DOI: 10.13718/j.cnki.xdzk.2026.04.013
朱宇峰，陈玲，黄嘉衍，等. L-CSNet：基于高效多尺度膨胀卷积的轻量级计数监督网络在小麦穗计数中的应用[J]. 西南大学学报(自然科学版)，2026，48(4)：182-194.

L-CSNet：基于高效多尺度膨胀卷积的轻量级计数监督网络在小麦穗计数中的应用

朱宇峰[1]，　陈玲[1,2]，　黄嘉衍[1]，　李传东[1]，　黄廷文[3]，　曾晓洋[2]

1. 西南大学 电子信息工程学院，重庆 400715；2. 复旦大学 集成芯片与系统全国重点实验室 上海 200433；
3. 深圳理工大学 计算机科学与控制工程学院，广东 深圳 518055

摘要：目前的深度学习方法解决小麦穗计数问题时都严重依赖昂贵的位置级标注(诸如边界框或密度图)，因而需要大量人工标注工作，也容易引入标注噪声，限制了其在农业领域中的实际应用。为了解决这一问题，提出了一种新型的轻量级计数监督网络，此网络使用图像级计数标签实现了高精度的小麦穗计数，摆脱了对位置信息的需求。其核心创新点是设计了一种高效多尺度膨胀卷积(Efficient Multi-Scale Dilated Convolution，EMDC)模块，用并行膨胀卷积替代传统的计算密集型结构，高效地提取了多尺度特征，且极大降低了模型参数量。在一个公开的麦穗检测数据集上进行了系统实验，结果显示其平均绝对误差与均方根误差都显著优于现有的位置监督方法，推理速度达到了 120 帧/秒(FPS)，适合在资源受限设备(如无人机或移动传感器)上进行实时部署。

关　键　词：小麦穗计数；计数监督；多尺度膨胀卷积；轻量级网络；农业自动化

中图分类号：TP29　　**文献标识码**：A　　开放科学(资源服务)标识码(OSID)：

文 章 编 号：1673-9868(2026)04-0182-13

L-CSNet：Application of a Lightweight Count-Supervised Network Based on Efficient Multi-Scale Dilated Convolution in Wheat Ear Counting

ZHU Yufeng[1]，　CHEN Ling[1,2]，　HUANG Jiayan[1]，
LI Chuandong[1]，　HUANG Tingwen[3]，　ZENG Xiaoyang[2]

收稿日期：2026-01-22
基金项目：国家自然科学基金区域创新发展联合基金重点项目(U25A20474)；国家自然科学基金项目(62373310)；重庆市自然科学基金面上项目(CSTB2024NSCQMSX0428)；重庆市教委项目(KJQN202300207)；集成芯片与系统全国重点实验室高级访问学者计划项目(SKLICS-G202503)。
作者简介：朱宇峰，硕士研究生，主要从事机器学习以及机器学习相关技术在农业上的应用研究。
通信作者：陈玲，副教授，硕士研究生导师。

1. College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China;
2. State Key Laboratory of Integrated Chips and Systems, Fudan University, Shanghai 200433, China;
3. Faculty of Computer Science and Control Engineering, Shenzhen University of Advanced Technology, Shenzhen Guangdong 518055, China

Abstract: Current deep learning methods for wheat ear counting predominantly rely on expensive location-level annotations such as bounding boxes or density maps, which require considerable manual annotation effort and are susceptible to annotation noise, thus limiting their practical use in agriculture. To address this issue, this study proposed a new lightweight count-supervised network for high-precision wheat ear counting based on image-level count labels, without requiring location information. The core innovation lay in the design of an efficient multi-scale dilated convolution (EMDC) module, which replaced the traditional computationally expensive structures with parallel dilated convolutions. This enabled efficient extraction of multi-scale features while keeping the number of model parameters to a minimum. Systematic experiments on a public wheat ear detection dataset demonstrated that the proposed method significantly outperforms existing position-supervised approaches in terms of both the mean absolute error (MAE) and root mean square error (RMSE). With an inference speed of 120 frames per second, the network was highly suitable for real-time deployment on resource-constrained devices such as unmanned aerial vehicles (UAVs) or mobile sensors.

Key words: wheat ear counting; counting supervision; multi-scale dilated convolution; lightweightnetwork; agricultural automation

自动化技术对提高育种效率和粮食产量意义重大。小麦作为全球重要的粮食作物，为人类提供了约20%的蛋白质与碳水化合物[1]，且在工业原料、生物燃料以及动物饲料等众多领域有着广泛应用。但小麦产量的增长速度已跟不上不断提升的社会发展需求。有数据表明，小麦需求的年增长率为1.7%，而其遗传增益的年均增长率仅为1%[2]。自动化技术在农业领域得到广泛运用，如利用软硬件协同，加速农作物病害鉴定[3]，或通过自动化技术替代人工对作物表型(如株高、颜色、麦穗数量等)进行统计分析[4]，减少了人力和时间成本，促进了高效育种的开展。

自动化计数筛选具有优良性状的品种是小麦育种中的核心环节。小麦产量作为关键的育种性状，由单位面积的麦穗数、单穗粒数和千粒质量这3个要素共同决定[5]。传统的育种方式不仅效率低下、耗费大量时间和人力，还容易因为人为操作产生较高误差。所以，实现麦穗的自动化计数对于提高育种效率、节省人力资源十分重要。为了实现这一目标，研究人员早期尝试通过图像处理技术来识别麦穗：文献[6]利用颜色和纹理特征处理技术实现了图像中小麦穗的分割；文献[7]结合Gabor滤波器与K-means聚类算法，完成了麦穗区域的检测与计数。不过，这些传统方法的泛化能力有限，容易受到光照、环境等干扰因素的影响，难以适应复杂场景。

随着深度学习技术的不断突破，以边界框监督和点监督为代表的位置监督方法在麦穗计数领域受到了广泛关注[8-13]。这些方法在一定程度上解决了一部分传统方法泛化能力差、抗噪声能力弱的问题，但都依赖高成本的位置级图像进行训练，并且边界框监督和点监督的麦穗计数模型大多基于局部感知的卷积神经网络[14]。然而麦穗密集且多样的位置信息不仅标注成本高昂，还会引入噪声制约模型性能。

