPC

1）可学习批量归一化模块（LS-BN）

该模块通过在传统BN（BatchNorm）的基础上引入自适应标准化参数 α 缓解训练过程中可能出现的归一化漂移问题。其核心目标是保证每个通道在训练过程中能稳定地进行标准化，从而提升训练稳定性和模型的性能，计算为

$$\hat{\boldsymbol{x}} = \left(\frac{\boldsymbol{x} - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \alpha\right) \cdot \gamma + \beta \tag{1}$$

式中，$\boldsymbol{x}$ 是输入特征图，μ 是输入的均值，σ 是标准差，ε 是为了避免除零错误的常数，γ 和 β 分别是可学习的缩放和偏移参数，α 为自适应标准化参数（初始化为0，使得初始行为等价于常规BN，训练中与网络参数端到端共同优化），$\hat{\boldsymbol{x}}$ 是归一化后的输出特征。

LS-BN在保持标准BN形式的同时，赋予小幅、可学习的通道尺度校正（初值为0，端到端学得），自适应标准化参数α对每个特征通道的标准化进行了优化，从而在弱/中/强3种增强视图之间更稳定地共享统计，减少早期不收敛与后续过度校准的风险，并有效避免了梯度消失或爆炸问题，提升了训练过程的稳定性。

2）双尺度并行卷积模块（DSPC）

双尺度并行卷积模块的作用是在同一层内提取多尺度信息并进行轻量融合，通过多尺度卷积核增强特征提取能力，捕捉不同尺寸的目标特征。在CIFAR-10/100、SVHN的32 × 32像素的小分辨率场景与ResNet骨干网络下，判别性线索主要集中于两段尺度：局部纹理/边缘与中尺度部件/轮廓。前者涵盖笔画、毛发以及物体边界等；后者涵盖数字圈/弧与物体局部形状块（如机翼/耳廓）的轮廓信息。模块以3 × 3聚焦细节纹理/边缘、以5 × 5覆盖中尺度部件/轮廓，在同一层同时使用这两段最有效的判别尺度；考虑到在骨干网络中采用多层3 × 3的叠加已能等效扩大感受野，额外再引入7 × 7等卷积层分支所带来的“全局”增益在小分辨率场景上十分有限。相对地，3 × 3/5 × 5的双尺度在成本—收益上更均衡，在实现中令两分支各产生通道并以1 × 1融合，保持输出维度与计算可控。

如图2所示，输入形状为$C \times H \times W$（特征图的三维形状：通道数 × 高度 × 宽度）的特征图，被并行送入两个卷积分支：一个使用3 × 3卷积，另一个使用5 × 5卷积，两分支均采用stride = 1、padding = $\lfloor k/2 \rfloor$（stride为步幅，k为卷积核的边长），以保持空间大小$H \times W$不变。每个分支后接ReLU（rectified linear unit）激活函数，得到两组不同感受野的特征。随后在通道维度进行拼接（concatenation，C），形成包含多尺度信息的组合特征。最后通过1 × 1卷积完成通道压缩与特征混合，将拼接后的通道数映射回目标维度，从而输出与输入空间尺寸一致、但更具多尺度判别性的特征图。该设计在不增加大量参数的前提下增强了网络对不同目标尺寸与纹理细节的敏感性，也为后续的内容/风格分离的双分支表征模块提供更加稳健的表征基础。

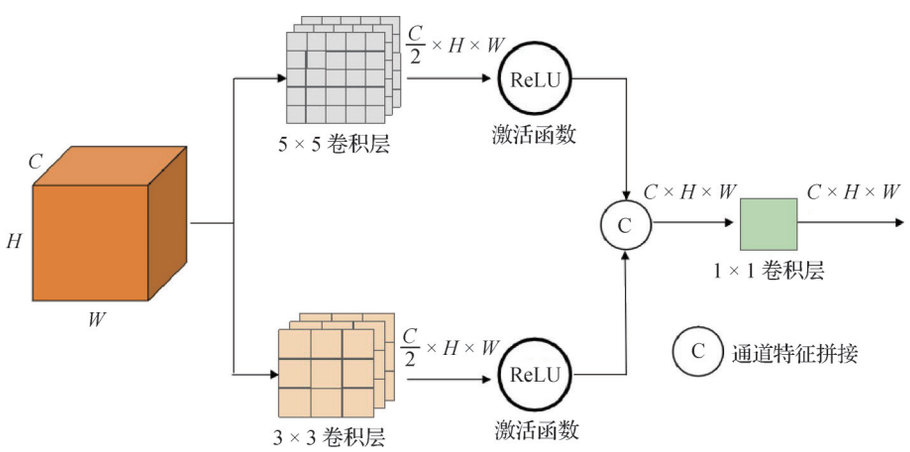


图2　双尺度并行卷积模块结构

Fig. 2　Dual-scale parallel convolution module structure

目前该方法在CIFAR-10/100数据集中32 × 32像素分辨率的图像上取得了良好的效果。对于其他更高分辨率的数据集，将进一步探索合适的卷积核组合和参数配置，以便在不增加过多计算开销的前提下，最大化特征提取能力。

1.2.2　CS-DBR

在基于半监督学习的有限图像标签数据扩展方法FixMatch++中，弱/中/强3种级别的增强及图像采集时引入的风格差异（亮度、对比度、色偏、纹理强弱）会导致图像特征出现通道尺度漂移，即各通道的整幅响应被近似按比例放大或压缩。通道尺度漂移会使softmax函数输出的最大分类概率下降或上升：当最大分类概率下降时，原本应被采纳的可信标签的最大分类概率因低于其类别阈值τ_c而被剔除；当最大分类概率上升时，错误标签的最大分类概率因超过类别阈值τ_c而被误采为可信标签，从而降低了标签的可信度并拖慢模型训练收敛。为此，提出内容/风格分离的双分支表征模块（CS-DBR），将特征

