

动态局部和边界感知的轻量级单域泛化人群计数方法研究

周丽芳^{1),2),3)} 张旭波^{1),2)}

¹⁾(重庆邮电大学计算机科学与技术学院(示范性软件学院) 重庆 400065)

²⁾(重庆邮电大学重庆市图像认知重点实验室 重庆 400065)

³⁾(重庆邮电大学大数据智能计算重点实验室 重庆 400065)

摘要 人群计数任务在单域泛化(Single Domain Generalization, SDG)领域长期面临模型效率与泛化性能难以平衡的挑战。现有的人群计数SDG方法通常依赖复杂的网络架构,这不仅增加了计算负担,也限制了模型在实际场景中的部署能力。针对这一关键问题,本文提出了一种创新的动态局部和边界感知轻量级单域泛化人群计数模型——LB-CountX,有效解决了效率与性能的协同优化问题。LB-CountX从四个维度进行了创新性设计。首先,设计了一个轻量级的人群计数网络架构CountX,降低了模型参数量。其次,引入交叉注意力机制(Criss-cross attention, CCA),在保证计算效率的同时有效提升了特征表达能力,使轻量化模型的训练过程更加稳定。然后,针对传统全局损失函数难以精确刻画局部区域人群分布的问题,创新性地设计了动态局部感知损失函数,通过动态加权机制区分背景与人群区域。最后,针对密集场景中人群边界检测模糊的难题,设计了边界感知损失函数,提升了边界检测的准确性。实验验证方面,该在SHA、SHB、UCF-QNRF和JHU-CROWD等主流人群计数数据集上进行了评估。结果表明,LB-CountX在SHA→SHB跨域测试中实现了参数量比最新单域泛化方法参数量减少了60.3%和平均绝对误差(Mean Absolute Error, MAE)降低了8.8%,在多数测试场景中取得了最优的性能表现,实现了模型效率与泛化性能的平衡。该的工作进展为单域泛化人群计数模型在实际应用中的部署提供了重要技术支持。

关键词 人群计数;轻量化模型;单域泛化;动态局部感知;边界感知

中图分类号 TP391

DOI号 10.11897/SP.J.1016.2026.00466

Lightweight Single-Domain Generalization Crowd Counting with Dynamic Local and Boundary Perception

ZHOU Li-Fang^{1),2),3)} ZHANG Xu-Bo^{1),2)}

¹⁾(School of Computer Science and Technology, Chongqing University of Posts and Telecommunications (National Exemplary Software School), Chongqing 400065)

²⁾(Chongqing Key Laboratory of Image cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065)

³⁾(Key Laboratory of Big Data Intelligent Computing, Chongqing University of Posts and Telecommunications, Chongqing 400065)

Abstract The task of crowd counting in the field of Single Domain Generalization (SDG) has still faced the challenge of balancing model efficiency and generalization performance. It is because the existing SDG methods for crowd counting typically rely on complex network architectures that increase computational burden and limit real-world deployment capability. To tackle this issue,

收稿日期:2025-05-31;在线发布日期:2025-11-06。本课题得到国家自然科学基金(Nos. 62276039、62331008)、贵州大学省部共建公共大数据国家重点实验室(No. KY[2022]376)、重庆市科卫联合医学科研项目(No. 2024ZDXM022)、基于人脸图像的金融反欺诈防伪方法研究(No. DFIK2025D05)的资助。周丽芳(通信作者),博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为模式识别、医学图像处理。E-mail: zhoulf@cqupt.edu.cn。张旭波,硕士研究生,主要研究领域为人群计数、计算机视觉。

this paper proposes an innovative Local and Boundary-aware Lightweight Single Domain Generalization model for crowd counting(LB-CountX), which effectively resolves the co-optimization of efficiency and performance. Firstly, A lightweight network architecture CountX is designed to reduce parameter significantly. Secondly, a Criss-cross Attention (CCA) mechanism is proposed to enhance feature while maintaining low computational and stabilizing training. Thirdly, a novel local-aware loss function employing dynamic weighting is designed to distinguish local crowd from background, which always fails by traditional global loss functions. Finally, a boundary-aware loss function is introduced that markedly improves boundary detection accuracy in dense scenes with blurred crowd boundary. Extensive evaluations on SHA, SHB, UCF-QNRF, and JHU-CROWD have been performed. The experimental results show that LB-CountX, in the context of cross-domain testing from SHA to SHB, achieves a 60.3% reduction in the number of parameters compared to the latest single-domain generalized methods. It also reduces the mean absolute error (MAE) by 8.8%, balancing between efficiency and generalization can be obtained, which is essential for real-world deployment of crowd counting model.

Keywords crowd counting; lightweight model; single domain generalization; dynamic local perception; boundary perception

1 引言

人群计数技术作为计算机视觉领域的重要研究对象,已广泛应用于公共安全、视频监控、集会管理、交通监控、场景理解等诸多领域。然而现有的研究普遍忽视了一个关键问题,在实际应用场景中,算法的实时性要求与模型复杂度之间的矛盾日益突出。尽管轻量化模型设计能够有效提升计算效率,但简单的模型压缩往往会导致性能的显著衰减^[1,2],这一问题在动态复杂环境下表现得尤为明显。

当前大多数人群计数研究方法存在一个普遍局限,通常假设训练集与测试集服从相同数据分布。但在实际应用中,由于场景差异、天气变化、光照条件、尺度变化、背景噪声等^[3-5]环境因素影响,测试数据(目标域)与训练数据(源域)往往存在显著分布差异,这会导致模型性能显著退化。尽管有研究尝试通过合成数据集来增加数据的多样性^[6],但这种方法难以完全模拟真实场景的复杂性。

针对性能退化问题,学界提出了域适应(Domain Adaptation, DA)和域泛化(Domain Generalization, DG)两类方法。Cai等人^[7]提出的域自适应方法通过利用少量目标域无标注/弱标注数据缩小域间差距,典型代表性方法^[8,9]对目标域数据具有依赖性,导致模型在实际应用部署时可能面临数据隐私等法律合规性挑战。

相较之下,不依赖目标域数据的DG方法展现

出更好的场景适应性。比如Qin等人^[10]提出的行为识别、Kim等人^[11]提出的LiDAR语义分割。其中最具挑战性的是单域泛化SDG场景,现有方法多数采用对抗域增强技术(如自适应归一化^[11])来提升模型泛化能力。这些技术的突破为轻量化人群计数模型跨域部署提供了新的研究思路和技术路线。

在单域泛化人群计数领域,尽管现有SDG方法在模型泛化性能方面取得了一定的进展,如Peng等人^[12]提出的MPCount(基于记忆库的特征重构)和Du等人^[13]提出的DCCUS(子域划分与正交约束)方法,但这些方法普遍依赖域计算复杂度较高的VGG模型^[14]作为主干网络,导致在资源受限场景下实际应用受到严重制约。针对这一关键性问题,本研究创新性地提出了一种融合动态局部和边界感知机制的轻量级单域泛化人群计数模型,通过系统性网络架构设计和损失函数优化,实现了模型效率与性能协同优化。本文的主要创新贡献可归纳为以下四个方面。

(1)本文设计了轻量级单域泛化人群统计模型LB-CountX,该框架引入了交叉注意力机制CCA^[15]和输出多尺度特征融合策略,有效缓解了轻量化模型性能骤降问题。实验结果表明,在SHA→SHB跨域测试中实现参数量减少了60.3%和MAE指标降低了8.8%的双优表现。

(2)为进一步提升轻量化网络模型CountX泛化性能,本文设计了动态局部感知损失,不仅仅提升了模型泛化性,同时缓解了全局感知损失对局部区域

感知不足导致的背景误识别问题。该损失通过建立区域敏感的动态加权机制,提升了模型对人群区域背景的判别能力,为准确的人群计数奠定了基础。

(3)虽然动态局部感知损失提升了模型泛化性,并加强了对局部区域的关注,但是已有的人群计数模型对密集的人群检测还存在边界检测模糊问题,导致最终的人群检测准确性下降和模型泛化性降低。为此,本文设计边界感知损失进一步提升模型对人群边界的识别,使得人群边界检测更加清晰和准确,同时模型泛化性也得到一定提升。

(4)本文在多个常用人群计数数据集上进行了验证,在 SHTech^[16]、UCF-QNRF^[17]和 JHU-CROWD++^[18]上的实验结果充分表明本文提出的方法在保持模型轻量化同时,性能在大部分数据集上超越现有最优方法(DCCUS/MPCount),进一步证明了本文方法在模型效率和泛化性能方面达到了一个较好的平衡。