本文借鉴了针对人群的计数监督方法[15-18]，构建了一个只需图像级计数标签即可做到高精度的麦穗计

数网络模型（A Lightweight Count-Supervised Network via Efficient Multi-Scale Dilated Convolution，L-CSNet），其主要的创新之处在于设计了高效多尺度膨胀卷积模块（Efficient Multi-Scale Dilated Convolution，EMDC），即用并行膨胀卷积[19]替代计算密集型结构，在高效捕获多尺度麦穗特征的同时将模型参数量大大降低，且没有削弱模型的性能。在 GWHD_2020[20]和 GWHD_2021[21]数据集上的实验结果可以清楚地证明，L-CSNet 在平均绝对误差（MAE）和均方根误差（RMSE）等指标上都优于现有方法，因此它也是自动化农业计数任务中更经济实用、更易于推广的理想选择，可以更好地推动深度学习在实际农业场景中的应用落地。通过以上论述，本文的主要研究工作有以下几点：

1）提出了轻量级计数监督网络 L-CSNet：该网络只需对图像级计数标签进行训练，并将模型参数量显著压缩，为模型在边缘设备上的实时部署提供实际支撑。

2）设计了高效多尺度膨胀卷积（EMDC）模块：该模块通过通道优化、并行多尺度膨胀卷积和轻量级注意力机制，实现了高效且鲁棒的多尺度特征感知。

3）构建架构与训练协同优化策略：从模型架构和训练策略两方面进行协同设计与优化，确保了训练的稳定性和模型最终性能。

1　材料与方法

1.1　数据集

小麦作为全球广泛种植的农作物之一，其品种因各种自然条件的区域差异呈现出显著多样性。为验证 L-CSNet 的有效性和泛化能力，本文选取农业表型领域覆盖范围最广且标注规范的全球小麦检测（GWHD）系列数据集作为实验基准（图 1）。GWHD_2020[20]涵盖世界多个地区的 4 700 张 RGB 图像，共标注 193 634 个小麦穗，采集后经数据协调处理确保了所有样本的可视化一致性。该数据集包含不同生长阶段与基因型的小麦品种，麦穗性状丰富多样，可实际检验模型对不同场景的适配能力。GWHD_2021[21]为 GWHD_2020 的扩展版本，新增了 5 个国家的 1 722 张图像与 81 553 个标注小麦穗，更适用于高密度、复杂背景下的模型性能测试。

GWHD_2020 和 GWHD_2021 的单张图像平均麦穗数量分别为 42.77 和 42.22，其中：中等密度图像占比约 75%，是数据集的核心组成部分，让模型能充分学习田间最常见场景的麦穗特征；低密度与高密度图像占比约 25%，为模型提供了极端场景的特征学习样本，保证了模型的泛化能力。本文从 GWHD_2020 和 GWHD_2021 中选取部分数据集进行筛选与划分，将每个数据集按 8∶1∶1 的比例划分为训练集、验证集与测试集（表 1）。

表 1　GWHD 数据集使用统计信息

数据集	数据子集	图像数量/副	*Min*	*Max*	*Avg*	*Total*
GWHD_2020	训练集	2 712	0	112	42.77	116 012
	验证集	339	0	88	44.31	15 024
	测试集	339	0	94	40.13	13 603
GWHD_2021	训练集	5184	0	190	42.22	218 917
	验证集	648	0	146	41.19	26 692
	测试集	648	0	159	42.34	27 434