按“位置相关（内容特征）/位置无关（风格特征提取风格向量）”进行结构性分离，随后用风格向量$\boldsymbol{s}$以通道级运算的方式对内容特征进行校准，从而提升标签可信度和鲁棒性。同时，风格向量$\boldsymbol{s}$将作为动态残差门控（DRG）的控制信号，用于对风格残差进行可控引入。

内容特征图$\boldsymbol{F}_c \in \mathbf{R}^{C\times H\times W}$：保留空间坐标$(x,y)$，代表与类别相关的形状、边缘以及部件/布局等位置相关的结构信息。

风格向量$\boldsymbol{s} \in \mathbf{R}^{C}$：不含空间坐标$(x,y)$，无形状、边缘以及部件/布局等结构信息，仅表示每个通道的平均响应幅度，即风格信息总体的强度。

如图3所示，输入形状为$C\times H\times W$的特征被并行送入两条编码分支：上支为内容编码器，下支为风格编码器。其中，内容分支采用 stride = 1、padding = $\lfloor k/2 \rfloor$（stride 为步幅，k为卷积核边长），不做全局池化，直接输出$\boldsymbol{F}_c$，保留空间坐标(x,y)；风格分支在 stride = 1、padding = $\lfloor k/2 \rfloor$（stride 为步幅，k为卷积核边长）的基础上再对风格特征图$\boldsymbol{F}_s$做全局平均池化（global average pooling，GAP），去掉空间坐标(x,y)得到风格向量$\boldsymbol{s}$。

通过风格向量$\boldsymbol{s}$生成每个通道的缩放参数，并对内容特征图$\boldsymbol{F}_c$逐元素运算进行校准，具体为

$$(\gamma_c,\beta_c)=G(s)\in\mathbf{R}^{C},\quad \tilde{\boldsymbol{F}}=(1+\gamma_c)\odot\boldsymbol{F}_c+\beta_c \tag{2}$$

式中，$G(\cdot)$为两层多层感知机（multi-layer perceptron，MLP），γ_c和β_c分别为通道级缩放参数，初始$\gamma\approx 0$、$\beta\approx 0$为恒等映射，便于训练，$\tilde{\boldsymbol{F}}$为校准后的特征图，$\odot$为 Hadamard（逐元素）乘，通道维度上各自独立计算。

为进一步降低风格干扰，使风格向量$\boldsymbol{s}$只表示当前通道的风格信息总体的强度而不携带类别相关信息，进一步在有标签图像的风格向量$\boldsymbol{s}$上加入梯度反转层$GRL_\lambda(\cdot)$与一个轻量风格判别器$D(\cdot)$，具体为

$$L_{\text{adv}}=CE\left(D\left(GRL_\lambda(\boldsymbol{s})\right),y\right) \tag{3}$$

式中，y代表有标签图像的真实标签，CE表示计算交叉熵损失。最终，内容特征图与风格向量在后续的动态残差门控（DRG）机制中自适应融合两路信息，形成“内容主导、风格可控”的表示。

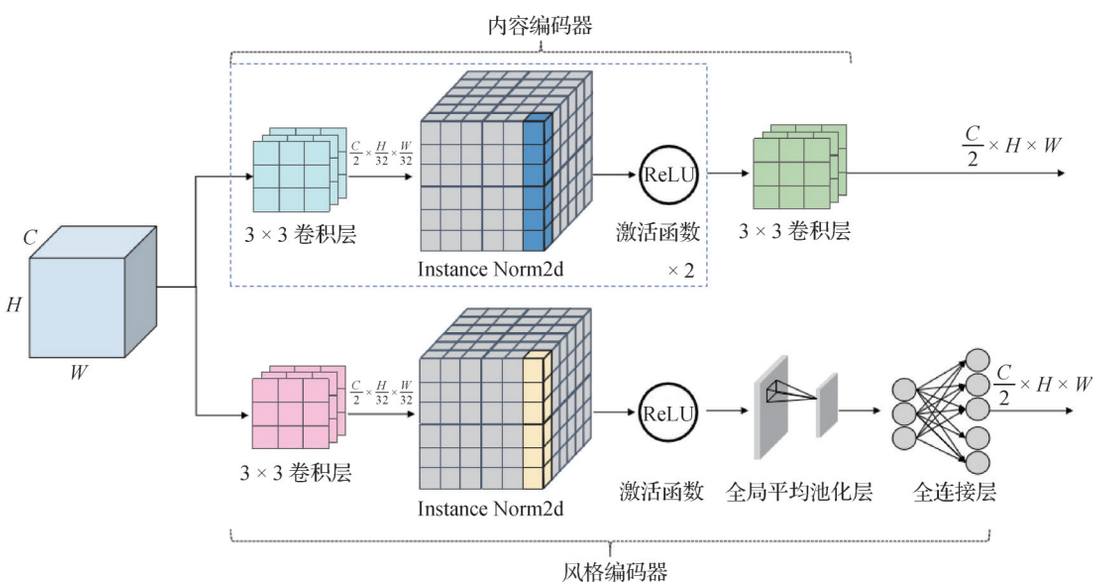


图3　内容/风格分离的双分支表征模块结构

Fig. 3　Content-style dual-branch representation module structure

1.2.3　用于特征融合的动态残差门控机制

动态残差门控机制通过引入自适应门控机制，动态调节内容和风格特征的融合比例，从而提高模型的鲁棒性。在该机制中，使用门控函数G控制内容和风格特征的结合方式。

将风格向量$\boldsymbol{s}$送入如图4所示的门控生成器（LeakyReLU→LayerNorm→LeakyReLU→Linear），得到通道级对数权重$w\in\mathbf{R}^{C}$。为使权重有界，采用 tanh 函数将其压缩为$(-1,1)$：$g=\tanh(w)$；当需要非负权重时使用$g=\tanh(w)+1/2$。