2 相关工作

2.1 轻量化人群统计模型

尽管已有大量工作致力于提升密度图预测精度,但轻量级模型设计这一关键问题尚未获得足够重视。值得注意的是,模型轻量化对于实现人群计数技术的实际部署具有决定性意义,特别是在资源受限的边缘计算场景中。然而,现有研究方法^[1, 2, 19]在轻量化方向上的探索仍显不足,如Huang等人^[19]提出的方法主要从知识蒸馏的角度提升轻量级模型的精度,但尚未探索更具挑战的单域泛化方向。单域泛化能力对于实现跨场景的鲁棒人群统计至关重要。在实际应用中,人群分布往往呈现出显著的空间异质性和场景依赖性,要求模型必须具备强大的域泛化能力。基于上述背景,本研究聚焦于单域泛化轻量级人群计数模型,致力于解决模型复杂度、预测精度和泛化能力之间的平衡问题。通过系统性的方法创新,本文期望为人群计数的实际应用提供更高效的解决方案。

2.2 人群计数的域自适应

域自适应技术在提升人群计数模型泛化性能方面展现出显著优势,He等人^[20]通过利用少量无标注或弱标注的目标域数据来有效缓解域偏移问题。其中Wang等人^[21]提出的方法不依赖目标域数据,仅在合成数据上训练,Wu等人^[22]提出的动量模板(MoT)通过离线训练时的动量更新方式学习,并在

在线跨数据集评估时对每张测试图像进行动态更新,在零样本跨域人群计数任务中取得了突破性进展。Liu等人^[23]提出的无监督人群计数方法引入了一种迭代自监督学习方案,通过目标域中的回归-检测双知识迁移,证明了该方法在迁移学习场景中的有效性。

尽管域自适应技术能显著提升模型在未知场景下的性能,但其对目标域数据的依赖性依然存在,这在隐私敏感或数据获取受限的应用场景中构成了主要瓶颈。这一局限性直接催生了更具挑战性的单域泛化研究课题,即在完全无法获取目标域数据的条件下实现模型的跨域泛化能力。

2.3 人群计数的单域泛化

在计算机视觉领域,因其完全无法利用目标域数据进行模型的优化,域泛化相比域域自适应更具挑战性。本研究聚焦于单域泛化SDG这一前沿方向,即在单一源域数据集上训练模型,使模型具备跨未知场景的泛化能力。SDG技术已在多个视觉任务中展现出重要价值,比如分类领域Zhang等人^[24]提出的单域泛化方法,语义分割领域方法^[11, 25],这些研究有效提升了模型的泛化能力。从归一化白化变换和通用视角来看,单域泛化方法被应用于分类、目标检测和分割等领域^[26-28]。在人群计数领域,已有方法^[12, 13]基于单域泛化思想提升了模型在未知场景下的性能表现。

针对人群计数任务,现有SDG方法主要呈现两种技术路线。DCCUS方法^[13]采用元学习策略,将源域动态划分为模拟元训练和元测试的子域,通过双记忆模块记录域不变和域特定信息。该方法虽创新性地设计了回归任务的记忆机制,但MPCount方法^[12]指出其存在两个关键局限,一是未充分考虑标签模糊问题,二是当源域服从窄分布时基于K-Means聚类的子域划分效果受限。相较之下,MPCount采用块级分类策略解决标签模糊性,并引入单一记忆库实现无需子域划分的泛化建模,这在窄分布场景下表现出更强的鲁棒性。

鉴于现有SDG方法均采用计算密集的VGG网络模型^[14]作为骨干网络,严重制约了实际部署可行性。本文首次系统性地探索了轻量级SDG人群计数模型的设计方法。本文通过引入交叉注意力机制稳定模型训练和提升泛化性,并创新性地设计了动态局部感知损失和边界感知损失,在SHA→SHB跨域测试中实现参数量减少了60.3%的同时MAE指标降低了8.8%。实验证明,该方法在SHTech^[16]、

UCF-QNRF^[17]等基准数据集上均达到最优的泛化性能,为资源受限场景下的实时人群分析提供了切实可行的技术方案。

3 方法

本文提出LB-CountX方法,通过设计轻量级网络模型并结合动态局部和边界感知损失,有效缓解了单域泛化人群计数中模型参数量过大的问题,通过融合动态局部与边界感知损失提升了模型性能,在模型效率和泛化性之间实现了较好的平衡,在大部分目标

数据集上超越了人群计数域泛化最新研究结果^[12, 13]。

如图1所示,LB-CountX架构包含三个核心步骤。首先,设计轻量级网络模型CountX,结合交叉注意力机制CCA^[15]实现模型训练的稳定性。其次,本研究是在DCCUS^[13]基础上进行改进,保留原有的域不变分支(Domain Invariant-Branch, DI-Branch)和特定域分支(Domain Specific-Branch, DS-Branch),这两个分支分别用于提取域不变信息和特定域信息。最后,本文创新性地设计了动态局部感知损失与边界感知损失,分别用于解决局部人群区域分布和人群边界识别模糊问题。

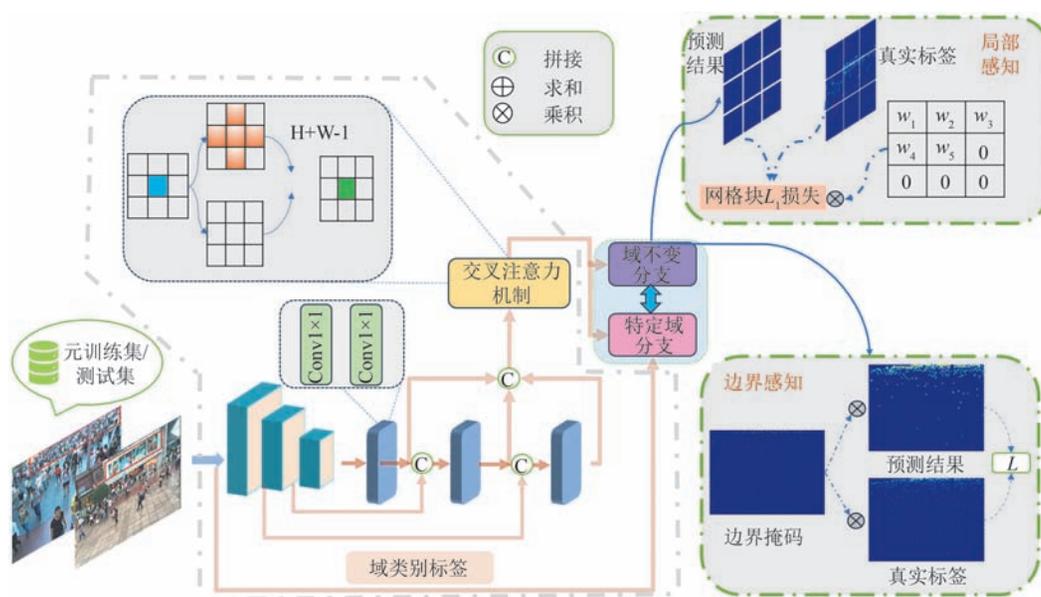


图1 整体模型架构

3.1 轻量化网络模型与交叉注意力融合

最新单域泛化人群计数研究存在一定局限性,现有代表性方法MPCount^[12]和DCCUS^[13]均依赖计算密集的VGG模型^[14]作为骨干网络,导致模型参数量庞大且推理效率低下,严重制约实际应用可行性。本研究通过设计轻量化模型,引入ConvNeXt-tiny^[29]模块作为构建高效骨干网络基础,并引入交叉注意力机制CCA^[15]以提升训练稳定性和模型泛化性。实验结果表明,该方案在保持模型精度的同时降低了计算复杂度,实现了模型效率和泛化性之间较好的平衡,为轻量级单域泛化人群计数研究提供了有效方案。为了进一步提升模型的泛化性,本文设计了动态局部感知损失和边界感知损失联合学习加强对局部区域的关注和人群边界的检测。

3.2 联合动态局部和边界感知损失

3.2.1 动态局部感知损失

在人群计数研究中,大部分方法采用均方误差作为全局损失,其数学表达式如下:

$$\mathcal{L}_{global} = \frac{1}{2N_{train}} \|den_{pred} - den_{gt}\|^2 \quad (1)$$

