Idea：空洞卷积、跨层连接、上采样

解决问题：

* 空洞卷积可以扩大感受野，可以适应人头的大小不一，即透视畸变/透视畸变问题
* 网络越深，感受野越大，细节信息往往会丢失，但人群计数属于像素级回归任务，往往需要的就是细节信息，所以为了弥补这一问题，采用跨层连接可以融合浅层特征与深层特征，而不仅仅是只针对深层特征做预测，加入了浅层特征融合细节信息
* 上采样是因为之前的各种网络都是将密度图下采样多少倍然后再乘以一个数，这样就会造成数值上的误差，上采样至相同尺寸就是为了避免这种问题。

Title

之前：基于融合多尺度上下文信息的密集人群计数方法（网络）

模板：用于拥挤人群计数**的**上下文感知多尺度聚合网络

Context-Aware Multi-Scale Aggregation Network for Congested Crowd Counting

题目1：用于拥挤人群计数**的**上下文感知空洞卷积残差网络

A context-aware Dilated Convolutional Residual Network for Congested Crowd Counting

题目2：用于拥挤人群计数**的**上下文信息跨层连接**的**空洞卷积神经网络

A Dilated Convolutional Neural Network for Cross-Layers of Contextual Information for Congested Crowd Counting

Abstract

人群计数任务旨在从人群图像中统计得到准确的人数，精确的计数结果可以为更高级的视觉任务如行人异常检测等提供有用信息。然而在高度拥挤的场景中，由于透视畸变以及尺寸变换问题，人群计数仍是一个具有挑战的问题。为了解决这些问题，我们提出了一种新的用于拥挤人群计数的上下文信息跨层连接的空洞卷积神经网络（CL-DCNN），它的前端是传统的深度卷积神经网络用于提取特征图，后端级联了单个或多个即插即用的DCM模块，其中的DCM模块采用多个空洞卷积扩大感受野以适应透视畸变问题，同时为了克服连续的空洞卷积造成的细节信息丢失的问题，DCM实现上下文信息的跨层连接以融合浅层特征的细节信息。在四个公开的人群计数数据集（即ShanghaiTech、Mall、 UCF\_CC\_50 和 UCF\_QNRF）上的广泛实验表明，我们的CL-DCNN提供了由于优于最先进方法的性能。

Keywords

1 Introduction

P1 人群计数的重要性/必要性

过去几十年以来，全世界人群数量呈指数级增长，快速增长的人口给公共安全带来了压力，在公共场所中统计人数对于社会安全和控制管理起着不可或缺的作用。且人群计数任务为更高级的人群分析任务如人流预警*[人流预警]*、交通管制*[交通管制]*、行人异常检测*[行人异常检测]*、安防监控*[安防监控]*等带来有益效果。可见，人群计数任务是非常重要且有必要的。

P2 人群计数任务的挑战/难点

然而，人群计数面临着一系列挑战。如图1所示，由于行人离摄像设备的远近不同从而造成人群图像内行人呈现不同的大小，称之为人群图像内的透视畸变（见图1.a）。其次，由于现实背景比较杂乱，容易对人群的检测造成影响，例如图1.b中的灌木丛被当作行人。最后，人群之间遮挡严重（见图1.c），这些问题都对人群计数任务带来困难和挑战。‘