门控函数定义为

$$G\left(\boldsymbol{f}_s\right)=g\cdot\boldsymbol{s} \tag{4}$$

式中，门控函数的作用为对风格特征按通道缩放：$g \geqslant 0$时该通道的风格特征放大/保留，$g < 0$时起到反向抵消该通道的风格特征。

最终输出按残差方式与内容特征f_c进行融合，得到融合特征$\boldsymbol{f}$，具体为

$$\boldsymbol{f} = f_c + G(f_s) \tag{5}$$

整个过程不改变空间尺寸H、W，增强体现在通道语义的自适应调节。

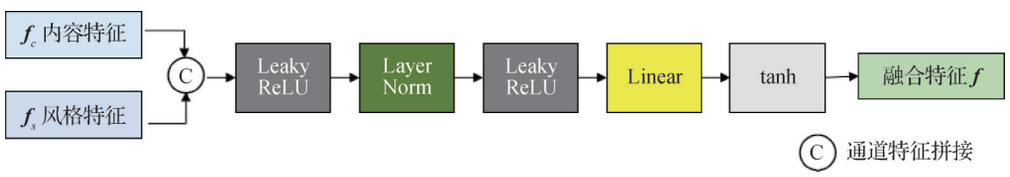


图4　动态残差门控机制结构

Fig. 4　Dynamic residual gating structure

1.3　多级别伪标签加权融合机制

多级别伪标签融合机制是图像分类标签生成的重要机制，其核心思想是通过结合不同级别视图的预测输出，利用加权融合生成高质量的伪标签，从而提升半监督学习的效果。

1.3.1　多级别图像增强

为提高模型在半监督学习中的表现，FixMatch++对输入图像应用了3种不同的增强策略，以产生多个增强视图。具体增强方式包括：

1）弱增强（weak augmentation，Aw）。轻度的数据增强操作，如随机裁剪和水平翻转等。

2）中增强（medium augmentation，Am）。中等强度的增强操作，如RandAugment。

3）强增强（strong augmentation，As）。使用更强的数据增强方法，如Cutout。

RandAugment是一种自适应数据增强方法，其通过在不同的增强操作中随机选择，动态调整增强强度。Cutout是一种通过遮挡图像区域来增强模型对局部信息的关注，防止模型对特定区域的过拟合。

3种增强方式对应的图像视图将分别输入到改进的ResNet网络中进行特征提取，得到每个视图的图像特征，如图1所示。

1.3.2　Logits输出与softmax函数

Logits是FixMatch++模型对图像输出的每个类别未经归一化的预测值，softmax函数将每个类别的n维logits（n为类别数）转化为概率，且所有类别的概率之和为1。

设一幅图像的logits向量为$\boldsymbol{z} = \left[z_1, \cdots, z_n\right]^{\mathrm{T}} \in \mathbf{R}^n$，则$softmax(z_i)$的第$i$类概率为

$$softmax(z_i) = \frac{\mathrm{e}^{z_i}}{\sum_{j=1}^{n}\mathrm{e}^{z_j}} \tag{6}$$

对同一幅图像进行多级别图像增强后，经改进的ResNet网络特征提取后得到弱、中、强3种级别增强视图下的$logits_{\mathrm{w}}$、$logits_{\mathrm{m}}$和$logits_{\mathrm{s}}$。对$logits_{\mathrm{w}}$、$logits_{\mathrm{m}}$和$logits_{\mathrm{s}}$进行softmax计算，可以得到各级别增强视图下的图像分类概率分布向量，具体为

$$\begin{cases} \boldsymbol{P}_{\mathrm{w}} = softmax\left(logits_{\mathrm{w}}\right) \\ \boldsymbol{P}_{\mathrm{m}} = softmax\left(logits_{\mathrm{m}}\right) \\ \boldsymbol{P}_{\mathrm{s}} = softmax\left(logits_{\mathrm{s}}\right) \end{cases} \tag{7}$$

式中，$\boldsymbol{P}_{\mathrm{w}}$、$\boldsymbol{P}_{\mathrm{m}}$、$\boldsymbol{P}_{\mathrm{s}}$分别是弱增强、中增强和强增强视图的softmax输出，分别为$n \times 1$的图像分类概率分布向量（n为类别数）。

1.3.3　最大分类概率的归一化

最大分类概率表示FixMatch++模型对于每个增强视图中最有信心的类别的预测概率值，P_{wmax}、P_{mmax}和P_{smax}分别表示一幅图像在弱增强视图、中增强视图和强增强视图的最大概率值。

为了将3种不同级别增强视图的分类概率进行统一处理并进行加权，将3种视图的最大分类概率进行归一化，得到每个增强视图的加权权重，使得它们的总和为1。具体为

$$\begin{cases} w_{\mathrm{w}} = \dfrac{P_{\mathrm{wmax}}}{total\_p} \\ w_{\mathrm{m}} = \dfrac{P_{\mathrm{mmax}}}{total\_p} \\ w_{\mathrm{s}} = \dfrac{P_{\mathrm{smax}}}{total\_p} \end{cases} \tag{8}$$

式中，w_{w}、w_{m}和w_{s}表示弱、中、强3个级别增强视图的归一化权重，$total_p$表示3种级别增强视图的最大

概率和。

1.3.4　加权融合

利用计算得到的加权值，对每个增强视图 softmax 输出的分类概率进行加权融合。具体计算式为

$$\boldsymbol{P}_{\text{avg}} = \boldsymbol{P}_{\text{w}} \times w_{\text{w}} + \boldsymbol{P}_{\text{m}} \times w_{\text{m}} + \boldsymbol{P}_{\text{s}} \times w_{\text{s}} \tag{9}$$

式中，$\boldsymbol{P}_{\text{avg}}$是加权融合后的最终概率分布，为$n \times 1$的向量（$n$为类别数），$w_{\text{w}}$、$w_{\text{m}}$、$w_{\text{s}}$分别是每个增强视图的加权系数。