注：*Min*、*Max*、*Avg*、*Total* 分别表示对单张图像标注的麦穗数进行统计后得到的麦穗最小数量、最大数量、平均数量与总数量。

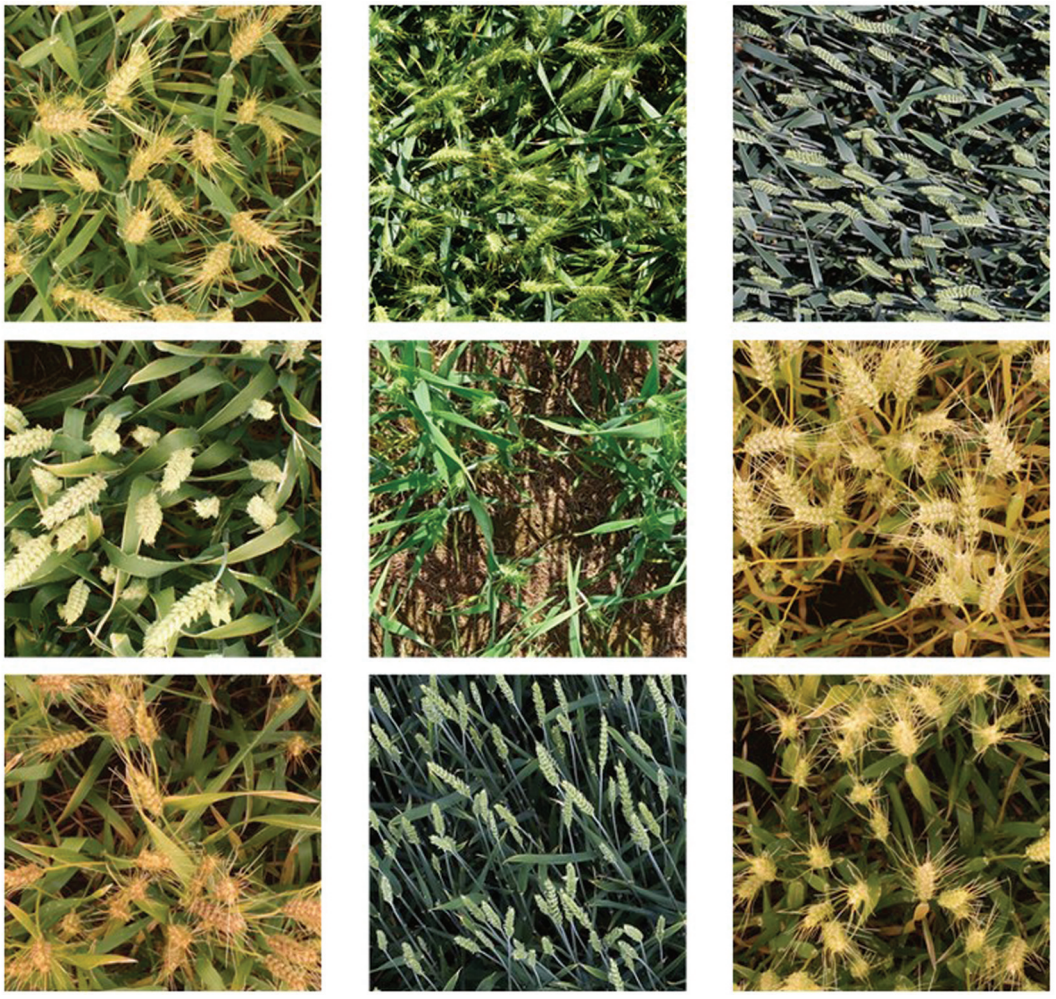


图1 GWHD_2020数据集图示

1.2 方法与L-CSNet网络架构

为实现“轻量级、高精度、低成本”的小麦穗计数目标，本文提出的轻量级计数监督网络(L-CSNet)以计数监督为核心设计理念，摆脱了传统位置监督对边界框或密度图的依赖，仅通过图像级计数标签完成训练。L-CSNet整体架构由改进型骨干网络(ModifiedBackbone)、高效多尺度膨胀卷积(EMDC)模块和简化计数头(CounterHead)3部分构成(图2)。

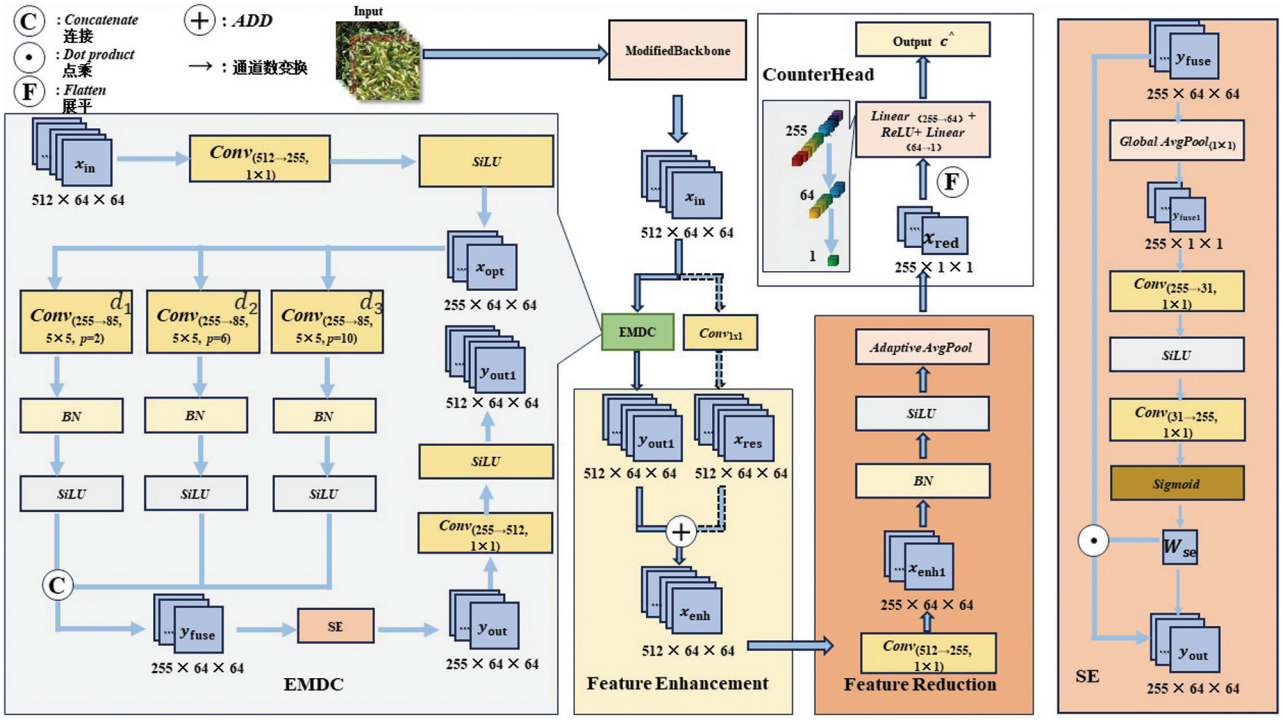
1.3 整体架构设计

1.3.1 改进型骨干网络

骨干网络是模型进行特征提取的基础，决定了后续模块的表达能力与计算效率。为能在“特征表达能力”与“参数量控制”之间取得平衡，本文基于VGG16[22]网络进行裁剪优化，构建了ModifiedBackbone。该ModifiedBackbone保留了对边缘、纹理等底层视觉特征的提取能力，同时剔除了冗余的全连接层与部分池化层，从而显著降低了计算开销。具体结构包括10个3×3卷积层与3个2×2最大池化层。该骨干网络的参数量仅为7.6×10^6个，相较完整VGG16降低了约65%，并且依托ImageNet预训练权重进行初始化[23]，减少了对大规模训练数据的依赖并加速模型收敛。

1.3.2 高效多尺度膨胀卷积模块

高效多尺度膨胀卷积模块(EMDC)如图2所示。模块依次完成通道优化、并行膨胀卷积、通道注意力与通道恢复及残差适配4个步骤。其中，X_{bb}与X_{in}均表示ModifiedBackbone输出的特征图，在残差连接



模块内数字表示特征图维度“通道数×高度×宽度”，虚线代表残差连接，灰色模块为核心创新组件；*Conv* 为卷积层卷积操作，*BN* 为批归一化，*GlobalAvgPool* 为全局平均池化，*p* 为 *padding*，*Linear* 为线性层，*d* 为膨胀分支，Feature Enhancement 为特征加强模块，Feature Reduction 为特征降维模块。