图1 人群计数任务的困难与挑战 (a)透视畸变，图中不同部位的人有不同的大小(b)背景干扰(c)人群遮挡严重

P3 最近人群计数方法总结（传统/深度学习）

近几年有许多方法被提出以解决人群计数任务，主要分为三类：基于检测的方法*[基于检测的方法]*、基于回归的方法*[基于回归的方法]*和基于深度学习*[基于深度学习的方法]*的方法。基于检测的方法使用一个移动窗口式检测器来识别图像中的人，并将此信息用于计数任务中，检测有两种方式，基于整体*[基于整体检测的方法]*和基于局部*[基于局部检测的方法]*的方式，用检测器提取出整体或局部特征后，训练分类器，例如支持向量机*[支持向量机]*、Boosting*[Boosting]*和随机森林*[随机森林]*等。例如Topkaya I S等人*[2010-2018, 2014]*通过采用基于狄利克雷过程混合模型的聚类方案来估计场景中的人数。这种方法不能很好处理遮挡严重的问题。基于回归的方法分为两步，第一步从人群图像中提取有效特征，早期利用手工特征，如方向梯度直方图HOG*[HOG]*、 尺度不变特征SIFT*[SIFT]*、局部二值模式LBP*[LBP]*、灰度共生矩阵GLCM*[GLCM]*等，第二步利用各种回归函数来估计人群数量，回归方法包括线性回归*[线性回归]*或高斯过程回归*[高斯过程回归]*等。例如Antoni B.Chan等人*[透视图，2008]*采用基于特征的回归来完成人群计数，用动态纹理的方法分割出运动的人群，用高斯过程回归将提取的特征回归到图像中的人群数量。这种方法由于直接根据特征回归计数结果，不能反映人群的分布情况。目前深度学习已经成功应用在各种计算机视觉任务中并取得了有益成果，例如目标检测*[目标检测]*、人群计数*[人群计数]*、目标跟踪*[目标跟踪]*等等。深度学习应用在人群计数任务中主要是生成人群图像的密度图*[密度图]*，密度图反应人群图像的密度分布情况，对密度图积分求和得出人群计数结果。

P4 基于深度学习的人群计数方法总结

近几年有很多深度学习模型被提出已解决人群计数中的各种问题，例如为了解决透视畸变问题，Zhang等人提出一种多列卷积神经网络MCNN*[MCNN，2016]*，采用不同大小的卷积核同时提取特征，不同列的卷积核大小不同，特征图的感受野不同，最后进行通道维度的拼接。这种方法也成为典型的多列结构，以后的多种方法也是借鉴于这种思想。文献*[CSRNet, 2018]*提出，这种多列结构复杂度高，计算量大，且每一列提取的特征信息相差无几，为了解决这一问题，文献*[SaCNN, 2018]*提出了一种自适应透视畸变SaCNN，通过融合浅层特征与深层特征，而不是多列结构。为了适应人群图像中不同大小的人头，文献*[CSRNet, 2018]*提出CSRNet，在后端子网中采用多个空洞卷积生成密度图，但这种方法只是对深层特征图做预测，忽略了浅层特征的细节信息。以上的网络在一定程度上提高了人群计数的准确率，但它们有一个共同问题，就是生成的密度图是原图的1/8或1/16， 这就需要将真实的密度图下采样至相同尺寸，再乘以一个因子，此时真实密度图的结构信息会被破坏，与预测密度图之间的SSIM指标会出现误差，从而对结果造成影响。

P5 本文提出的方法、解决了什么问题

基于以上问题，本文提出一种用于拥挤人群计数的上下文信息跨层连接的空洞卷积神经网络CL-DCNN，主干采用VGG16的前10层卷积层提取特征图。为了克服人群图像的透视畸变问题，我们提出了一个新的空洞卷积模块（DCM），DCM有两个分支，一个分支对输入做1\*1卷积降低通道数保留上下文信息，另一个分支对输入采用连续的空洞卷积进一步扩大感受野，提取不同尺度大小的特征。将两个分支进行通道维度的拼接，实现深层特征与浅层特征的跨层连接。值得注意的是，DCM作为一种即插即用模块，我们所提出的CL-DCNN通过在CNN主干级联多个DCM形成。最后采用1\*1卷积将通道数降为1，生成预测密度图并上采样至与原图相同尺寸，这样可直接与真实密度图做损失计算，避免因下采样带来的计数误差。

P6 本文做出贡献（2-3个点）、文章结构安排

综上所述，本文做出以下几点贡献：

1)提出一种即插即用的空洞卷积模块（DCM），DCM有两个分支，一个分支对输入做1\*1卷积降低通道数保留上下文信息，另一个分支对输入采用连续的空洞卷积进一步扩大感受野，提取不同尺度大小的特征。将两个分支进行通道维度的拼接，实现深层特征与浅层特征的跨层连接。

2)通过在CNN的主干上级联多个DCM，构建用于拥挤人群计数的上下文信息跨层连接的空洞卷积神经网络CL-DCNN；

3)在几个主流人群计数数据集上与最先进的方法做出对比，实验结果证明所提出的CL-DCNN的有效性。

2 Related work

人群计数方法大致可以分为三类，即基于检测的方法、基于回归的方法和基于CNN的方法。在本节中，我们只回顾与我们的方法最相关的基于CNN的方法，包括基于多列结构*[多列结构]*和空洞卷积*[空洞卷积]*的人群计数方法。