1.3.5　标签生成与采纳

最终，伪标签通过从加权融合后的概率分布中选择最大概率所对应的类别生成。计算式为

$$\hat{y} = Argmax(P_{\max}) \tag{10}$$

式中，$\hat{y}$为生成的伪标签。$Argmax$用于选择从加权融合后的概率分布中，最大融合概率$P_{\max}$对应的类别。

伪标签的采纳则依赖于阈值机制，$\tau_{\text{c}} = classwise[\hat{y}]$表示当前图像的伪标签$\hat{y}$对应的类别的阈值。若$P_{\max} \geqslant \tau_{\text{c}}$，则采纳该伪标签为可信标签。

最终，输出图像标签扩充后的数据集以及与图像对应的可信标签，舍弃不可信伪标签。

2　实验与结果分析

2.1　数据集与实验设置

本文在 CIFAR-10、CIFAR-100 和 SVHN 3 个公开的半监督学习图像数据集上进行了实验。CIFAR-10 图像数据集中含有 10 个类别的物体，包括飞机、汽车和猫等，每个类别有 6 000 幅彩色图像；CIFAR-100 图像数据集含有 100 个类别的物体，包含更具体的类别，如：不同品种的花朵、蔬菜等，其中每个类别有 600 幅彩色图像。SVHN 是一个街景房屋号码图像数据集，包含了从 Google 街景图像中提取的 10 类数字图像，用于识别房屋号码。3 个数据集的详细信息如表 1 所示。

在经过基本数据增强的数据集上进行实验。不同的增强方式应用于有标签图像和无标签图像，具体的增强策略如图 1 所示。基本的数据增强方法包括水平翻转、随机裁剪以及颜色抖动等常见操作，而对于无标签图像，还使用了 RandAugment 和 Cutout 方法。实验中的主要评价指标为验证集上的分类错误率。

表 1　实验数据集

Table 1　Experimental datasets

数据集	图像尺寸/像素	类别数	训练集数量	测试集数量
CIFAR-10	32 × 32	10	50 000	10 000
CIFAR-100	32 × 32	100	50 000	10 000
SVHN	32 × 32	10	73 257	26 032

实验均在 NVIDIATeslaP100 显卡（16 GB 内存）上进行，操作系统为 Linux5.15.133，使用 PyTorch2.0.0 框架，CUDA 版本为 11.4，cuDNN 版本为 8.0.9，并使用 Python3.10.12 作为编程语言。训练使用了随机梯度下降法（stochastic gradient descent，SGD）优化器，并对其进行了学习率调度。训练轮数设置为 300 轮，批量大小为 64。初始学习率设置为 0.03，并采用指数衰减策略，分别在第 60、120、160 和 240 轮时将学习率减小为原来的 0.2 倍。超参数设置为 $\lambda = 1$、$\beta = 0.9$、$\tau = 0.95$、$\mu = 7$、$B = 64$、$K = 220$。关于 w_{w}、w_{m} 和 w_{s}：三者为式（8）定义的图像样本级自适应权重，在每次前向由当前图像样本的 3 种增强视图的最大类别概率归一化得到，非固定超参数，因此不在超参数表中单列。

所有实验方法和设置遵循相同的训练配置，以确保实验环境的一致性。

2.2　模型对比实验

为了全面评估 FixMatch++ 框架的性能，选择了 7 种当前主流的半监督学习模型作为对比模型，分别是 MeanTeacher、UDA、ReMixMatch、FixMatch、FlexMatch、SoftMatch 和 UES（uncertainty-aware ensemble structure）。这些对比模型涵盖了基于一致性学习、伪标签生成以及增强一致性等策略。通过与这些方法进行对比，能够全面评估 FixMatch++ 在不同半监督学习设置下的性能表现，进一步验证其在有限图像标签数据条件下的优势。

实验数据集标签量的选择与 7 个基线模型保持一致，以保证可比性与可复现性。表 2 列出了 FixMatch++ 与 7 种基线方法在 CIFAR-10（250 个标签、4 000 个标签）、CIFAR-100（400 个标签、2 500 个标签）和 SVHN（1 000 个标签）数据集的分类错误率对比。

根据表 2 结果显示，相比其他方法，FixMatch++ 方法在 3 个数据集上几乎全部优于 7 个基线方法产生最佳的分类性能，分类错误率有了一定的下降。

表 2　FixMatch++方法在3个数据集上与7种基线方法的分类错误率对比结果

Table 2　Comparison of classification error rates between the FixMatch++ method and seven baseline methods on three datasets

/%

方法	CIFAR-10（250个标签）	CIFAR-10（4 000个标签）	CIFAR-100（400个标签）	CIFAR-100（2 500个标签）	SVHN（1 000个标签）
MeanTeacher	32.32	9.19	67.18	53.91	3.42
UDA	8.82	4.88	59.28	33.13	2.46
ReMixMatch	5.44	4.72	44.28	27.43	2.65
FixMatch	5.07	4.26	48.85	28.29	<u>2.28</u>
FlexMatch	4.99	**3.95**	32.44	23.95	2.86
SoftMatch	<u>4.91</u>	4.82	**31.10**	26.66	2.33
UES	4.97	4.21	42.27	<u>22.18</u>	–
FixMatch++（本文）	**4.56**	<u>4.12</u>	<u>31.61</u>	**21.89**	**2.19**
FixMatch++分类错误率降低程度	0.35%～27.76%	0.09%～5.07%	0.83%～35.57%	0.29%～32.02%	0.09%～1.23%