图 2　L-CSNet 模型架构流程图

部分记为 X_{bb}，在 EMDC 输入部分记为 X_{in}，以区分其在不同模块中的作用。

为了防止特征图 X_{in} 直接进入多分支卷积造成冗余计算，ModifiedBackbone 输出的特征图 X_{in} 先通过 1×1 的卷积将通道压缩至 255(512→255，其中“→”为通道数变换操作)，用 *SiLU*(Sigmoid Linear Unit)激活函数提升非线性表达得到特征图 X_{opt}：

$$X_{opt}=SiLU(Conv_{1\times1}(X_{in},\ 512\rightarrow255)) \tag{1}$$

模型将压缩后的特征均匀分配至三路，通过设计的三路并行 5×5 卷积后得到特征图 Y_k，膨胀率分别为 $d_1=1$、$d_2=3$、$d_3=5$，捕获不同尺度的小麦穗特征，三路通道中的 d_1 用于聚焦细节，d_2 用于适配中等尺度，d_3 则用于捕获全局信息。三路的输出最后拼接(*Concat*)得到融合特征 Y_{fuse}：

$$Y_k=SiLU(BN(Conv_{5\times5,\ d_k}(X_{opt},\ 255\rightarrow85))),\ k\in1,2,3 \tag{2}$$

$$Y_{fuse}=Concat(Y_1,\ Y_2,\ Y_3)\in R^{255\times64\times64} \tag{3}$$

如图 2 所示，在模型中引入改造后的轻量化 SE 注意力机制[24]，目的是突出模型的有效通道，抑制背景噪声的干扰。模块在对 Y_{fuse} 进行全局平均池化(Global AvgPool，GAP)之后得到了融合特征 Y_{fuse1}，再经过两层 1×1 的卷积与 *SiLU* 激活，最后使用 Sigmoid 归一化得到通道权重向量 $\boldsymbol{W}_{se}$，$\boldsymbol{W}_{se}$ 与融合特征逐元素相乘并以 1×1 的卷积与 *SiLU* 进行整合得到特征图 Y_{out}：

$$\boldsymbol{W}_{se}=\sigma(Conv_{1\times1}(SiLU(Conv_{1\times1}(GAP(Y_{fuse}))))) \tag{4}$$

$$Y_{out}=SiLU(Conv_{1\times1}(Y_{fuse1}\otimes W_{se})) \tag{5}$$

公式(5)中的⊗表示逐元素乘法，经过加权后的特征图为 255 通道。对特征图使用 1×1 的卷积将通道恢复至 512，和残差分支对齐得到特征图 Y_{out1} 再与残差映射相加得到增强的特征图 X_{enh}：

$$Y_{out1}=SiLU(Conv_{1\times1}(Y_{out},\ 255\rightarrow512)) \tag{6}$$

$$X_{enh}=Y_{out1}+X_{res} \tag{7}$$

这种设计将原始的语义信息与多尺度增强后的特征信息有效融合，保证了梯度的稳定传递，大幅提升了模型在复杂田间背景下的检测鲁棒性。

1.3.3 残差连接与特征降维层

如图2中Featuer Reduction模块所示，在ModifiedBackbone与EMDC之间引入了残差连接，此设计有效地缓解了深层训练中的梯度消失，确保了特征的稳定分布。模型的ModifiedBackbone输出X_{bb}后进行1×1的卷积形成了恒等映射X_{res}，再与EMDC的输出逐元素相加得到X_{enh}：

$$X_{res}=Conv_{1\times1}(X_{bb},\ 512\rightarrow 512) \tag{8}$$

$$X_{enh}=Y_{out}+X_{res} \tag{9}$$

模型在最后进行了特征降维操作：在通过1×1的卷积后将通道由512压缩至256，经过BN层与$SiLU$优化分布后，再通过自适应平均池化操作将空间维度压缩至1×1，得到全局特征向量$\boldsymbol{X}_{red}$：

$$\boldsymbol{X}_{red}=AdaptiveAvgPool_{1\times1}(SiLU(BN(Conv_{1\times1}(X_{enh},\ 512\rightarrow 256)))) \tag{10}$$

1.3.4 简化计数头

如图2所示，将控制模型输出的计数头(CounterHead)设计为“扁平化—全连接回归”结构，这种设计将全局特征映射为计数结果。先通过扁平化操作将$\boldsymbol{X}_{red}$转化为一个256维的特征向量，再经过两层全连接层与$ReLU$(Rectified Linear Unit)激活函数输出的最终预测计数为$\hat{C}$：

$$\hat{C}=FC_{64\rightarrow1}(ReLU\ (Flatten(\boldsymbol{X}_{red}))) \tag{11}$$

使用了该设计的简化计数头参数量仅约为42 K，只占模型总参数量的极小部分，在最大程度上保证了预测的准确性，而且大幅降低了推理延迟与模型参数量。

1.4 模型训练策略与机制讨论

设计并采用了一种分阶段训练策略的模型训练方式，目的是确保L-CSNet在计数监督场景下的稳定性与精度。在第一阶段的训练中冻结了ModifiedBackbone的浅层参数，仅微调了EMDC模块、特征降维层与计数头，让新设计的模块能够快速学习小麦穗的特征分布。第一阶段的训练轮数为40，学习率设为1×10^{-5}。第二阶段的训练则是解冻所有层进行全局微调，进一步优化了整体网络结构的参数，保证了模型能够在多样化场景下保持稳定表现。设置第二阶段的训练轮数为460，初始学习率为1×10^{-5}，结合MultiStepLR学习率调度策略，控制学习率在训练到第200轮时衰减至1×10^{-6}。

采用Huber损失函数，让模型在面对异常样本时具有更强的鲁棒性。Huber损失函数结合了L1损失(绝对误差)与L2损失(均方误差)的优点，在误差较小时表现为平方损失，在误差较大时转化为线性损失，Huber损失函数的使用避免了L2损失对异常值过度敏感的问题：

$$L(\hat{C},\ C)=\begin{cases}\dfrac{1}{2}(\hat{C}-C)^2 & |\hat{C}-C|\leqslant\delta\\ \delta\cdot(|\hat{C}-C|-\dfrac{1}{2}\delta) & |\hat{C}-C|>\delta\end{cases} \tag{12}$$