2.1 Multi-column CNN-based Methods

由于摄像设备距离行人远近不同，行人头部大小在人群图像的不同位置有很大差异，这称为人群图像内的尺度变化问题。

许多早期成功的工作通常采用多列网络结构来解决这一问题，例如，Zhang等人*[MCNN，2016]*提出了有三个感受野不同的分支组成的多列卷积神经网络MCNN。但后来的研究*[CSRNet，2018]*证明该方法有两个明显的缺点：大量的训练时间和无效的分支结构，作者通过实验证明MCNN中的每一列都学习了几乎相同的特征，这违背了MCNN设计的初衷，即每列学习不同的特征。Hydra-CNN*[Hydra-CNN, 2016]*使用对应于不同尺度的图像片金字塔来学习用于最终密度图估计的多尺度非线性回归模型。*[CrowdNet-2016]*结合了不同列的浅层和深层网络，其中浅层网络捕获对应于大尺度变化的低层特征，而深层网络捕获高层语义信息。*[Switching CNN，2017]* 在图像块上训练几个独立的 CNN 人群密度回归器，回归器与 MCNN具有相同的结构。 一个Switch分类器也在回归上进行交替训练，以选择最佳的一个用于密度估计。

在上述工作的基础上，Wu等人*[ASD，2019]*发现通过CNN预测密度图统计计数结果与场景有关：稀疏场景预测结果要比真实结果偏小，密集场景预测结果比真实结果偏大。作者提出一种用于人群计数的自适应场景发现框架ASD，由两条平行路径构成，这些路径使用不同大小的感受野进行训练，将场景看作两条路径的线性组合，通过离散化的权重，设计第三个自适应分支来学习感知场景的响应，以服务于不同的规模和人群密度。Shi等人*[PACNN，2019]*设计出PACNN，该网络包括两个密度图回归分支，并从透视感知的角度加权自适应地组合两个密度输出，其中权重通过网络中预测透视映射的非线性变换来学习。同样基于透视感知，Gao等人*[PCCNet，2019]*提出一个透视人群计数网络PCCNet，该网络作为一个多列结构，由密度图估计器、随即高级密度分类、前背景分割模块组成。其中(R-HDC)提取全局特征以预测图像中随机斑块的粗略密度标签，前景/背景分割(FBS)编码中级特征以分割前景和背景。Wei等人提出的MSPNet*[MSPNet，2020]*在训练过程中进行了多次监督，可以补充池化和上采样操作中丢失的细节，从而提高密度图的质量。

进一步的，Yang等人*[2020-3]*通过从估计的密度图中探索透视信息，并将透视空间量化为几个单独的场景并将透视分析嵌入到具有循环连接的多列框架中，所提出的网络有效地将各种尺度与不同的感受野相匹配。Sajid等人*[2021-4]*提出了一个新的基于PRM的多分辨率多任务人群计数网络，该模型由三个层次较深的分支组成，每个分支生成不同分辨率的特征图，这些分支执行彼此之间的特征级融合，以构建用于最终群体评估的重要集体知识。Zand等人*[2022-2]*提出了一种基于点监督的多尺度多任务结构，能够有效地估计出图像中人数变化较大时的数量和位置。

综上所述，不论何种多列结构，他们都有一个共同缺点，每个分支提取的特征大多没有明显差别，反而增加了计算量，所以现在很多结构已经摒弃了这种用法，转而使用高效的单列结构。

2.2 Dilated Convolution-based Methods

空洞卷积首次在语义分割任务*[空洞卷积首次被提出的语义分割]*中被提出，它是为了代替池化层，可以很好的聚合上下文信息而不会丢失分辨率。现在已广泛应用于各种计算机视觉任务。