注：加粗和下划线字体分别表示各列最优和次优结果。“–”表示未提供实验结果。

1）在CIFAR-10（250个标签）的情况下，FixMatch++错误率为4.56%，相较对比方法分类错误率下降0.35%～27.76%。

2）在CIFAR-10（4 000个标签）的情况下，FixMatch++分类错误率为4.12%，相较对比方法分类错误率下降0.09%～5.07%。

3）在CIFAR-100（400个标签）的情况下，FixMatch++分类错误率为31.61%，相较对比方法分类错误率下降0.83%～35.57%。

4）在CIFAR-100（2 500个标签）的情况下，FixMatch++分类错误率为21.89%，相较对比方法分类错误率下降0.29%～32.02%。

5）在SVHN（1 000个标签）的情况下，FixMatch++错误率为2.19%，相较对比方法分类错误率下降0.29%～1.23%。

2.3　模块消融实验

为深入分析所提各模块对整体模型性能的贡献，分别对A（可学习批量归一化模块LS-BN和双尺度并行卷积模块DSPC）、B（内容/风格分离的双分支表征模块CS-DBR + 动态残差门控机制DRG）和C（多级别伪标签融合机制）进行模块移除消融实验。表3列出了不同标签量配置下的消融实验结果。

表3　模块消融实验结果（不同模块组合对比）

Table 3　Module ablation experiment results (comparison of different module combinations)

/%

方法	A	B	C	CIFAR-10@250	CIFAR-10@400	SVHN@1 000
FixMatch++	√	√	√	**4.56**	**4.12**	**2.19**
–A	–	√	√	4.71	4.19	2.21
–B	√	–	√	4.69	4.15	2.25
–C	√	√	–	4.97	4.23	2.22
–全部	–	–	–	5.07	4.26	2.28

注：加粗字体表示各列最优结果。“–”表示去除该模块，“√”表示使用该模块。

从表3可以看出，增加A模块会导致模型的分类错误率有所下降，尤其是在CIFAR-10（250个标签样本）和SVHN（1 000个标签样本）数据集上，模型分类错误率分别下降了0.15%和0.02%。此外，增

加 B 模块同样导致了分类错误率下降，表明该模块在增强模型表现方面的关键作用。增加 C 模块对模型性能也产生了不小的影响，验证了该模块在 FixMatch++中的重要性。可见，三大模块均对模型性能提升产生了积极贡献。

2.4 Grad-CAM 类激活图分析

为了分析各模块对 FixMatch++的影响，实验使用 Grad-CAM 对 FixMatch++模型进行可视化。

图 5 中选取了 CIFAR-10 数据集中的 10 个常见类别进行实验，对于每个类别，展示了原始图像（CIFAR10 数据集）、完整模型（FixMatch++）、去除模块 A（可学习批量归一化模块（LS-BN）+双尺度卷积模块（DSPC））；去除模块 B（内容/风格分离的双分支表征模块（CS-DBR）+动态残差门控机制（DRG））；去除模块 C（多级别伪标签融合机制的 Grad-CAM 图）的可视化效果。

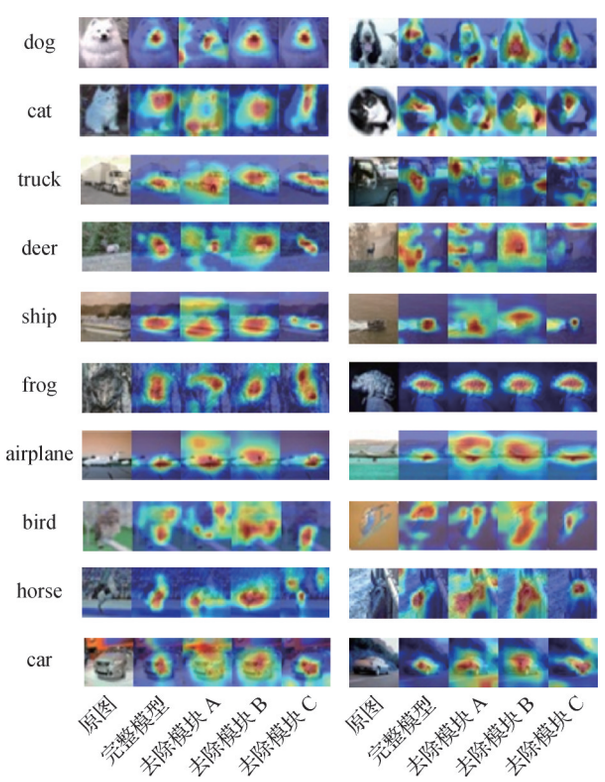


图 5　CIFAR10 数据集上 Grad-CAM 类激活图可视分析

Fig. 5　Grad-CAM class activation map visual analysis on the CIFAR10 dataset

在完整模型下，Grad-CAM 关注区域大多数集中在目标物体的核心部分。例如，dog 图像中，关注的核心区域是狗的头部和身体，图像中的其他部分没有过多的干扰。airplane 图像也显示出飞机的机身和机翼被清晰地标记为关注区域，表明模型对飞机特征的学习较为精确。truck 和 car 等类别的图像显示出车辆的轮廓被较好地识别，关注区域主要集中在车身和车轮部分。

当去除模块 A（用于特征提取的可学习批量归一化模块 + 双尺度并行卷积模块）时，Grad-CAM 图显示出注意力从核心区域扩展到了图像的其他部分。例如，在 dog 图像中，关注区域不仅在狗的头部，还扩展到周围的背景区域，说明特征提取模块对于正确聚焦目标区域起到了关键作用。对于 airplane 类别，注意力也变得更加分散，不仅集中在飞机机身，还扩展到周围的天空区域，导致模型对目标的准确识别有所下降。

当去除模块 B（内容/风格分离的双分支表征模块+动态残差门控机制）后，Grad-CAM 图的变化在 cat 和 frog 等类的图像中尤为明显。注意力虽然依然集中在目标物体上，但焦点相对较弱且较为模糊。例如，frog 图像的关注区域不再集中在青蛙的身体部分，而是扩展到背景和其他不相关区域。对于 truck 图像，尽管目标的核心区域依然受到关注，但背景的干扰更加明显，影响了目标的清晰识别。