其中, N_{train} 代表训练图像总数, den_{pred} 表示模型输出的密度图, den_{gt} 对应真实密度图。

虽然现有全局损失难以认识到局部背景区域和人群区域。为此,本文提出动态局部感知损失,使模型能够同时关注全局和局部人群区域特征。相关研究方法^[30]虽已采用类似方法增强模型对局部人群信息的关注,但未能解决局部区域无人群分布(纯背景)时的训练难题。针对这一缺陷,本方法引入动态加权机制,通过动态权重指导模型学习当前局部区域人群分布,从而提升模型检测精度。如

图2(a-d)所示,引入动态局部感知损失后,人群分布的局部区域定位误差显著降低,验证了该损失函数的有效性。

动态局部感知损失的具体实现分为三个步骤。

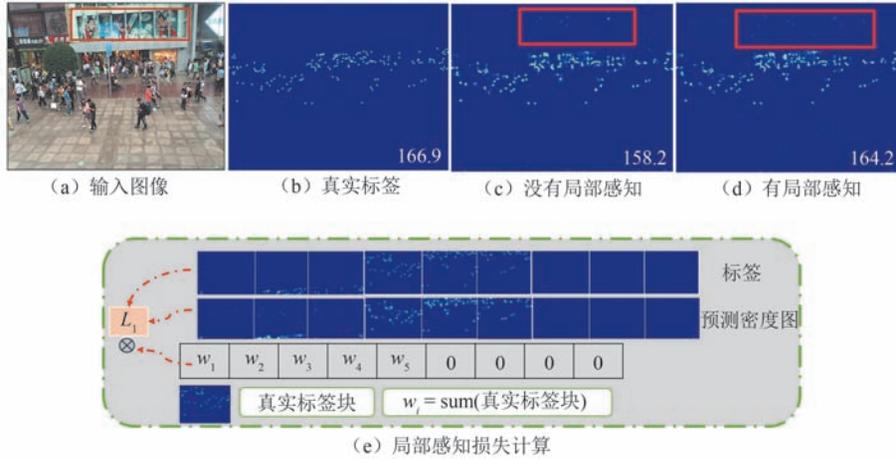


图2 动态局部感知实现

$$\mathcal{L}_{\text{local}} = \frac{1}{K \times K} \sum_{k=1}^{K \times K} \omega_k \cdot \left(\frac{1}{|R_k|} \sum_{i \in R_k} |D_{\text{pred}}(i) - D_{\text{gt}}(i)| \right) \quad (2)$$

其中, R_k 表示第 k 个局部区域, $|R_k|$ 代表该区域的像素数量, ω_k 表示当前真实密度图划分为 $H/K \times W/K$ 个局部区域之后每一个区域对应的人群计数值, 作为局部区域关注的动态权重, D_{pred} 为预测密度图, D_{gt} 对应真实密度图。动态局部感知损失提升了模型关注局部区域的重要性, 同时也提升了模型的泛化性, 但是模型在检测密集人群边界时存在模糊问题, 为了提升模型对人群边界的关注, 本文设计了边界感知损失。

3.2.2 边界感知损失

在人群密集场景中, 个体间的边界因严重遮挡而变得模糊, 导致密度图边界区域的预测不准确。传统全局损失(如均方根误差 RMSE)和动态局部感知损失对所有像素一视同仁, 未能突出边界区域的重要性。为解决边界检测模糊问题, 本文提出边界感知损失函数, 以增强模型处理人群边界分布的能力。在本研究中, 本文使用 Canny 边缘检测算法^[31]获取真实密度图的边缘信息。本文计算损失时, 将处理后的边缘密度图分别与真实密度图和模型输出的密度图相乘。这种方法能有效抑制非边缘信息, 更好地区分人群边界分布信息。

Canny 边缘检测算法实现主要包含五个步骤。第一步对图像进行高斯平滑处理以降低噪声; 第二

如图2(e)所示, 首先将真实和预测密度图分别划分为 $H/K \times W/K$ 个局部区域, 随后计算每个区域的 L_1 误差。最后基于当前局部区域的人群数量动态分配权重。其数学定义为

步计算水平和垂直方向的梯度; 第三步在梯度方向上进行非极大值抑制, 仅保留局部最大值; 第四步使用高阈值和低阈值对边缘进行分类; 第五步将弱边缘连接到强边缘, 若弱边缘与强边缘相邻, 则保留该弱边缘。

由于本文的真实密度图是在模型训练之前使用高斯函数和人群坐标点生成, 所有值都是在 0-1 之间的实数, 在使用 Canny 检测真实密度图中人群边缘之前, 会将所有 0-1 之间的实数值缩放至 0-255。本文对 SHTechA^[16]训练集密度图像素值统计, 如图4所示可以看出缩放之后密度图像素值大部分都集中在更小值区域, 并且计算其像素值的均值为 3.90。因此, Canny 边缘检测算法第四步高低阈值选择设置低阈值为 0, 高阈值为 20。

基于 Canny 边缘检测算法, 本文最终定义的边界感知损失如下:

$$\mathcal{L}_{\text{boundary}} = \frac{1}{S_{\text{boundary}}} \sum_{i,j} M_{\text{boundary}}(i,j) \cdot |D_{\text{pred}}(i,j) - D_{\text{gt}}(i,j)|^2 \quad (3)$$

其中, $S_{\text{boundary}} = \sum_{i,j} M_{\text{boundary}}(i,j)$, $M_{\text{boundary}}(i,j)$ 表示通过 Canny 算法获得的真实边缘密度图, D_{pred} 表示模型输出的密度图, D_{gt} 指对应的真实密度图。

如图3所示对比结果来看, 使用本文设计的边界感知损失提升了模型对人群边界的关注。关于动态局部感知和边界感知损失的详细步骤, 实现细节请参考 3.3 节。

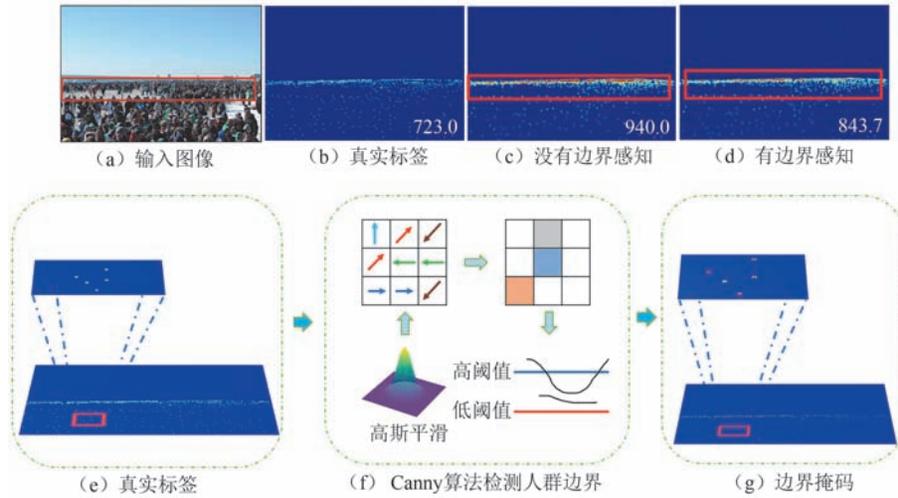


图3 边界感知实现

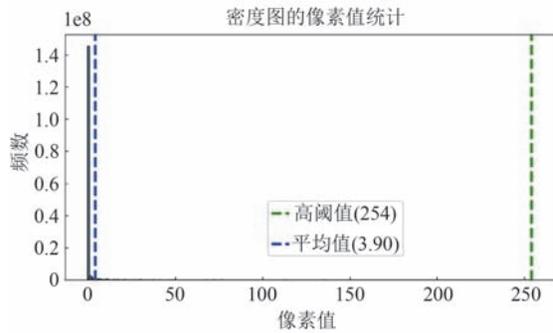


图4 密度图像素值的统计

模型整体训练过程使用的损失函数定义如下：

$$\mathcal{L}_{total} = \mathcal{L}_{global} + \mathcal{L}_{local} + \mathcal{L}_{boundary} + \mathcal{L}_{rec} + \mathcal{L}_{orth} \quad (4)$$

鉴于本文是基于DCCUS^[13]改进而来,需要保留原始的重建损失 \mathcal{L}_{rec} 和正交损失 \mathcal{L}_{orth} 。因此,各损失项的默认权重设置为 $\lambda_{global} = 0.1$, $\lambda_{local} = 1.0$, $\lambda_{bound} = 1.0$, $\lambda_{rec} = 1.0$, $\lambda_{orth} = 1.0$ 。