空洞卷积首次应用在人群计数任务中是在CSRNet*[CSRNet，2018]*中，作者在后端子网中使用多个空洞卷积扩大感受野得到密度图，以适应人群图像的透视畸变。DADNet*[DADNet-2019]*采用具有不同扩张率的扩张卷积神经网络来捕获更多的上下文信息作为前端，并采用自适应可变形卷积作为后端来准确定位对象的位置。Yan等人提出的PGCNet*[PGCNet-2019]*旨在通过考虑透视信息来解决连续尺度变化问题，PGCNet 是单列的，计算量适度增加。进一步的，作者在2021年提出了PFDNet*[PFDNet-2021]*，通过对连续尺度变化进行建模，PFDNet 能够选择适当的分数扩张核来适应不同的空间位置，显著提高了仅考虑离散代表尺度的最新技术的灵活性。Sooksatra等人*[2020-2]*通过将高级特征传递给浅层并强调其低级特征，提出了一种估计网络。为了保留语义信息，使用扩张卷积而不调整特征图的大小。为了提高具有大感受野的列的评估精度，Yang等人*[2020-3]*提出了一种变换扩张卷积，变换扩张卷积打破了深度网络的固定采样结构。Wang等人*[2021-5]*提出了一种用于精确人群计数的金字塔扩张深度卷积神经网络，称为PDD-CNN。采用两个金字塔扩张模块，每个模块由四个不同速率的平行扩张卷积层和一个平行平均池化层组成，以捕获多尺度特征。

空洞卷积虽然可以扩大感受野，提取更深层次的特征，但在深度网络中感受野越大，丧失的细节信息就越多，而人群计数任务是像素级的回归任务，在密集场景中最关键的就是细节信息，所以上述结构单纯的采用空洞卷积扩大感受野，丧失了很多浅层特征中的细节信息，造成计数精度不高。

3 Proposed method

在这一节中，我们首先描述了DCM的原理，然后介绍CL-DCNN的网络结构，最后简单介绍Ground truth 的生成、损失函数以及模型的评价指标。其中CL-DCNN的网络结构图如x图所示。



Figure 2 CL-DCNN网络结构图

3.1 Dilated Convloutional Module

空洞卷积已被证明在分割任务中产生了显著提高的精度*[空洞卷积首次被提出的语义分割]*，它是池化层的理想替代方案，其中池化层（平均池化、最大池化）是被用来扩大感受野、减少参数，池化之后的特征图往往会变成原来的1/2，这样不可避免的会造成像素丢失，且这种损失是不可逆的，要想恢复原来的尺寸，就需要上采样，这时又会造成二次损失。空洞卷积之所以可以替代池化层是因为它可以在不改变参数量的前提下扩大感受野，且可以保持原来的尺寸。空洞卷积涉及空洞率，空洞卷积通过将特征图或卷积核内插（空洞率-1）个0，再进行普通卷积。当空洞率=1时，空洞卷积和普通卷积相同，当空洞率>1时，空洞卷积可以扩大特征图的感受野， 2D空洞卷积可以定义如下。



其中，和分别表示空洞卷积的输入与输出，代表长度为、宽度为的卷积核，是空洞卷积中的参数空洞率。

连续的空洞卷积在一步步扩大感受野的同时会显著降低空间分辨率，这意味着特征图的空间信息会丢失。为了弥补这个问题，所提出的DCM有两个分支，一个分支对输入做1\*1卷积降低通道数保留上下文信息，另一个分支对输入采用连续的空洞卷积进一步扩大感受野，具体的，我们先对输入采用1\*1卷积降维，然后采用3个连续的空洞卷积扩大感受野，图x中的CBR代表Convolution-Batch Normailzation-Relu操作，下面数字的含义分别为kernel size-channels-padding-stride 。DBR代表Dialted Convolution-Batch Normalization-Relu操作,下面数字的含义分别为kernel size-channels-padding-stride-dilatation，将两个分支进行通道维度的拼接，实现深层特征与浅层特征的跨层连接。值得注意的是，DCM的输出与输入的特征图尺寸、通道数相同，这个特性使得DCM成为一个即插即用的模块。

3.2 CL-DCNN

图 2 说明了我们的CL-DCNN的架构。在这项工作中，为了建立一个可以接受任意大小输入的网络，我们去掉了VGG-16的全连通层，并在VGG-16中建立了具有全卷积层的CL-DCNN。如图2所示，我们保留了VGG-16的前10个卷积层作为特征提取层，只保留了三个最大池层，这意味着输入图像将减少8倍(即，每个最大池化层将特征图减少到其原始大小的一半)。我们所提出的CL-DCNN在CNN主干后级联多个DCM。最后采用1\*1卷积将通道数降为1，生成预测密度图并上采样至与原图相同尺寸，这样可直接与真实密度图做损失计算，避免因下采样带来的计数误差。