其中：$\hat{C}$表示预测计数；C表示真实计数；$\delta=1$为分段阈值。L-CSNet以计数监督为核心设计理念，其核心设计逻辑为麦穗计数结果与麦穗多尺度视觉特征的空间密度存在强线性关联，因此无需位置标注，仅通过预测计数与真实计数的误差损失，即可通过梯度反向传播指导网络自发捕获麦穗特征。该损失并非直接学习“麦穗位置”，而是通过标量误差的空间化梯度分配与多尺度特征的密度感知拟合，让网络聚焦于麦穗的边缘、纹理、麦穗密度等核心视觉特征，同时抑制土壤、杂草等背景干扰；而EMDC模块的多尺度膨胀卷积与轻量化注意力机制，为特征学习过程提供了高效的提取框架，这是实现“无位置标注但有特征精准聚焦”的计数监督学习方法的核心原因。

1.5　评估指标

在评估指标的选择上，本文采用平均绝对误差（*MAE*）衡量模型的计数准确性，采用均方根误差（*RMSE*）衡量模型的稳定性，采用决定系数（R^2）衡量模型的拟合程度：

$$MAE=\frac{1}{N}\sum_{i=1}^{N}|\hat{C}_i-C_i| \tag{13}$$

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{C}_i-C_i)^2} \tag{14}$$

$$R^2=1-\frac{\sum_{i=1}^{N}(\hat{C}_i-C_i)^2}{\sum_{i=1}^{N}(C_i-\overline{C})^2} \tag{15}$$

其中：N 表示测试集图像数量；$\hat{C}_i$ 为第 i 张图像的预测计数；C_i 为真实计数；$\overline{C}$ 为测试集所有图像的平均真实计数。

2　结果与分析

2.1　实验细节

2.1.1　训练配置

L-CSNet 的优化算法采用随机梯度下降算法，并设置动量为 0.9，权重衰减系数为 5×10^{-4}。分两阶段调整学习率：前 40 轮冻结 ModifiedBackbone 浅层参数，设学习率为 1×10^{-5}，仅微调 EMDC 模块与计数头；41－500 轮解冻所有层，学习率保持 1×10^{-5}，采用 MultiStepLR 调度器（学习率在第 200 轮衰减至 1×10^{-6}）。

硬件环境为 NVIDIA 3070Ti GPU（8GB 显存）、AMD Ryzen 55600 CPU，单卡训练总时长约 325 h。

2.1.2　对比方法选择

选取 3 类主流小麦穗计数方法与 L-CSNet 进行对比实验，其中：边界框监督方法 Faster R-CNN[8]、SSD[25]、DETR[26]、YOLOv11[10]均为不同时期目标检测领域的代表性模型，需人工标注边界框；点监督方法 MCNN[27]、CSRNet[28]、WheatNet[12]通过预测密度图实现计数，需人工标注麦穗中心点；计数监督方法 TransCrowd[16]、CSNet[18]仅需图像级计数标签，与 L-CSNet 监督方式一致。

2.2　性能对比实验

2.2.1　GWHD 数据集上的定量对比

通过观察表 2 中的数据发现，L-CSNet 网络在 GWHD_2020 数据集上表现优异，*MAE* 和 *RMSE* 在所有对比方法中仅次于 CSNet。在 GWHD_2021 更具挑战性的密集场景数据集上，L-CSNet 网络的 *MAE* 与 *RMSE* 表现稳定，模型性能与最佳的 CSNet 预测结果非常接近。该实验结果证明了 EMDC 模块通过多尺度膨胀卷积能够有效地捕捉不同密度场景下的麦穗特征，模型具有较强的泛化能力。

L-CSNet 网络的模型参数量仅为 9.96×10^6，远低于其他计数监督方法，且优于多数轻量化的点监督模型。同时，L-CSNet 网络具备高速的推理能力，其推理速度达到了 120 FPS，在表 2 的所有对比方法中速度最快。推理速度快与参数量低的特点，有利于 L-CSNet 在资源受限的设备上实时部署。

通过对比两个数据集的不同结果可得出结论：由于 GWHD_2021 因包含更多密集和复杂场景的图像，表中所有模型的 *MAE* 和 *RMSE* 都存在明显上升，但 L-CSNet 的增幅平稳。这验证了 L-CSNet 模型的鲁棒性。

表2 L-CSNet与主流方法在GWHD数据集上的性能对比

方法	年份	监督类型	参数量/$\times10^6$个	推理速度/FPS	GWHD_2020		GWHD_2021	
					MAE	*RMSE*	*MAE*	*RMSE*
FasterR-CNN[8]	2015	边界框	42.3	18	3.52	4.47	5.48	9.78
SSD[25]	2016	边界框	28.7	25	3.98	5.50	7.32	12.11
DETR[26]	2022	边界框	86.5	12	3.76	4.97	8.43	16.23
YOLOv11[10]	2025	边界框	36.9	32	4.10	6.01	8.21	12.30
MCNN[27]	2016	点	15.2	45	5.05	6.56	7.34	10.19
CSRNet[28]	2018	点	10.8	38	3.87	5.01	5.88	8.18
WheatNet[12]	2022	点	4.04	85	3.85	5.19	6.03	7.89
TransCrowd-GAP[16]	2022	计数	56.8	22	12.23	15.02	10.09	12.63
CSNet[18]	2024	计数	109.2	68	2.96	3.98	3.77	5.66
L-CSNet	2026	计数	9.96	120	3.04	3.94	3.96	5.73

2.2.2 计数拟合程度分析

为了进一步探究L-CSNet模型的拟合程度，采用R^2来衡量模型预测值与真实值的线性拟合程度。实验结果如表3所示。

表3 各模型在GWHD_2020数据集上的R^2拟合程度

模型名称	年份	方法类别	R^2值
Faster R-CNN	2015	边界框	0.882 8
SSD	2016	边界框	0.901 6
DETR	2022	边界框	0.925 5
YOLOv11	2025	边界框	0.942 9
MCNN	2016	点	0.916 4
CSRNet	2018	点	0.934 9
WheatNet	2022	点	0.948 1
TransCrowd-GAP	2022	计数	0.951 2
CSNet	2024	计数	0.956 8
L-CSNet	2025	计数	0.953 2