3.3 动态局部和边界感知损失具体算法实现

算法1: 动态局部感知算法

输入:定义划分网格大小 $grid_h = h/grid_{size}$ 和 $grid_w = w/grid_{size}$ 以及记录动态局部损失变量 \mathcal{L}_{local}

FOR i IN $grid_{size}$ DO:

FOR j IN $grid_{size}$ DO:

step1: $gt_{patch} \leftarrow crop(gt_{map}, i, j)$ #根据网格大小对真实密度图裁剪

step2: $pred_{patch} \leftarrow crop(pred_{map}, i, j)$ #根据网格大小对预测密度图裁剪

step3: $W_{count} \leftarrow sum(gt_{patch})$ #根据网格大小求解当前区域人群数

step4: $\mathcal{L}_{local} += W_{count} \times \mathcal{L}_1(gt_{patch}, pred_{patch})$ #局部损失求和

END FOR

END FOR

输出: \mathcal{L}_{local}

算法2: 边界感知算法

输入:设置Canny检测的低阈值 $low_{threshold}$,高阈值 $high_{threshold}$,记录边界损失变量 \mathcal{L}_{bound} ,真实标签 gt_{map} 以及预测密度图 $pred_{map}$

step1: $gt_{map}^{Norm} \leftarrow normalize(gt_{map})$ #对真实密度图标签进行归一化

step2: $M_{boundary} \leftarrow Canny(gt_{map}^{Norm}, low_{threshold}, high_{threshold})$ #使用canny检测算法检测人群边界

step3: $M_{boundary} \leftarrow norm(M_{boundary})$ #对边界掩码图进行归一化

step4: $gt_{map} \leftarrow gt_{map} \times M_{boundary}$ #计算真实密度图的人群边界

step5: $pred_{map} \leftarrow pred_{map} \times M_{boundary}$ #计算预测密度图的人群边界

step6: $S_{boundary} \leftarrow sum(M_{boundary})$ #边界求和

step7: $\mathcal{L}_{bound} \leftarrow \mathcal{L}_{bound} / S_{boundary}$ #边界感知损失归一化

输出: \mathcal{L}_{bound}

算法3: 模型整体训练流程

输入:迭代周期 $epoch_{max}$,超参数 α, β, γ ,初始化的基准值 $base_{\alpha}, base_{\beta}, base_{\gamma}$ 以及损失值的权重参数 $\lambda_{global}, \lambda_{local}, \lambda_{bound}, \lambda_{rec}$ 和 λ_{orth}

FOR each epoch up to $epoch_{max}$ DO:

Step: 划分训练集作为多个子域

IF $epoch \geq epoch_{max} \cdot epoch_{ratio}$ THEN:

step1: 更新 $base_{\alpha}$: $base_{\alpha} \leftarrow base_{\alpha} \cdot \alpha$

step2: 更新 $base_{\beta}$: $base_{\beta} \leftarrow base_{\beta} \cdot \beta$

step3: 更新 $base_{\gamma}$: $base_{\gamma} \leftarrow base_{\gamma} \cdot \gamma$

END IF

FOR each iteration up to $total_{iterations}$ DO:

#元训练和元测试

step4: $\mathcal{L}_{global} = \mathcal{L}_{global} \cdot \lambda_{global} \cdot base_{\alpha}$
 step5: $\mathcal{L}_{local} = \mathcal{L}_{local} \cdot \lambda_{local} \cdot base_{\beta}$
 step6: $\mathcal{L}_{bound} = \mathcal{L}_{bound} \cdot \lambda_{bound} \cdot base_{\gamma}$
 step7: $\mathcal{L}_{total} = \mathcal{L}_{rec} + \mathcal{L}_{orth} + \mathcal{L}_{bound} + \mathcal{L}_{local} + \mathcal{L}_{global}$
 step8: 更新模型参数

END FOR

END FOR

4 实 验

4.1 数据集

本研究采用三个具有代表性的基准数据集进行模型验证。SHTech^[16]数据集包含482张(SHTechA)和716张(SHTechB)图像,其中SHTechA部分具有更高的人群密度,显著增加了检测难度;UCF-QNRF^[17]数据集规模更大(1,535张图像),且包含更密集的人群场景。JHU-CROWD++^[18]数据集则采用与MPCount^[12]相同的划分方法。所有数据集均采用标准化的头部标注方式,具体统计特征详见表1。

表1 数据集详细介绍

数据集	数据集大小	平均分辨率 (宽×高)/张	平均人数/张
SHTechA ^[16]	482	586 × 868	501
SHTechB ^[16]	716	768 × 1024	124
UCF-QNRF ^[17]	1535	2013 × 2902	815
JHU-CROWD++ ^[18]	4372	1450 × 900	262

4.2 实验设置

在数据预处理阶段,实施标准数据增强策略,包括随机裁剪至固定尺寸和水平翻转。训练过程采用180个epochs(每个epoch包含100次迭代)的优化方案,配合OneCycle学习率调度策略(最大学习率0.0001,最小学习率0.00001),对于真实密度图的生成,高斯平滑生成密度图的方差 σ 设置为2。其余超参数设置均与DCCUS^[13]保持一致。

实验环境配置,硬件平台采用两张NVIDIA GeForce RTX 3090 GPU(24 GB显存/卡),软件环境基于Python 3.8和PyTorch 1.11.0框架。训练阶段设置batch size为1,启用8个工作线程进行数据加载。针对UCF-QNRF数据集的高分辨率特性,验证阶段采用分块检测策略(1024×1024裁剪尺寸)以避免显存溢出。SHTechA/B, JHU-CROWD++数据集裁剪为320×320,UCF-QNRF

裁剪为512×512。测试阶段SHTechA/B数据集保持原始分辨率,其余数据集均采用1024×1024裁剪大小。

4.3 实验评价指标

本文采用人群计数领域两个通用评估指标,平均绝对误差(Mean Absolute Error, MAE)和均方根误差(Root Mean Squared Error, RMSE)。

这两个评估指标的定义如下:

$$MAE = \frac{1}{N_{test}} \sum_{i=1}^N |C_i - C_i^{GT}| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N_{test}} \sum_{i=1}^N |C_i - C_i^{GT}|^2} \quad (6)$$

其中, N_{test} 表示测试图像数量, G_i 表示第*i*张图像的预测人群数, G_i^{GT} 表示第*i*张图像的真实人群标注数量。

4.4 实验结果分析

需要特别说明的是,其中S和T分别表示源域和目标域数据集(A: SHTechA, B: SHTechB, Q: UCF-QNRF),DA和DG分别代表域适应和域泛化方法。除DCCUS和MPCount的部分结果通过自行训练获得外,其余对比方法数据均直接引用原始文献。

表2和表3结果表明,本文方法在多个目标域任务上取得较大提升:A→B(+1.0 MAE)、A→Q(+1.9 MAE)和B→A(+10.4 MAE)。虽然在B→Q和Q→B任务上性能略有下降(主要归因于模型轻量化导致的域不变特征学习能力轻微减弱),但综合考量模型参数量和推理速度等关键指标,本方法在保持轻量级架构优势的同时,仍实现了实质性的性能提升,充分验证了所提方法有效性。

表4的实验设置严格遵循JHU-CROWD++^[18]数据集的标准划分方案,其中SD(Stadium,体育场)、SR(Street,街道)、SN(Snow,雪)和FH(Fog/Haze,雾/霾)代表四种典型场景。基于MPCount^[12]框架的单域泛化实验结果表明,本方法在SR→SD和FH→SN目标域任务上取得了最好的性能提升,在SN→FH任务上也实现了7.6%的提升。虽然在SD→SR任务上的泛化性能略低于当前最优结果,但本模型在参数量和推理速度方面具有明显优势(详见4.5节分析)。

可视化结果(如图5所示)进一步验证了本方法的有效性,在Q→A、B→A、SR→SD和A→B四个典型场景中,模型对密集人群的检测表现出色。本文通过设计动态局部和边界感知机制,即使在轻量