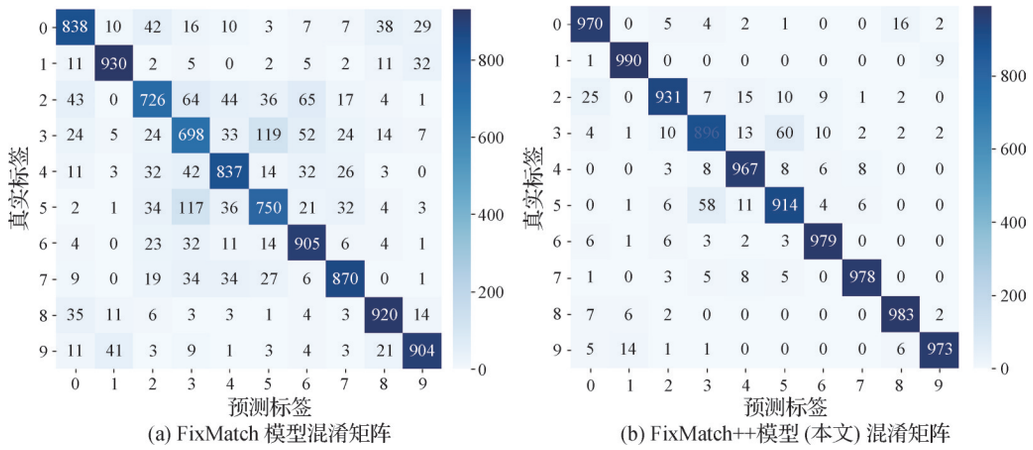
当去除模块 C（多级别伪标签加权融合机制）时，模型的关注区域变得更加广泛和不精确。例如，在 airplane 类别中，Grad-CAM 显示出飞机的机身部分被更广泛的区域覆盖，其他无关区域的干扰增多，表明没有多视图伪标签融合机制时，模型对图像的细节识别能力减弱。类似地，在 bird 和 horse 类别的图像中，Grad-CAM 的关注点扩展到不相关的区域，导致了模型的判别力下降。

通过分析，可以验证每个模块在特征提取和分类决策中的具体作用，进一步证明了 FixMatch++框架在处理多视图伪标签和特征提取时的有效性。

2.5 FixMatch 和 FixMatch++的混淆矩阵对比

图 6 展示了 FixMatch 和 FixMatch++在 CIFAR-10 数据集上的混淆矩阵对比情况。混淆矩阵能够反映模型在不同类别上的分类效果，具体到每个类别的准确性及误分类情况。通过这种对比，能够更清晰地看到模型在不同类别上的表现差异。

从 FixMatch 模型（图 6（a））和 FixMatch++模型（图 6（b））可以明显看出，FixMatch++在多个类别上显著提高了分类精度，从对角线上的变化可以看到，FixMatch++模型大多数类别的正确分类数显著提



(a) FixMatch 模型混淆矩阵　　(b) FixMatch++模型 (本文) 混淆矩阵

图6　FixMatch与FixMatch++模型的混淆矩阵

Fig. 6　Confusion matrices of FixMatch and FixMatch++ models

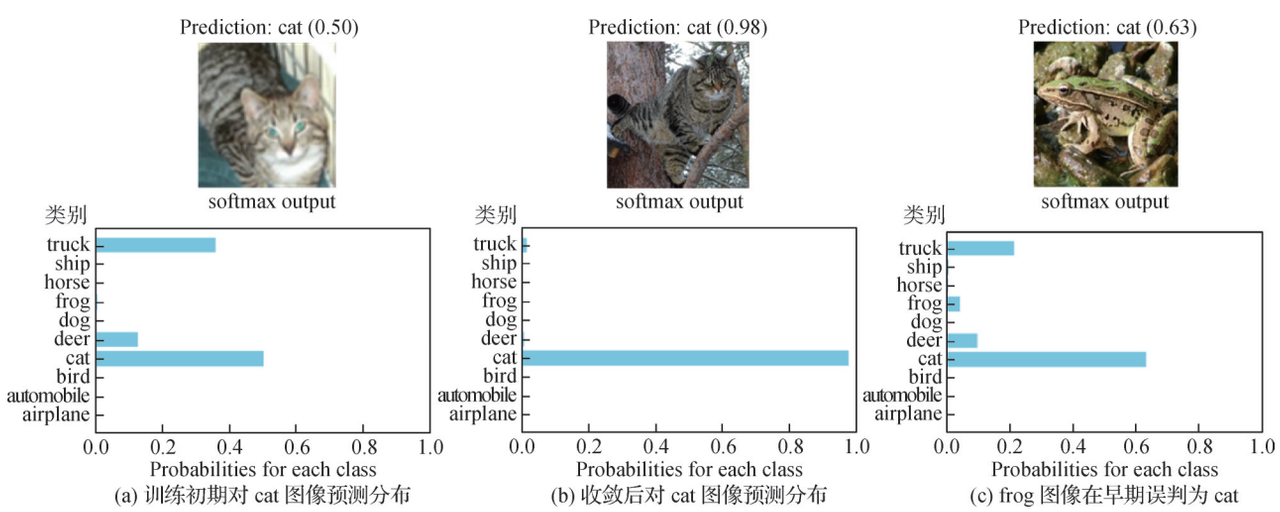
((a) confusion matrix of FixMatch model; (b) confusion matrix of FixMatch++ model (ours))