由表3可知，L-CSNet在小麦穗计数任务中具有优秀的可靠性与泛化能力，在各类方法中整体拟合度位于前列。

2.2.3 不同骨干的影响

为了进一步验证骨干网络对L-CSNet性能的影响，选取了多种经典的网络进行对比，其中包括VGG16[22]、ResNet34[29]、ResNet50[29]、MobileNetV3[30]、EfficientNet-B0[31]，以及Swin-Transformer[32]。所有骨干网络均在imageNet上预训练。实验结果如表4所示。

由表4可知，采用VGG16时，L-CSNet获得了最优的计数精度。这与任务特性密切相关：小麦穗计数主要依赖边缘与纹理特征，而VGG16的浅层卷积结构更适合这一场景。而ResNet50虽然参数量更大，但表现不佳，说明过深的残差结构可能在简单任务中导致过拟合。

在轻量化骨干方面，MobileNetV3在GWHD_2021上的精度不佳，说明其更适合简单场景需求。EfficientNet-B0表现较为均衡，但精度略低于VGG16。值得注意的是，Swin-Transformer在两个数据集上的表现均不理想，说明该类结构可能更适用于复杂度更高的大规模数据。

实验结果表明骨干网络选择对 L-CSNet 性能影响显著，合理选择骨干网络是平衡准确性、泛化性与效率的关键。VGG16 是当前任务的最优选择。

表 4　L-CSNet 在不同骨干下的性能比较(GWHD 数据集)

骨干网络	参数量/$\times10^6$ 个	GWHD_2020		GWHD_2021	
		MAE	*RMSE*	*MAE*	*RMSE*
VGG16	7.6	3.04	3.94	3.96	5.73
ResNet18	21.2	3.21	4.36	4.89	6.92
ResNet50	25.6	4.41	5.88	6.12	8.97
MobileNetV3	0.8	3.12	4.28	7.21	9.84
EfficientNet-B0	5.3	3.36	4.51	4.62	6.77
Swin-Transformer	28.4	8.94	11.87	13.82	17.64

2.3　关键组件消融实验

为了验证 L-CSNet 各核心组件的必要性与有效性，在 GWHD_2020 数据集上进行了系统性的消融实验。实验以“ModifiedBackbone＋CounterHead”为基础方案(Baseline)，所有实验均在统一条件下运行，结果如下。

2.3.1　核心组件消融

如表 5 所示，EMDC 模块是性能提升的关键，多尺度膨胀卷积能够有效捕捉不同大小的小麦穗特征。而残差连接可以进一步降低误差，说明该措施缓解了梯度消失，使底层边缘与纹理特征能够更好地传递。两阶段训练带来了额外的准确性提升，原因在于冻结浅层保护了预训练特征，避免破坏底层信息，而后期全局微调则优化了整体适配。三者结合使 L-CSNet 在保持轻量化的同时实现了高精度与高鲁棒性。

表 5　L-CSNet 核心组件消融实验结果

配置	EMDC 模块	残差连接	两阶段训练	参数量/$\times10^6$ 个	*MAE*	*RMSE*
Baseline	×	×	×	7.6	5.08	6.79
Baseline＋EMDC(三分支，d＝1/2/3，SE＝on)	√	×	×	9.9	3.21	4.47
Baseline＋EMDC＋残差(1×1 维度匹配)	√	√	×	10.2	3.7	4.18
Baseline＋EMDC＋残差＋两阶段训练	√	√	√	9.96	3.04	3.94

2.3.2　EMDC 微结构消融

如表 6 所示，三分支设计 e＝1/3/5(e 为膨胀率，1/3/5 表示三分支膨胀率分别为 1，3，5)能够覆盖麦穗的多样性。四分支并未带来额外收益，原因在于过多分支导致特征冗余，增加了参数量并降低速度。并且 SE 注意力机制不可或缺，在其关闭后误差显著上升，说明了通道权重调整对抑制背景干扰至关重要。$\mathrm{Conv}_{1\times1}$ 替代 FC(全连接层)更高效，可以解释为保持张量结构减少了计算开销。而通道压缩过度显示出性能下降，说明通道过少不足以捕捉复杂纹理。

表 6　L-CSNet EMDC 微结构消融实验结果

EMDC 设计	参数量/$\times10^6$ 个	*MAE*	*RMSE*
两分支(e＝1/3，SE＝on)	8.95	3.36	4.61
三分支(e＝1/3/5，SE＝on)	9.96	3.04	3.94
四分支(e＝1/2/4/6，SE＝on)	12.8	3.10	4.23
三分支(e＝1/3/5，SE＝off)	8.32	3.41	4.69
三分支(SE：FC 替换 Conv1×1)	10.6	3.16	4.31
通道压缩(255→192)	10.2	3.24	4.52

2.3.3 计数头与池化策略消融

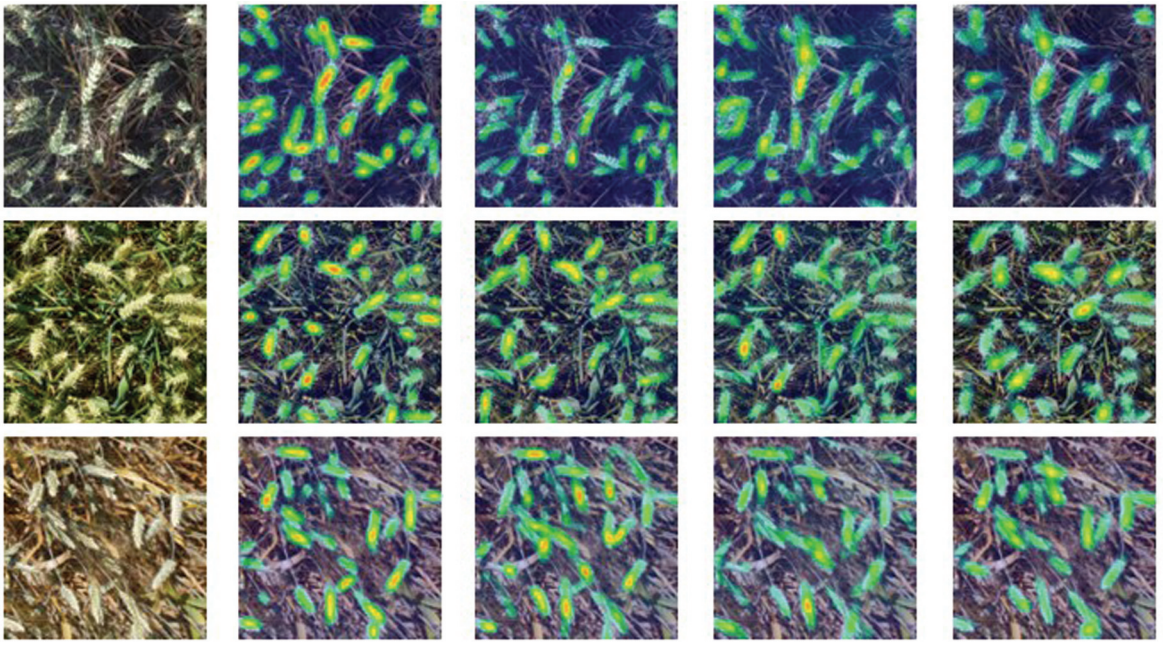
如表7所示，CounterHead宽度增加(64→128)并未提升性能，结果显示由于任务本身较为简单，宽度过宽反而浪费参数。*GeM-pool*(Generalized Mean Pooling，广义平均池化)略优于*GAP*，说明其能更灵活地调整聚合方式，适合密集场景。$AvgPool_{2\times2}$表现较差，原因在于空间压缩过早导致信息丢失。综上，*GAP*+*Lineard*的组合是最好的计数头方案，既轻量又准确。