表2 在三个目标数据集上MAE和RMSE综合实验结对比结果

方法	S → T		A → B		A → Q		B → A		B → Q	
	DA	DG	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
BL ^[32]	×	×	15.9	25.8	166.7	287.6	138.1	228.1	226.4	411.0
DMCount ^[33]	×	×	23.1	34.9	134.4	252.1	143.9	239.6	203.0	386.1
SASNet ^[30]	×	×	21.3	33.2	211.2	418.6	132.4	225.6	273.5	481.3
ChfL ^[34]	×	×	18.7	29.1	122.3	218.0	121.3	200.8	197.1	357.9
MAN ^[35]	×	×	22.1	32.8	138.8	266.3	133.6	255.6	209.4	378.8
RBT ^[23]	✓	×	13.4	29.3	175.0	294.8	112.2	218.2	211.3	381.9
C ² Mot ^[22]	✓	×	12.4	21.1	125.7	218.3	120.7	192.0	198.9	368.0
FGFD ^[36]	✓	×	12.7	23.3	124.1	242.0	123.5	210.7	209.7	384.7
DAOT ^[37]	✓	×	10.9	18.8	113.9	215.6	-	-	-	-
FSIM ^[38]	✓	×	11.1	19.3	105.3	191.1	120.3	202.6	194.9	324.5
IBN ^[27]	×	✓	19.1	30.8	280.2	561.0	125.9	202.3	183.5	317.4
SW ^[28]	×	✓	20.2	30.4	285.5	431.0	126.7	193.8	200.7	333.2
DG-MAN ^[39]	×	✓	23.9	37.7	215.6	399.7	156.2	291.5	263.1	442.2
ISW ^[40]	×	✓	17.3	28.7	129.1	238.2	130.7	225.1	182.4	325.8
DCCUS ^[13]	×	✓	12.6	24.6	119.4	216.6	121.8	203.1	179.1	316.2
MPCCount ^[12]	×	✓	11.4	19.7	115.7	199.8	99.6	182.9	165.6	290.4
LB-CountX	×	✓	10.4	18.2	113.8	202.8	89.2	162.9	177.5	333.3

注：加粗表示最好结果。

表3 两个目标数据集上MAE和RMSE综合对比结果

方法	S → T		Q → A		Q → B	
	DA	DG	MAE	RMSE	MAE	RMSE
DMCount ^[33]	×	×	73.4	135.1	14.3	27.5
SASNet ^[30]	×	×	73.9	13.0	22.1	-
ChfL ^[34]	×	×	68.7	118.5	14.0	27.0
MAN ^[35]	×	×	67.1	122.1	12.5	22.2
FGFD ^[36]	✓	×	70.2	118.4	12.5	20.3
DAOT ^[37]	✓	×	67.0	128.4	11.3	19.6
FSIM ^[38]	✓	×	66.8	111.5	11.0	19.7
IBN ^[27]	×	✓	105.9	174.6	16.0	25.5
SW ^[28]	×	✓	102.4	168.8	19.0	32.9
DG-MAN ^[39]	×	✓	83.4	136.0	22.1	34.7
DCCUS ^[13]	×	✓	67.4	112.8	12.1	20.9
MPCCount ^[12]	×	✓	65.5	110.1	12.3	24.1
LB-CountX	×	✓	63.3	107.3	14.0	25.1

注：加粗表示最好结果。

化架构下，模型仍能保持精确的人群密度检测能力，特别是在个体边界识别方面展现出更好的表现。这些实验结果充分证实了本方法在保持模型效率的同时，有效提升了模型单域泛化场景下的计数精度。

4.5 参数量以及推理时间对比

虽然 Yang 等人^[41]提出的方法作为半监督人群计数领域的最新研究成果，在 A → Q 和 Q → A 目标测试集上报告了实验结果，但并没有报告在其他目标数据集上的泛化性能，因此未将其纳入对比实

验。其中 ISW、SW 和 IBN 均采用与 MPCCount 相同的模型架构(具体实现细节请参考 MPCCount 原文)，因此除 MAE 指标外，这三个变体在其他评估指标上呈现完全一致的实验结果。针对 DG-MAN 方法用于图像分类的单域泛化，并且现有单域泛化^[12-13]没有具体介绍 DG-MAN 模型适应到人群计数具体方案，本文不对 DG-MAN 模型进行太多改动，在网络模型最后加上和 ISW 一样的解码器模块。

如图 6 所示，本研究 LB-CountX 模型在模型复杂度与推理效率之间实现了更优的平衡。虽然本方法单帧推理时间略高于 ISW/SW/IBN，但泛化性能优于这些方法。与当前主流的 DCCUS 和 MPCCount 方法相比，LB-CountX 展现出更好的表现：1) 模型参数量仅为二者的 1/3；2) 推理速度分别提升 0.4 倍和 14.5 倍；3) 在 A → B 跨域测试中取得最优 MAE。这些结果充分验证了本方法在保持模型轻量化特性的同时，有效提升了模型单域泛化性能。

如图 7 所示第一行和第二行是在 A → B 数据集上可视化结果，第三行和第四行是在 B → A 数据集上可视化结果。从可视化结果可以看出，相比于 ISW、SW 和 IBN 三个方法，本方法的检测结果更优。原因在于本文提出的动态局部感知损失和边界感知损失，其中动态局部感知损失缓解了全局感知损失对局部区域感知不足导致的背景误识别问题，边界感知损失提升了模型对人群边界的识别，使得

表4 JHU-CROWD++数据集上域泛化和域自适应对比结果

方法	S → T		SD → SR		SR → SD		SN → FH		FH → SN	
	DA	DG	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
BL ^[32]	×	×	42.1	79.0	262.7	1063.9	48.1	129.5	343.8	770.5
MAN ^[35]	×	×	45.1	79.0	246.1	950.8	38.1	68.0	445.0	979.3
DAOT ^[37]	✓	×	45.3	88.0	278.7	1624.3	42.3	73.0	151.6	273.9
IBN ^[27]	×	✓	92.2	178.0	318.1	1420.4	109.7	267.6	491.8	1110.4
SW ^[28]	×	✓	110.3	202.4	312.6	1072.4	131.5	306.6	381.3	825.0
ISW ^[40]	×	✓	108.1	212.4	385.9	1464.8	151.6	365.7	276.6	439.8
DCCUS ^[13]	×	✓	90.4	194.1	258.1	1005.9	54.5	125.8	399.7	945.0
MPCount ^[12]	×	✓	37.4	70.1	218.6	935.9	31.3	55.0	216.3	421.4
LB-CountX	×	✓	45.8	89.5	179.1	549.1	29.1	45.7	106.2	228.2

注:加粗表示最好结果。

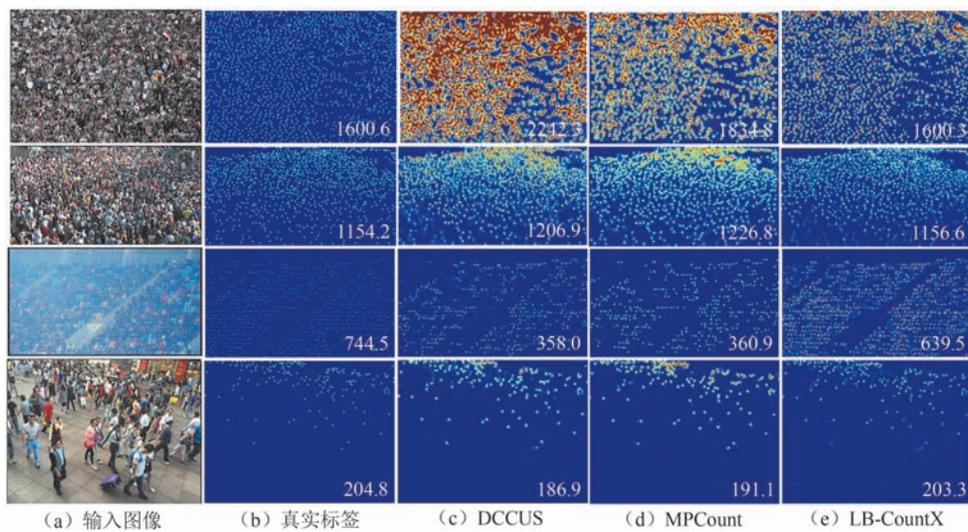


图5 不同目标数据集的可视化结果



图6 不同模型参数量以及推理时间对比

人群边界检测更加清晰和准确,再一次证明了本文方法在参数量更低的情况下,泛化能力更优。

表5系统比较了各单域泛化人群计数模型在多个关键指标上的性能表现,所有实验均在

NVIDIA GeForce RTX 3090 GPU平台上进行标准化测试。首先,本方法在平均绝对误差(MAE)指标上取得了最优性能,优于现有方法。其次,虽然本文模型参数量比DG-MAN参数量大,但管从推理速