3.3 Ground truth generation

在本节中，我们将介绍将手动标记的图像转换为密度图的方法。在一张人群图像中，假设像素处有一个人头目标，将其表示为函数。创建一张与原图大小一致的全0矩阵，在矩阵中将该点处的值置为1，则带有N个人头标记的图像可以用如下公式表示：



其中代表二维图像中的坐标，代表人头标记总数。对单一人头标记而言，将人头中心位置利用高斯核平滑处理，使得每个高斯核的像素和为1，通过对密度图整体积分求和得到总人数，人群密度图可由二值矩阵与标准2D高斯核卷积可得，二维高斯核定义如公式所示。



其中表示二维坐标，σ表示高斯核的标准差。则密度图可用如下公式表示：



其中，\*表示卷积操作，是系数因子，代表个最近邻人头距离的平均距离，不同数据集由于人群分布的稀疏程度不同，采用的高斯核的标准差也不同，在稀疏场景中，人群分布较均匀，采用固定大小尺寸的高斯核，在密集场景中，由于视角畸变严重，采用个最近邻人头间的平均距离作为高斯核的标准差，生成自适应高斯核，一般采用。

3.4 Loss Function

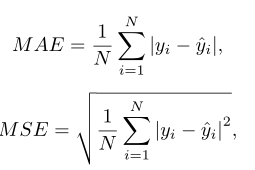
基于CNN的密度图估计人群计数方法大多是一种回归任务，通常采用欧氏距离作为损失函数来衡量估计的密度图与地面真实度之间的差异。定义如下：



其中是CL-DCNN生成的地面真值密度图，而是输入图像的地面真值；是CL-DCNN中的一组可学习参数，是训练批的大小，是地面真实值和估计密度图之间的损失。

3.5 Evaluation metrics

在人群计数领域，主流的评价指标是平均绝对误差（MAE）和均方误差（MSE），定义如下：



其中 N 是测试集中的样本数，yi 是计数标签，yi 是第 i 个测试样本的估计计数值。

4 Experiments

在本节中，我们将评估所提的方法。我们首先介绍实验中使用的四个基准数据集和实验细节。然后，我们将所提方法在四个数据集上与现有方法进行比较。最后，我们根据所提方法做出消融实验。

4.1 Datasets and implementation details

Our experiments are mainly conducted on four crowd counting benchmark datasets, which are ***ShanghaiTech****[ShanghaiTech]* ***Mal******l****[Mall]*, ***UCF\_CC\_50****[UCF\_CC\_50]* and ***UCF\_QNRF****[UCF\_QNRF]*. The datasets specific statistics of them are listed in Table x. Some samples from the representing datasets are depicted in Fig. x.

TABLE I: Statistics of the four crowd counting benchmark datasets. Total, Min, Avg and Max represent the total number, the minimum, average number and maximum number of instances in the datasets, respectively.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset | Year | Number of Imgaes | Trainging/Test | Average Resolution | Count Statistics | | | |
| Total | Min | AvgSH | Max |
| SHTA*[ShanghaiTech]* | 2016 | 482 | 300/182 | 589 \* 868 | 241,677 | 33 | 501.4 | 3,139 |
| SHTB*[ShanghaiTech]* | 2016 | 716 | 400/316 | 768 \* 1024 | 88,488 | 9 | 123.6 | 578 |
| Mall*[Mall]* | 2012 | 2000 | 800/1200 | 320 \* 240 | 62,325 | 13 | 31 | 53 |
| UCF\_CC\_50  *[UCF\_CC\_50]* | 2013 | 50 | — | 2101 \* 2888 | 63,974 | 94 | 1,280 | 4,543 |
| UCF\_QNRF  *[UCF\_ QNRF]* | 2018 | 1,535 | 1201/334 | 2013 \* 2902 | 1,251,642 | 49 | 815 | 12,865 |

***ShanghaiTech****[ShanghaiTech]*是一个近年来应用广泛的大规模的人群计数数据集之一，由1198幅图像和330165条注释组成。根据密度分布的不同，将数据集分为两部分：A部分（SHTA）和B部分（SHTB）。SHTA是从互联网上随机选择的图像，B部分是从上海大都市的一条繁忙街道拍摄的图像。SHTA中的密度比SHTB中的密度大得多。

***Mall****[Mall]*是从购物中心的监控视频中收集的数据集。数据集中的视频序列由 2000 帧组成，大小为 320×240，总共包含 62325 名行人。***Mall***在更显著的光照条件下覆盖了更多的多样性密度以及不同的活动模式（静止和运动的人），并且由于场景对象而具有严重的遮挡。