表7 计数头与池化策略消融实验结果

计数头与池化策略	参数量/$\times10^6$个	*MAE*	*RMSE*
GAP+*Linear*(256→64→1)	9.96	3.04	3.94
GAP+*Linear*(256→128→1)	10.1	3.32	4.03
GeM-pool+*Linear*(256→64→1)	9.96	2.90	3.93
AvgPool 2×2 *stride* 2+*Linear*	10.2	3.06	4.21

2.4 注意力热力图对比

为直观对比L-CSNet与CSNet、YOLOv11以及TransCrowd-GAP在小麦穗区域的关注能力，采用了Grad-CAM技术[33]生成模型最后一层卷积层的注意力热力图(图3)。



a. 原始图像 b. L-CSNet c. CSNet d. Yolo11 e. TransCrowd-GAP

红色区域表示模型高注意力，蓝色区域表示低注意力。

图3 不同模型的注意力热力图

如图3所示，L-CSNet的热力图红色区域精准覆盖麦穗主体，对杂草、土壤等背景的关注度较低，这表明EMDC模块中的SE通道注意力能够有效强化麦穗相关通道并抑制背景干扰。CSNet的热力图整体表现优于YOLOv11，但在高密度区域仍存在部分背景区域被错误激活的情况，说明其MLP-Mixer结构在复杂场景下存在一定局限性。YOLOv11受限于边界框预测机制，热力图在杂草区域出现较多红色响应，说明检测类方法在复杂背景下容易受到干扰，且对重叠麦穗的注意力分散。TransCrowd-GAP的热力图能够在全局范围内捕捉麦穗分布，但在边界及局部细节上注意力不足，导致部分小麦穗响应偏弱。

综上，L-CSNet 在密集场景下的注意力分布最为集中和准确，能够有效区分麦穗与背景，表现优于其他对比模型。

2.5　EMDC 模块多尺度特征可视化

为验证 EMDC 模块中不同膨胀率的分支在特征提取中的作用，可视化了 3 个分支的特征响应，结果如图 4 所示。

a. 原始图像　　b. e=1　　c. e=3　　d. e=5　　e. 完整响应图

图 4　EMDC 模块不同膨胀率分支的特征响应图

从图 4 可知：当 $e=1$ 时，分支(感受野为 5×5)的特征响应集中于麦穗边缘，亮度较高，能够捕捉小尺寸麦穗的纹理细节；$e=3$ 时，分支(感受野为 13×13)的响应集中于单个麦穗区域，适配中等尺寸麦穗的特征提取；$e=5$ 时，分支(感受野为 21×21)的响应覆盖密集麦穗全局，能够避免漏检重叠麦穗。三者融合后，EMDC 模块实现了细节、局部、全局的多尺度特征互补，这是 L-CSNet 在密集场景下性能优越的核心原因。

3　结论

本文针对农业场景下小麦穗计数任务标注成本高、现有模型计算复杂度高难以实时部署的挑战，提出了一种轻量级计数监督网络 L-CSNet。该网络模型摒弃了对边界框或密度图等位置标注的依赖，仅需图像级计数标签即可完成训练，显著降低了数据标注成本。

高效多尺度膨胀卷积(EMDC)模块是本文的核心创新。该模块通过并行膨胀卷积与通道注意力机制，实现了在较低的参数量下对多尺度小麦穗特征的高效捕获。在公开数据集 GWHD 上的大量实验表明，L-CSNet 在计数精度、模型轻量化和推理速度方面取得了平衡，在 GWHD 数据集上，模型的综合性能均优于现有的位置监督和计数监督方法，模型参数量大幅压缩至 9.96×10^{6}，在单 GPU 上实现了 120 FPS 的实时推理速度。

但本论文工作仍然存在可以改进的地方，未来的工作将集中于以下几个方向：一是探索模型在不同作物(如水稻、高粱)计数任务上的泛化能力；二是研究如何将 L-CSNet 与无人机等移动平台深度集成，实现真正端到端的实时田间作物表型分析；三是借鉴深度学习与类脑计算[34]的最新进展，提升模型的能效与可扩展性，探究更前沿的农业自动化方法。

参考文献：

[1] ASSENG S, GUARIN J R, RAMAN M, et al. Wheat Yield Potential in Controlled-Environment Vertical Farms [J]. Proceedings of the National Academy of Sciences of the United States of America, 2020, 117(32): 19131-19135.

[2] TADESSE W, SANCHEZ-GARCIA M, ASSEFA S G, et al. Genetic Gains in Wheat Breeding and Its Role in Feeding the World [J]. Crop Breeding, Genetics and Genomics, 2019, 1(1): e190005.

[3] WEN H X, LI C D, WANG X P, et al. Software and Hardware Synergy for Accelerated Plant Disease Identification [J]. Applied Soft Computing, 2025, 174: 112926.