高,误分类数量明显减少,尤其在airplane和cat类别上,误分类数量减少了50%以上。

特别地,FixMatch模型的误分类主要集中在一些具有较高混淆度的类别。例如,bird类别在FixMatch中存在明显误分类情况,其中,bird样本错误分类为cat和dog。而在FixMatch++模型中,bird类别的误分显著减少,更多的bird类图像被正确分类到bird类别,反映了FixMatch++在细化类别之间的边界方面的优势。此外,在FixMatch++模型中,car和ship类别的误分类也得到了有效减少,模型的稳定性和泛化能力得到了提升。

2.6　FixMatch++在清晰无标签图像上的应用

许多真实业务场景中存在大量清晰且语义明确的无标签图像。若能为此类图像生成高置信度伪标签并扩充有标签训练集,将显著降低人工标注成本并提升小样本场景下的模型表现。如图7所示,将训练好的FixMatch++直接用于清晰图像的标签增强任务。图7(a)显示一幅cat图像在训练初期的预测分布。虽然预测类别为cat,但最大置信度仅约50%,其他类别概率分布分散,不满足阈值要求,不采纳。图7(b)对同一幅cat图像在FixMatch++收敛后再次推断,概率突增至约98%,超过类别阈值,被采纳为可信伪标签,可直接用于扩充训练集。图7(c)示意一幅真实类别为frog的图像在早期被误判为cat(最大置信度约63%),由于未达到该类别的阈值,自动拒绝该伪标签,避免错误样本污染训练。



(a) 训练初期对 cat 图像预测分布　(b) 收敛后对 cat 图像预测分布　(c) frog 图像在早期误判为 cat

图7　FixMatch++在清晰无标签图像上的应用

Fig. 7　Application of FixMatch++ on clear unlabeled images ((a) prediction distribution of a cat image at early training; (b) prediction distribution of the same cat image after convergence; (c) a frog image misclassified as cat at early stage)

综上，FixMatch++可作为一套有限图像标签数据扩展方法，在几乎无人工成本的情况下，为大量易判别样本生成可信伪标签，扩充训练数据、提升模型性能。

3 结 论

本文提出了FixMatch++框架，并在FixMatch模型的基础上进行了多方面的创新和优化。通过设计可学习批量归一化模块（LS-BN）提升训练稳定性，结合双尺度卷积模块（DSPC）增强多尺度特征提取能力；引入内容/风格分离的双分支表征模块（CS-DBR）实现内容与风格特征分离，并采用动态残差门控机制（DRG）优化特征融合过程，提出多级别伪标签融合机制，通过加权融合3种增强视图的分类概率并配合类别阈值筛选策略生成可信标签，有效提升了标签的数量和质量。实验结果表明，FixMatch++在3个公开图像数据集上的表现优于7个近期主流半监督学习方法。通过详细的消融实验和可视化分析，证明了每个模块对提升模型性能的重要性。此外，通过清晰的理论推导、公式计算和实验对比，为半监督学习的研究提供了有力的支持，并为未来的无标签数据利用提供了新方向。未来的工作中，可以探索将本框架应用于更大规模数据集、复杂跨域迁移场景以及其他视觉任务。尤其是在处理不平衡数据时，类似BSGAN-GP（胡静 等，2025）的方法可以进一步提升模型鲁棒性与性能。此外，结合自监督预训练和模型可解释性机制，如胡非易等人（2024）在显微图像分类中的研究，也将有助于提升FixMatch++模型的跨领域适应能力。同时，在安全敏感应用中，例如人脸深度伪造防御任务，研究者也强调半监督方法的重要性（瞿左珉 等，2024），这从应用层面进一步印证了本研究的价值。未来还可结合持续学习等方向（吕凡 等，2025），推动半监督学习方法在动态和多源环境下的广泛应用。

参考文献（References）

Berthelot D, Carlini N, Cubuk E D, Kurakin A, Sohn K, Zhang H, et al. 2020. ReMixMatch: semi-supervised learning with distribution alignment and augmentation anchoring//Proceedings of 2020 International Conference on Learning Representations. Addis Ababa, Ethiopia: OpenReview.net [DOI: 10.48550/arXiv.1911.09785]

Berthelot D, Carlini N, Goodfellow I, Oliver A, Papernot N and Raffel C. 2019. MixMatch: a holistic approach to semi-supervised learning//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #454

Chen H, Tao R, Fan Y, Wang Y D, Wang J D, Schiele B, et al. 2023. SoftMatch: addressing the quantity-quality trade-off in semi-supervised learning [EB/OL]. [2025-08-28]. https://arxiv.org/pdf/2301.10921.pdf

Fan Y, Kukleva A, Dai D X and Schiele B. 2023. Revisiting consistency regularization for semi-supervised learning. International Journal of Computer Vision, 131(3): 626-643 [DOI: 10.1007/s11263-022-01723-4]

Ganin Y and Lempitsky V. 2015. Unsupervised domain adaptation by backpropagation//Proceedings of the 32nd International Conference on Machine Learning. Lille, France: PMLR: 1180-1189

Han D, Kim J and Kim J. 2017. Deep pyramidal residual networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 6307-6315 [DOI: 10.1109/CVPR.2017.668]

Hu F Y, Zhang H, Yuan X F, Liu J X and Chen Y R. 2024. Self-supervised extraction of spectral sequence and semantic information for microscopic cholangiocarcinoma hyperspectral image classification. Journal of Image and Graphics, 29(12): 3817-3832 (胡非易, 张辉, 袁小芳, 刘嘉轩, 陈煜嵘. 2024. 自监督提取光谱序列和语义信息的胆管癌显微高光谱图像分类. 中国图象图形学报, 29(12): 3817-3832) [DOI: 10.11834/jig.230812]

Hu J, Zhang R M and Lian B Q. 2025. BSGAN-GP: a semi-supervised image recognition model driven by class balancing. Journal of Image and Graphics, 30(1): 95-109 (胡静, 张汝敏, 连炳全. 2025. BSGAN-GP: 类别均衡驱动的半监督图像识别模型. 中国图象图形学报, 30(1): 95-109) [DOI: 10.11834/jig.230881]

Hu Z J, Yang Z Y, Hu X F and Nevatia R. 2021. SimPLE: similar pseudo label exploitation for semi-supervised classification//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 15099-15108 [DOI: 10.1109/CVPR46437.2021.01485]

Huang Z Y, Wang H H, Xing E P and Huang D. 2020. Self-challenging improves cross-domain generalization//Proceedings of the 16th European Conference European Conference on Computer Vision. Glasgow, UK: Springer: 124-140 [DOI: 10.1007/978-3-030-58536-5_8]

Laine S and Aila T. 2017. Temporal ensembling for semi-supervised learning//Proceedings of the 5th International Conference on Learning Representations. Toulon, France: ICLR.