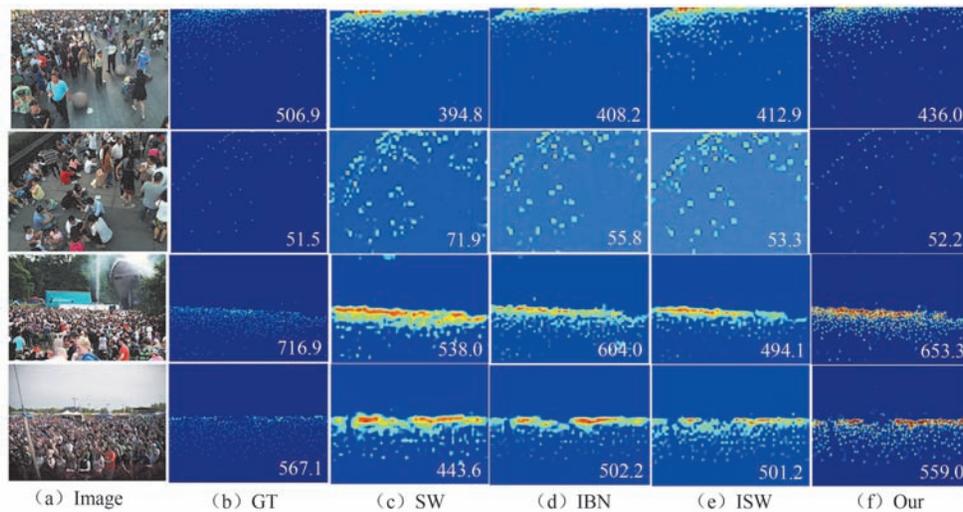


图7 单域泛化模型在不同数据集上可视化对比

表5 单域泛化模型效率及性能对比(橙色表示次优)

方法	MAE	参数量/M	推理时间/ms	乘加运算次数/G	浮点运算量/G
ISW ^[40]	17.3	14.4	36	9.1	18.3
SW ^[28]	20.2	14.4	36	9.1	18.3
IBN ^[27]	19.1	14.4	36	9.1	18.3
DG-MAN ^[39]	23.9	3.945	120	1.5	3.1
DCCUS ^[13]	12.6	32.6	151	121.2	242.3
MPCount ^[12]	11.4	33.0	1670	56.0	112.1
LB-CountX	10.4	13.1	108	10.9	21.9

度还是MAE指标上,效果都优于DG-MAN方法。最后,在推理时间、乘加运算次数和浮点运算量等计算效率指标上,尽管本方法并非各项均为最优,但在推理效率与MAE指标之间实现了最佳平衡。这种平衡使得本方法在保持轻量化优势的同时,仍能确保计数精度,充分验证了所提方法在模型复杂度、计算效率和预测精度三者之间的协同优化能力。

如图8所示第一行和第二行是在A→B数据集上可视化结果,第三行和第四行是在B→A数据集上可视化结果。本文的方法在检测人群边界清晰度以及准确度均优于DG-MAN。更进一步,DG-MAN的梯度手术方法基于强假设——域间梯度冲突可以通过简单的投影操作解决。然而在人群计数任务中,不同域的人群密度分布存在根本性差异,导致梯度方向存在本质性分歧。简单的投影操作会丢失重要的域特异性信息,特别是在人群边界区域,本文如图13所示也说明了不同域人群密度存在明显差异。相比之下,LB-CountX的动态局部和边界感知损失机制能够更好地自适应处理这种复杂情况,

图8可视化结果也充分说明了这一点。本文引入动态局部和边界感知损失函数,能更好地保持跨域场景下的人群检测清晰度和计数准确性。通过平均绝对误差MAE,均方根误差RMSE以及如图8所示可视化结果综合评估可知,所提方法在模型效率和性能两方面达到了动态平衡。

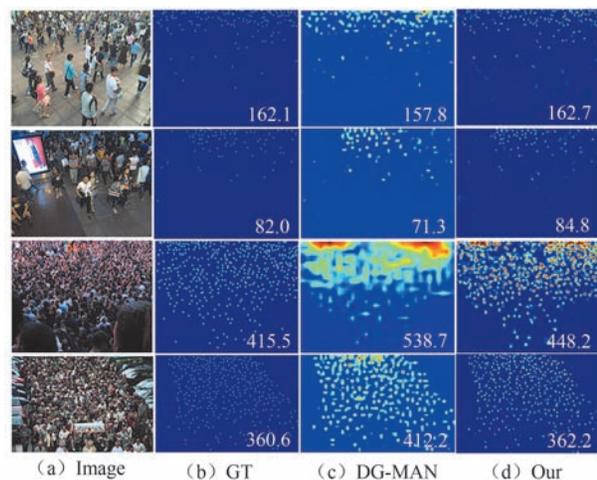


图8 DG-MAN和本文方法(LB-CountX)对比

4.6 消融实验

表6系统呈现了本研究的消融实验结果,通过对比分析验证了各模块和损失设计的有效性。实验数据表明,每个组件都对模型性能提升做出了显著

贡献。特别需要指出的是,本研究采用ConvNeXt-tiny^[32]模块作为基础,取最后三个层级的输出进行多尺度特征融合作为最终输出,以此构建更鲁棒的轻量级主干网络。

表6 LB-CountX消融实验对比

数据集				A→B		A→Q	
基准模型	交叉注意力	动态局部感知	边界感知	MAE	RMSE	MAE	RMSE
✓				11.6	20.2	124.1	219.3
✓	✓			11.0	19.9	121.8	214.1
✓	✓	✓		10.7	19.9	125.8	218.7
✓	✓		✓	10.9	20.5	125.6	223.9
✓	✓	✓	✓	10.4	18.2	113.8	202.8

从表6实验结果可以看出本文的基准模型在SHTechA数据集上泛化性,在没有引入交叉注意力,改进的动态局部和边界感知损失情况下,本文的基准模型和最新方法MPCount^[12]相差0.2个点,这也间接说明了本文设计的轻量化网络模型有效性。

4.6.1 交叉注意力

由于交叉注意力机制CCA^[15]在捕获空间信息方面具有重要作用,本文引入该机制进一步增强了模型在学习过程中专注力和稳定性。表6实验结果表明,加入交叉注意力后,模型在A→B和A→Q任务上的性能确实得到了一定的提升,并且从图9的可视化结果来看,相比于基准模型有了较大的提升。如图10所示,在SHTechA^[16]上训练结果可以看出,使用CCA注意力机制之后,在训练过程中的第4代到第19代和第70代到第75代两个区间模型训练明显更加稳定,没有出现明显抖动现象。使用CCA交叉注意力机制之后,并没有对模型的总参数量和推理速度产生太大的影响,兼顾了模型参数量和模型效率。如图9所示,加入注意力之后检测结果更准确。因此,本文决定引入CCA交叉注意力机制。尽管基准模型的泛化性和训练稳定性得到了提升,但是模型对于局部区域关注较少,然而局部区域对于模型最终性能具有较大的影响,为了进一步提升轻量化网络模型CountX泛化性以及局部区域关注度,设计了动态局部感知损失,不增加模型参数量的同时提升模型性能。

4.6.2 动态局部感知损失作用

表6的消融实验结果表明,动态局部感知损失的引入使模型在A→B任务上的MAE指标提升0.3点,验证了动态局部区域对于轻量化模型CountX泛化提升具有重要作用。然而,该损失在

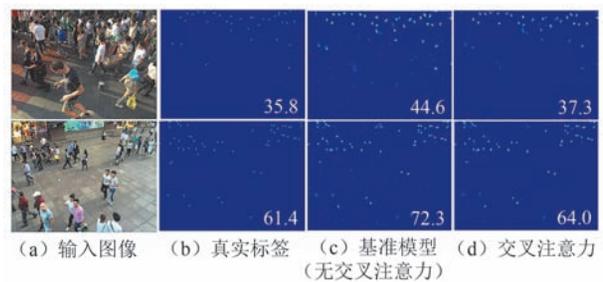


图9 交叉注意力的影响

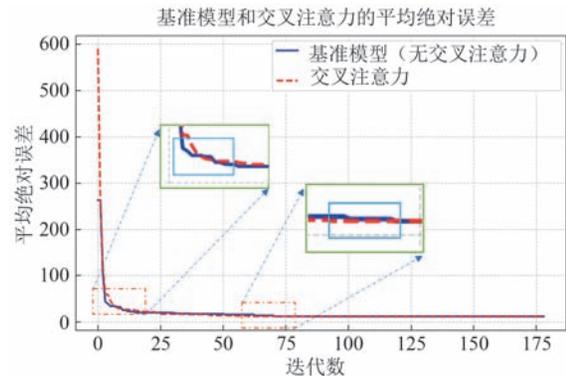


图10 基准模型和加入交叉注意力的MAE对比结果

A→Q任务(UCF-QNRF^[17]数据集)上性能轻微下降,分析认为这种下降源于局部感知损失与原始重建损失、正交损失间的优化目标冲突,在更高分辨率、更密集的场景中表现稍微明显一点。值得注意的是,可视化分析如图11所示清晰显示,该损失有效抑制了背景干扰,在多数场景中保持了稳定的泛化性能。这种精度提升与特定场景性能波动的权衡,证实了动态局部感知机制的设计合理性。动态局部感知损失提升了轻量化模型CountX对局部区域的关注,但是模型在检测人群边界依然存在模糊问题,为进一步提升模型泛化性和对人群边界的检测,设计了边界感知损失并通过实验证明了有效性。