[4] DONG X Y, ZHAO K J, WANG Q, et al. PlantPAD: A Platform for Large-Scale Image Phenomics Analysis of Disease in Plant Science [J]. Nucleic Acids Research, 2024, 52(D1): D1556-D1568.

[5] PASK A, PIETRAGALLA J, MULLAN D, et al. Physiological Breeding II: A Field Guide to Wheat Phenotyping [M]. 景蕊莲, 译. 北京: 科学出版社, 2017.

[6] COINTAULT F, GUERIN D, GUILLEMIN J P, et al. In-Field Triticum Aestivum Ear Counting Using Colour-Texture Image Analysis [J]. New Zealand Journal of Crop and Horticultural Science, 2008, 36(2): 117-130.

[7] ALHARBI N, ZHOU J, WANG W. Automatic Counting of Wheat Spikes from Wheat Growth Images [C] //7th International Conference on Pattern Recognition Applications and Methods. Faro: SciTePress-Science and Technology Publications, 2018: 346-355.

[8] LI L, HASSAN M A, YANG S R, et al. Development of Image-Based Wheat Spike Counter through a Faster R-CNN Algorithm and Application for Genetic Studies [J]. The Crop Journal, 2022, 10(5): 1303-1311.

[9] LI R F, SUN X H, YANG K, et al. A Lightweight Wheat Ear Counting Model in UAV Images Based on Improved YOLOv8 [J]. Frontiers in Plant Science, 2025, 16: 1536017.

[10] LI X X, ZHANG Z H, WANG J Y, et al. Research on Wheat Spike Phenotype Extraction Based on YOLOv11 and Image Processing [J]. Agriculture, 2025, 15(21): 2295.

[11] XIONG H P, CAO Z G, LU H, et al. TasselNetv2: In-Field Counting of Wheat Spikes with Context-Augmented Local Regression Networks [J]. Plant Methods, 2019, 15(1): 150.

[12] KHAKI S, SAFAEI N, PHAM H, et al. WheatNet: A Lightweight Convolutional Neural Network for High-Throughput Image-Based Wheat Head Detection and Counting [J]. Neurocomputing, 2022, 489: 78-89.

[13] WU W, ZHONG X C, LEI C K, et al. Sampling Survey Method of Wheat Ear Number Based on UAV Images and Density Map Regression Algorithm [J]. Remote Sensing, 2023, 15(5): 1280.

[14] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation Applied to Handwritten Zip Code Recognition [J]. Neural Computation, 1989, 1(4): 541-551.

[15] YANG Y F, LI G R, WU Z, et al. Weakly-Supervised Crowd Counting Learns from Sorting rather than Locations [C] //Computer Vision-ECCV 2020. Cham: Springer, 2020: 1-17.

[16] LIANG D K, CHEN X W, XU W, et al. TransCrowd: Weakly-Supervised Crowd Counting with Transformers [J]. Science China Information Sciences, 2022, 65(6): 160104.

[17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is All You Need [J]. Advances in Neural Information Processing Systems, 2017, 30: 6000-6010.

[18] LI Y X, WU X C, WANG Q, et al. CSNet: A Count-Supervised Network via Multiscale MLP-Mixer for Wheat Ear Counting [J]. Plant Phenomics, 2024, 6: 0236.

[19] YU F, KOLTUN V. Multi-scale Context Aggregation by Dilated Convolutions [EB/OL]. (2015-11-23) [2025-12-26]. https://www.semanticscholar.org/venue?name=International%20Conference%20on%20Learning%20Representations.

[20] DAVID E, MADEC S, SADEGHI-TEHRAN P, et al. Global Wheat Head Detection (GWHD) Dataset: A Large and Diverse Dataset of High-Resolution RGB-Labelled Images to Develop and Benchmark Wheat Head Detection Methods [J]. Plant Phenomics, 2020, 2020: 3521852.

[21] DAVID E, SEROUART M, SMITH D, et al. Global Wheat Head Detection 2021: An Improved Dataset for Benchmarking Wheat Head Detection Methods [J]. Plant Phenomics, 2021, 2021: 9846158.

[22] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-scale Image Recognition [EB/OL]. (2014-09-04) [2025-12-26]. https://www.semanticscholar.org/paper/Very-Deep-Convolutional-Networks-for-Large-Scale-Simonyan-Zisserman/eb42cf88027de515750f230b23b1a057dc782108?p2df.

[23] DONG X Y, WANG Q, HUANG Q D, et al. PDDD-PreTrain: A Series of Commonly Used Pre-Trained Models

Support Image-Based Plant Disease Diagnosis [J]. Plant Phenomics, 2023, 5: 0054.

[24] HU J, SHEN L, SUN G. Squeeze-and-Excitation Networks [C] //2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Pres, 2018: 7132-7141.

[25] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector [C] //Computer Vision-ECCV 2016. Cham: Springer, 2016: 21-37.

[26] HE L Q, TODOROVIC S. DESTR: Object Detection with Split Transformer [C] //2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Pres, 2022: 9367-9376.

[27] ZHANG Y Y, ZHOU D S, CHEN S Q, et al. Single-Image Crowd Counting via Multi-Column Convolutional Neural Network [C] //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Pres, 2016: 589-597.

[28] LI Y H, ZHANG X F, CHEN D M. CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes [C] //2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Pres, 2018: 1091-1100.

[29] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition [C] //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Pres, 2016: 770-778.

[30] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3 [C] //2019 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE Pres, 2019: 1314-1324.

[31] MAHASIN M, DEWI I A. Comparison of CSPDarkNet53, CSPResNeXt-50, and EfficientNet-B0 Backbones on YOLO V4 as Object Detector [J]. International Journal of Engineering, Science and Information Technology, 2022, 2(3): 64-72.

[32] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows [C] //2021 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE Pres, 2022: 9992-10002.

[33] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization [J]. International Journal of Computer Vision, 2020, 128(2): 336-359.

[34] SHAN H X, WEI C Y, RAMOS N, et al. Neuromorphic Computing in the Era of Large Models [J]. Artificial Intelligence Science and Engineering, 2025, 1(1): 17-30.

责任编辑　张枸