***UCF\_CC\_50****[UCF\_CC\_50]*是由中佛罗里达大学的Idrees等人开发的，它只包含50张图片，但有63,075个被标记的个体，包括音乐会、抗议、体育场和马拉松等不同场景的各种密度和不同的视角扭曲。它包括一个很宽的密度范围(个体范围从94到4543)，以不同的视角失真覆盖不同的场景。由于只有50幅图像，我们使用标准的5倍交叉验证来评估算法，这也被其他先进的方法采用*[使用UCF\_CC\_50的模型]*。

***UCF-QNRF****[UCF\_QNFR，2018]*是一个大规模的极其拥挤的人群计数数据集，由从Flickr、网络搜索和朝圣镜头捕获的1,535张人群图像组成, 约125万条注释。它是最大规模的拥挤数据集，计数在49到12,865之间。该数据集中的图像包含更多种类的场景，并包含最多样化的视点、密度和照明变化。与***ShanghaiTech*** *[ShanghaiTech]*不同的是，这里的人群密度和图像分辨率都有很大的变化**错误!未找到引用源。**。

The training and evaluation are performed on NVIDIA RTX 2080 Ti GPU using PyTorch framework*[C-3-Framework，2019]*. 原始训练图像（大小为 768 × 1024）被随机裁剪为 576 × 768 大小。相应地，密度图和分割图被相应地裁剪，并重新计算计数标签。此外，在训练阶段采用随机水平翻转。The Adam optimizer [50] with a small learning rate of 1e-5 was used to train the model. 文献中*[C-3-Framework，2019]*提出，当密度图乘以一个因子时，收敛的更快，借鉴此想法，我们在实验中使用了相同的设置。All experiments were trained for 600 epochs.

4.2 Comparison with state of the arts

在本节中，我们展示了所提出的CL-DCNN在四个数据集上的表现，并与最先进的方法做出对比。最先进方法包括MCNN，Switching-CNN，CSRNet，ASD，PACNN，PGCNet，2022-2，Bidirectional ConvLSTM，ACSPNet，MSPNet，PFDNet，CL-DCNN，DT-CNN，CAN，SFCN and DADNet。

如表x所示，所提出的方法针对ShanghaiTech数据集的最新方法进行了评估。对于SHTA，我们的方法达到了最低的55.1的mae，相对于MCNN提高了50%，相对于CSRNet提高了19%。mse仅次于PGCNet，达到了91.0。对于SHTB，所提的CL-DCNN实现了最低的8.3/13.0的mae/mse，相对于MCNN提高了68%，相对于CSRNet提高了21%。这表明我们的方法相比多列结构（MCNN）和简单的堆叠多个空洞卷积（CSRNet）实现了更好的效果，进一步验证了所提方法的合理性。另外，从表1可以看出，不论是在密集场景，还是在稀疏场景，我们的方法都实现了良好的性能。

表1 Estimation errors on the ShanghaiTech dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Methods | Year | SHTA | | SHTB | |
|  | MAE | MSE | MAE | MSE |
| Zhang et al. | 2015 | 181.8 | 277.7 | 32.0 | 49.8 |
| MCNN | 2016 | 110.2 | 173.2 | 26.4 | 41.3 |
| Switching-CNN | 2017 | 90.4 | 135.0 | 21.6 | 33.4 |
| CSRNet | 2018 | 68.2 | 115.0 | 10.6 | 16.0 |
| ASD | 2019 | 65.6 | 98.0 | 8.5 | 13.7 |
| PACNN | 2019 | 66.3 | 106.4 | 8.9 | 13.5 |
| PGCNet | 2019 | 57.0 | **86.0** | 8.8 | 13.7 |
| 2022-2*[2022-2]* | 2022 | 71.4 | 110.7 | 9.6 | 15.0 |
| CL-DCNN(ours) |  | **55.1** | 91.0 | **8.3** | **13.0** |

表x显示了所提出的CL-DCNN与最先进方法在Mall数据集上的效果对比，结果表明，我们的方法在Mall数据集上达到了1.55的MAE和2.01的MSE，分别相比MCNN提高了30%和76%。