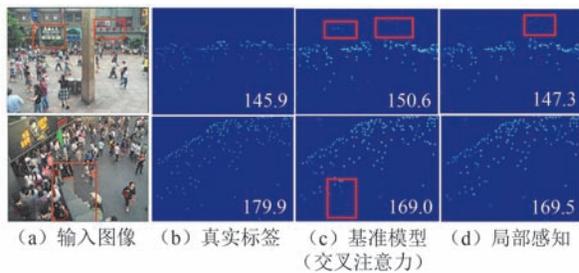


图11 动态局部感知损失的影响

4.6.3 边界感知损失作用

表6表明边界感知损失和局部感知损失的消融实验结果相似。相比“基准模型+交叉注意力”的配置,该损失在A→B数据集上实现了泛化性能的提升。此外,根据消融实验的可视化结果,引入边界感知损失能使检测人群边界结果更加清晰准确,尤其在密集人群场景中表现突出。如图12所示,可视化结果表明,边界感知损失显著增强了密集场景下的人群边界处理能力,从而进一步提升了模型在密集人群中的检测性能。

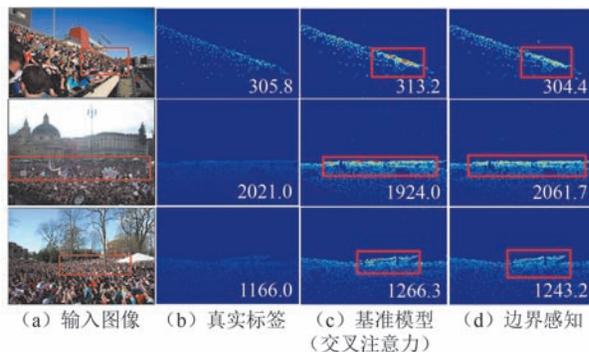


图12 边界感知损失的影响

4.6.4 联合动态局部和边界感知损失

表6实验表明,从设计轻量化网络模型CountX和引入交叉注意力CCA^[15],到设计动态局部感知损失和边界感知损失,每一步的消融实验结果都证明了提出方法的有效性。表6最后一行综合实验结果,表明了模型在A→B和A→Q任务上均取得了最佳性能表现。实验效果证明了CountX模型在效率和泛化性方面达到了一个很好的平衡。

4.7 局部感知损失网格大小影响

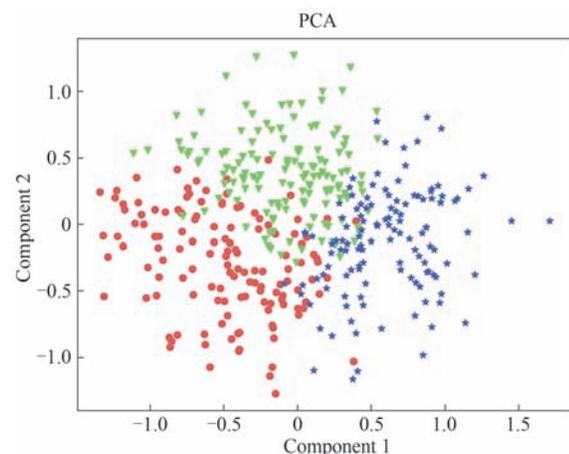
表7表示该文对动态局部感知损失中的网格尺寸参数进行了消融实验。该参数决定了输入图像在高度和宽度维度上的划分数量。通过对比16和32两种尺寸的实验结果,表明该参数会显著影响模型最终性能。基于实验结果,最终选择32作为默认网格尺寸大小。

表7 动态局部感知损失网格大小选择

参数	A→B		A→Q	
	MAE	RMSE	MAE	RMSE
网格大小				
16	11.7	23.5	117.8	209.6
32	10.4	18.2	113.8	202.8

4.8 子域划分

MPCount^[12]指出当源域服从窄分布时基于K-Means聚类的子域划分效果受限。为了说明本文方法在窄分布上基于K-Means聚类的子域划分仍然有效,本文对SHTechB^[16]测试数据集进行子域划分,子域划分数量设置为3。选择SHTechB测试集,主要是该数据集都是来自街道场景,并且图像中场景分布更加相似,导致数据分布更窄。首先,本文获取每一张图像并通过模型特定域分支(DS-Branch)之后获得域特征。其次计算均值和方差并拼接起来,维度为[1,512],将计算的均值和方差作为子域划分依据,这一点遵循了DCCUS^[13]。最后通过PCA(Principal Component Analysis)主成分分析方法^[42]将上面计算的均值和向量维度减小到2维,分别表示为Component1和Component2,使用Python工具进行可视化。如图13所示,可视化实验效果表明,在数据分布更窄的情况下,模型依然能划分出清晰的子域边界,说明本文提出的方法在保持模型轻量化的同时,模型的在窄域划分上依然有效。

图13 SHTechB^[16]测试数据集划分子域可视化

5 结论

本文针对现有单域泛化人群计数模型存在的参数量过大和推理速度慢的挑战,提出了轻量级单域泛化框架LB-CountX。首先,CountX轻量化模型相

比最新单域泛化人群计数模型具有更少的参数量。其次,引入交叉注意力模块CCA丰富了特征表达并稳定了训练过程。最后,双重感知机制结合了用于局部密度估计的动态局部感知损失和用于提高人群边缘检测的边界感知损失。经SHTechA、SHTechB、UCF-QNRF和JHU-CROWD数据集的大量评测结果和可视化分析表明,LB-CountX在减少参数量和加快推理速度的同时,实现了最好的MAE和RMSE性能,加速推动了单域泛化人群计数模型在实际场景中的部署。

致 谢 我们衷心感谢各位专家在审稿过程中对本论文提出的宝贵意见!