Li J, Socher R and Hoi S C H. 2020. DivideMix: learning with noisy labels as semi-supervised learning//Proceedings of the 8th Interna-

tional Conference on Learning Representations. Addis Ababa, Ethiopia: ICLR.

Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017. Feature pyramid networks for object detection//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 2117-2125 [DOI: 10.1109/CVPR.2017.106]

Liu Y, Bao N, Cao K R, Chen J and Wang X. 2025. Dynamic adaptive super-resolution for degraded images in continuous testing scenarios. Journal of Image and Graphics, 30(8): 2645-2659 (刘烨, 鲍娜, 曹克让, 陈吉, 王星. 2025. 连续测试场景中退化图像的动态自适应超分辨率. 中国图象图形学报, 30(8): 2645-2659) [DOI: 10.11834/jig.240764]

Lyu F, Wang L, Li X, Zheng W S, Zhang Z, Zhou T, et al. 2025. Comprehensive survey of continual learning. Journal of Image and Graphics, 30(8): 2599-2632 (吕凡, 王亮, 李玺, 郑伟诗, 张彰, 周涛, 等. 2025. 持续学习研究进展. 中国图象图形学报, 30(8): 2599-2632) [DOI: 10.11834/jig.240661]

Miyato T, Maeda S I, Koyama M and Ishii S. 2019. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(8): 1979-1993 [DOI: 10.1109/TPAMI.2018.2858821]

Pham H, Dai Z G, Xie Q Z and Le Q V. 2021. Meta pseudo labels//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 11552-11563 [DOI: 10.1109/CVPR46437.2021.01139]

Qi T H, Fang S C, Wu Y Z, Xie H T, Liu J W, Chen L, et al. 2024. DEADiff: an efficient stylization diffusion model with disentangled representations//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 8693-8702 [DOI: 10.1109/CVPR52733.2024.00830]

Qu Z M, Yin Q L, Sheng Z Q, Wu J Y, Zhang B L, Yu S R, et al. 2024. Overview of Deepfake proactive defense techniques. Journal of Image and Graphics, 29(2): 318-342 (瞿左珉, 殷琪林, 盛紫琦, 吴俊彦, 张博林, 余尚戎, 等. 2024. 人脸深度伪造主动防御技术综述. 中国图象图形学报, 29(2): 318-342) [DOI: 10.11834/jig.230128]

Sohn K, Berthelot D, Li C L, Zhang Z Z, Carlini N, Cubuk E D, et al. 2020. FixMatch: simplifying semi-supervised learning with consistency and confidence//Proceedings of the 34th International Conference on Neural Information Processing System. Vancouver, Canada: Curran Associates Inc.: #51

Tarvainen A and Valpola H. 2017. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results//Proceedings of the 31st International Conference on Neural Information Processing System. Long Beach, USA: Curran Associates Inc.: 1195-1204

Wang J D, Sun K, Cheng T H, Jiang B R, Deng C R, Zhao Y, et al. 2021. Deep high-resolution representation learning for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43 (10): 3349-3364 [DOI: 10.1109/TPAMI.2020.2983686]

Wang W H, Xie E Z, Li X, Fan D P, Song K T, Liang D, et al. 2022. PVT v2: improved baselines with pyramid vision transformer. Computational Visual Media, 8(3): 415-424 [DOI: 10.1007/s41095-022-0274-8]

Wang X, Chen H, Tang S A, Wu Z H and Zhu W W. 2024. Disentangled representation learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12): 9677-9696 [DOI: 10.1109/TPAMI.2024.3420937]

Wang Y D, Chen H, Heng Q, Hou W X, Fan Y, Wu Z, et al. 2023. FreeMatch: self-adaptive thresholding for semi-supervised learning//Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: ICLR

Xie Q Z, Dai Z H, Hovy E, Luong M T and Le Q V. 2020. Unsupervised data augmentation for consistency training//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #525

Zhang B, Wang Y, Hou W, Wu H, Wang J and Okumura M. 2021. FlexMatch: boosting semi-supervised learning with curriculum pseudo labeling//Proceedings of the 35th Conference on Neural Information Processing Systems. Virtual Event: Neural Information Processing Systems Foundation [DOI: 10.48550/arXiv.2110.08263]

Zhou Y F, Du K L, Lyu F, Hu F Y and Liu G C. 2025. Class activation map replay and minimum entropy sampling-based multilabel class-incremental learning. Journal of Image and Graphics, 30 (8): 2633-2644 (周怡凡, 杜凯乐, 吕凡, 胡伏原, 刘光灿. 2025. 类激活图回放和最小熵采样的多标签类增量学习. 中国图象图形学报, 30(8): 2633-2644) [DOI: 10.11834/jig.240643]

作者简介

杨雨晴,女,硕士研究生,主要研究方向为可视分析、数据增强和图像分类。E-mail:yangyuqing@st.btbu.edu.cn

陈谊,通信作者,女,教授,主要研究方向为可视分析、机器学习、图像数据增强。E-mail:chenyi@th.btbu.edu.cn

吕程,女,博士后,主要研究方向为深度学习、模型可解释性和可视分析。E-mail:lvcheng@btbu.edu.cn