表2 Estimation errors on the Mall dataset

|  |  |  |  |
| --- | --- | --- | --- |
| Methods | Year | Mall | |
|  | MAE | MSE |
| Zhang et al. | 2015 | 3.59 | 19.0 |
| MCNN,2016 | 2016 | 2.24 | 8.5 |
| Bidirectional ConvLSTM | 2017 | 2.10 | 7.6 |
| ACSPNet,2019 | 2019 | 1.76 | 2.24 |
| CL-DCNN(ours) |  | **1.55** | **2.01** |

我们按照参考文献[]的标准将UCF\_CC\_50数据集随机分为五个部分，每个部分包括十个图像，表x列出了CL-DCNN的结果，分别达到了181.8/240.6的MAE/MSE。表x也给出了之前在MAE和MSE评估的UCF\_CC\_50数据集的性能最好的方法。与最先进的方法相比，我们的计数系统实现了卓越的性能。例如，与次优方法(ASD)相比，CL-DCNN的MAE提高了7.3%，MSE提高了11.2%，这表明CL-DCNN可以很好地处理少量的训练数据。

表3 Estimation errors on the UCF\_CC\_50 dataset

|  |  |  |  |
| --- | --- | --- | --- |
| Methods | Year | UCF\_CC\_50 | |
|  | MAE | MSE |
| Zhang et al. | 2015 | 467.0 | 498.5 |
| MCNN | 2016 | 377.6 | 509.1 |
| Switching-CNN | 2017 | 318.1 | 439.2 |
| CSRNet | 2018 | 266.1 | 397.5 |
| MSPNet | 2020 | 206.7 | 299.3 |
| PFDNet | 2021 | 205.8 | 289.3 |
| ASD | 2019 | 196.2 | 270.9 |
| CL-DCNN(ours) |  | **181.8** | **240.6** |

表x给出了在UCF\_QNRF数据集上的结果，我们的方法在UCF\_QNRF上达到了96.4/168.7的mae/mse，相比于CSRNet，我们的方法提高了23.9个点

表4 Estimation errors on the UCF\_QNRF dataset

|  |  |  |  |
| --- | --- | --- | --- |
| Methods | Year | UCF\_QNRF | |
|  | MAE | MSE |
| MCNN | 2016 | 277 | 426 |
| Switching-CNN | 2017 | 228 | 445 |
| CSRNet | 2018 | 120.3 | 208.5 |
| CL-CNN | 2018 | 132 | 191 |
| DT-CNN | 2021 | 112 | 191 |
| CAN | 2019 | 107 | 183 |
| SFCN | 2019 | 102.0 | 171.4 |
| CL-DCNN(ours) |  | **96.4** | **168.7** |

表5 Evalution of the quality of density maps on ShanghaiTech Part A dataset

|  |  |  |
| --- | --- | --- |
| Methods | SHTA | |
| PSNR | SSIM |
| MCNN | 21.4 | 0.52 |
| Swicthing-CNN | 21.91 | 0.67 |
| CSRNet | 23.79 | 0.76 |
| DADNet | 24.16 | 0.81 |
| CL-DCNN(ours) | **26.18** | **0.83** |

4.3 Ablation experiments on SHTA dataset

为了验证DCM模块对CL-DCNN的有效性，我们对ShanghaiTech Part A进行了消融实验，并对结果进行了分析。如表x所示，我们分别试验了DCM的数量为0，1，2时对结果的影响，结果表明，当DCM的数量为1时，达到了最好的55.15/91.04的MAE/MSE值。当数量为0时模型只达到了63.82/104.02的MAE/MSE值，此时网络只有特征提取部分，没有考虑图像内的透视畸变问题。当DCM的数量为2时，又由于连续多个空洞卷积丧失了太多的细节信息从而造成了较大的计数误差，只达到了63.57/104.32的MAE/MSE。实验结果表明，将DCM数量设置为1是合适的，既利用DCM的空洞卷积扩大了感受野很好地解决了透视畸变问题，同时DCM实现了上下文信息的跨层连接避免丧失过多的细节信息。4.2节中的实验都是按照此设置进行的。

表6 on ShanghaiTech Part A dataset

|  |  |  |
| --- | --- | --- |
| Number of DCM | SHTA | |
| MAE | MSE |
| 0 | 63.82 | 104.02 |
| 1 | **55.15** | **91.04** |
| 2 | 63.57 | 104.32 |

5 Conclusion

Acknowledgements

References