参 考 文 献

- [1] Shi X, Li X, Wu C, et al. A real-time deep network for crowd counting//Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain, 2020: 2328-2332
- [2] Wang P, Gao C, Wang Y, et al. MobileCount: An efficient encoder-decoder framework for real-time crowd counting. *Neurocomputing*. 2020, 407: 292-9
- [3] Huang Z-K, Chen W-T, Chiang Y-C, et al. Counting crowds in bad weather//IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France 2023: 10.1109/ICCV51070.2023.02130
- [4] Kong W, Shen J, Li H, et al. RestoCrowd: Image restoration coadjutant learning for hazy-weather crowd counting without weather labels. *Pattern Recognition*. 2025, 161(3): 111291
- [5] Zhang Shi-Hui, Zhao Wei-Bo, Wang Lei, et al. MSIANet: Multi-scale interactive attention crowd counting network. *Journal of Electronics & Information Technology*. 2023, 45(6): 2236-2245
(张世辉, 赵维勃, 王磊, et al. MSIANet:多尺度交互注意力人群计数网络. *电子与信息学报*. 2023, 45(06): 2236-45)
- [6] Wang Z, Li Y, Wan J, et al. Diffusion-based data augmentation for object counting problems//Proceedings of the ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Fukuoka, Japan, 2025: 1-5
- [7] Cai Y, Chen L, Guan H, et al. Explicit invariant feature induced cross-domain crowd counting//Proceedings of the AAAI Conference on Artificial Intelligence. Washington, USA, 2023, 37(1): 259-267
- [8] Guo M, Chen B, Yan Z, et al. Virtual classification: Modulating domain-specific knowledge for multidomain crowd counting. *IEEE Transactions on Neural Networks and Learning Systems*. 2024, 36(2):1-15
- [9] Tan X, Ishikawa H. Dataset-Level Directed Image Translation for Cross-Domain Crowd Counting//Proceedings of the 2023 IEEE International Conference on Image Processing (ICIP). Kuala Lumpur, Malaysia, 2023:400-404
- [10] Qin X, Wang J, Chen Y, et al. Domain generalization for activity recognition via adaptive feature fusion. *ACM Transactions on Intelligent Systems and Technology*. 2022, 14(1): 1-21
- [11] Kim H, Kang Y, Oh C, et al. Single domain generalization for lidar semantic segmentation//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Vancouver, Canada, 2023: 17587-17598
- [12] Peng Z, Chan S-H G. Single domain generalization for crowd counting//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2024: 28025-28034
- [13] Du Z, Deng J, Shi M. Domain-general crowd counting in unseen scenarios//Proceedings of the AAAI Conference on Artificial Intelligence. Washington, USA, 2023, 37(1): 561-570
- [14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition//International Conference on Learning Representations. San Diego, USA, arXiv preprint arXiv:1409.1556, 2015
- [15] Huang Z, Wang X, Huang L, et al. Ccnet: Criss-cross attention for semantic segmentation//Proceedings of the IEEE/CVF international conference on computer vision. Seoul, Republic of Korea, 2019: 603-612
- [16] Zhang Y, Zhou D, Chen S, et al. Single-image crowd counting via multi-column convolutional neural network//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas, USA, 2016: 589-597
- [17] Idrees H, Tayyab M, Athrey K, et al. Composition loss for counting, density map estimation and localization in dense crowds//Proceedings of the European conference on computer vision (ECCV). Munich, Germany, 2018:532-546
- [18] Sindagi V A, Yasarla R, Patel V M. Jhu-crowd++: Large-scale crowd counting dataset and a benchmark method. *IEEE transactions on pattern analysis and machine intelligence*. 2020, 44(5): 2594-609
- [19] Huang Z, Sinnott R O. Improved Knowledge Distillation for Crowd Counting on IoT Devices//Proceedings of the 2023 IEEE International Conference on Edge Computing and Communications (EDGE). Chicago, USA, 2023: 207-214
- [20] He T, Xia Z, Chen J, et al. Target-agnostic source-free domain adaptation for regression tasks//Proceedings of the 2024 IEEE 40th International Conference on Data Engineering (ICDE). Utrecht, Netherlands, 2024: 1464-1477
- [21] Wang Q, Gao J, Lin W, et al. Learning from synthetic data for crowd counting in the wild//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Long Beach, USA, 2019: 8198-8207
- [22] Wu Q, Wan J, Chan A B. Dynamic momentum adaptation for zero-shot cross-domain crowd counting//Proceedings of the 29th ACM international conference on multimedia. Chengdu, China, 2021: 658-666
- [23] Liu Y, Wang Z, Shi M, et al. Towards unsupervised crowd

- counting via regression-detection bi-knowledge transfer// Proceedings of the 28th ACM international conference on multimedia. Seattle, USA, 2020: 129-137
- [24] Zhang Y, Li W, Sun W, et al. Single-source domain expansion network for cross-scene hyperspectral image classification. *IEEE Transactions on Image Processing*. 2023, 32: 1498-512
- [25] Xu Q, Yao L, Jiang Z, et al. Dirl: Domain-invariant representation learning for generalizable semantic segmentation// Proceedings of the AAAI conference on artificial intelligence. Vancouver, Canada, 2022, 36(3): 2884-2892
- [26] Jin X, Lan C, Zeng W, et al. Style normalization and restitution for domain generalization and adaptation. *IEEE Transactions on Multimedia*. 2021, 24: 3636-51
- [27] Pan X, Luo P, Shi J, et al. Two at once: Enhancing learning and generalization capacities via ibn-net// Proceedings of the european conference on computer vision (ECCV). Munich, Germany, 2018:464-479
- [28] Pan X, Zhan X, Shi J, et al. Switchable whitening for deep representation learning// Proceedings of the IEEE/CVF international conference on computer vision. Seoul, Republic of Korea, 2019: 1863-1871
- [29] Liu Z, Mao H, Wu C-Y, et al. A convnet for the 2020s// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New Orleans, USA, 2022: 11976-11986
- [30] Song Q, Wang C, Wang Y, et al. To choose or to fuse? scale selection for crowd counting// Proceedings of the AAAI conference on artificial intelligence. 2021, 35(3): 2576-2583
- [31] Canny J. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*. 1986, (6): 679-98
- [32] Ma Z, Wei X, Hong X, et al. Bayesian loss for crowd count estimation with point supervision// Proceedings of the IEEE/ CVF international conference on computer vision. Seoul, Republic of Korea, 2019: 6142-6151
- [33] Wang B, Liu H, Samaras D, et al. Distribution matching for crowd counting. *Advances in neural information processing systems*. 2020, 33: 1595-607
- [34] Shu W, Wan J, Tan K C, et al. Crowd counting in the frequency domain// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New Orleans, USA, 2022: 19618-19627
- [35] Lin H, Ma Z, Ji R, et al. Boosting crowd counting via multifaceted attention// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New Orleans, USA, 2022: 19628-19637
- [36] Zhu H, Yuan J, Yang Z, et al. Fine-grained fragment diffusion for cross domain crowd counting// Proceedings of the 30th ACM International Conference on Multimedia. Lisboa, Portugal, 2022: 5659-5668
- [37] Zhu HZ, Yuan J, Zhong X, et al. Daot: Domain-agnostically aligned optimal transport for domain-adaptive crowd counting// Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, Canada, 2023: 4319-4329
- [38] Zhu H, Yuan J, Zhong X, et al. Find gold in sand: Fine-grained similarity mining for domain-adaptive crowd counting. *IEEE Transactions on Multimedia*. 2023, 26: 3842-55
- [39] Mansilla L, Echeveste R, Milone D H, et al. Domain generalization via gradient surgery// Proceedings of the IEEE/ CVF international conference on computer vision. 2021: 6630-6638
- [40] Choi S, Jung S, Yun H, et al. Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 11580-11590
- [41] Yang M, Li Z, Zhang J, et al. Taste More, Taste Better: Diverse Data and Strong Model Boost Semi-Supervised Crowd Counting// Proceedings of the Computer Vision and Pattern Recognition Conference. Nashville, Tennessee, USA, 2025 : 24440-24451
- [42] Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*. 1987, 2(1): 37-52



ZHOU Li-Fang, Ph. D. professor. Her research interests mainly focus on pattern recognition and medical image processing.

ZHANG Xu-Bo, M. S. candidate. His research interests are crowd counting and compute vision.

Background

Domain generalization for crowd counting (SGD Crowd Counting) represents a broader research area within crowd counting, which aims to enable models trained on a specific scene to effectively generalize to other unseen crowd scenarios. This capability is particularly crucial for real-time, low-latency, and accurate crowd analysis in practical applications such as intelligent traffic management, public space planning, and the supervision of densely crowd areas.

Significant advancements have been achieved internationally in addressing domain generalization challenges, with methodologies primarily categorized into single-domain and multi-domain approaches. Single-domain methods involve training models on a single dataset and subsequently applying them to multiple unseen scenes, though research in this area remains relatively limited. Conversely, multi-domain approaches train models on multiple datasets and attempt to generalize across diverse unseen environments. However, existing single and multi-domain crowd counting methods often overlook the requirements of low latency and real-time inference, as the latest models tend to employ large-parameter architectures that result in slow inference speeds, thereby restricting practical real-time deployment.

To improve the practical application of single-domain

generalization models in crowd counting, this paper proposes a lightweight framework, LB-CountX, with the design of dynamic local and boundary perception losses to mitigate these limitations. A dynamic local perception loss is designed to co-optimize global context and fine-grained density estimation, and a boundary perception loss based on Canny edge-derived density maps to sharpen crowd margin localization in dense scenes, which improving accuracy in densely crowd environments. Extensive evaluations on SHA, SHB, UCF-QNRF, and JHU-CROWD show that LB-CountX achieves state-of-the-art MAE and RMSE with significantly fewer parameters and faster inference, making it ideal for real-time deployment in unseen environments.

The primary objective of this research is to advance the development of lightweight, single-domain generalization methods in crowd counting, providing both empirical evidence and theoretical insights to inspire future research and facilitate broader practical applications.

This work was supported by the National Natural Science Foundation of China (Nos. 62276039, 62331008), the Guizhou University Key Laboratory of Advanced Manufacturing Technology MOE (No. KY[2022]376), the Chongqing Science Health Joint Medical Scientific Research Project(No. 2024ZDXM022).