

CPEPF: 通过估计人流进行人数统计

1.概述

目前最先进的方法大多采用回归器来估计单张图像中的局部人群密度,然后通过对图像部分区域进行积分来进行人数统计。这些回归器通常使用随机森林、高斯过程,或近年来更常见的深度神经网络。本文提出通过估计人员流量(即图像平面中人员在不同位置间的移动数量)而非直接计算密度的新方法。

这种方法能够施加更强的人数守恒约束,显著提升性能而且无需更复杂的网络架构。此外,还可以利用人流与光流的相关性进一步优化结果。实验表明,结合时间守恒约束的主动学习策略能够大幅减少标注需求,仅需 6.25%的标注即可达到全监督模型的性能。

2.技术细节

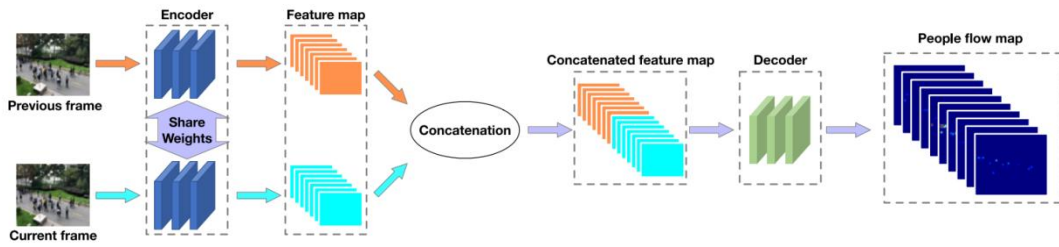


图 1 模型架构图

模型架构：基于 CAN 网络，编码器—解码器结构

将连续两帧 RGB 图像输入到同一个编码器网络,该网络基于 CAN 尺度感知特征提取器。输出后将这些多尺度特征经过进一步拼接并输入解码器网络,最终生成人员流动预测图。

2.1 约束条件

核心约束条件：时刻 t 位于区域 j 的人数，等于 $t-1$ 时刻已在该区域并保持静止的人数，加上 $t-1$ 到 t 时刻从相邻区域流入该区域的人数之和。

公式如下：

$$\sum_{i \in N(j)} f_{ij}^{t-1,t} = m_j^t = \sum_{k \in N(j)} f_{jk}^{t,t+1}$$

示意图：

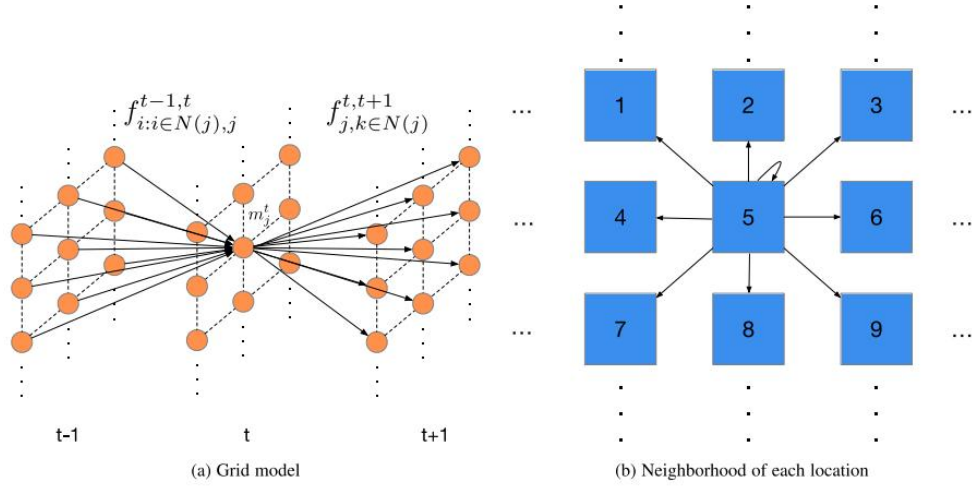


图 2 人数守恒约束示意图

- (a) 给定位置在时间 t 的人群密度只能来源于时间 $t-1$ 的相邻网格位置，并流向时间 $t+1$ 的相邻网格位置，这两种情况都包含该位置本身(即没有移动的)；
- (b) 对于不在图像边界的位置，存在九个可在一个时间步内到达的网格，包括该网格本身。对于图像平面边界的位置，添加第十个代表外部的的位置，该位置允许人员流出图像或从外部进入。

2.2 流量回归分析

训练回归器，该回归器需在遵守上述约束条件并妥善处理边界网格位置的同时，预测与观测数据相符的流量。我们将这个从图像 I_{t-1} 和 I_t 预测流量的回归器记为 F ，其待训练参数为 θ 。训练过程目标是寻找参数 θ 的最优值，使得：

$$\bar{m}_j^t = \sum_{i \in N(j)} f_{i,j}^{t-1,t} = \sum_{k \in N(j)} f_{j,k}^{t,t+1} \text{ and } f_{i,j}^{t-1,t} = f_{j,i}^{t,t+1}$$

引入损失函数：

$$L_{combi} = \sum_t L_{flow}^t + \alpha L_{cycle}^t \text{ (组合损失函数)}$$

$$L_{flow}^t = \sum_{j \in I^t} \left[\left(\bar{m}_j^t - \sum_{i \in N(j)} f_{i,j}^{t-1,t} \right)^2 + \left(\bar{m}_j^t - \sum_{k \in N(j)} f_{j,k}^{t,t+1} \right)^2 \right]$$

$$L_{cycle}^t = \sum_{j \in I^t} \left[\sum_{i \in N(j)} (f_{ij}^{t-1,t} - f_{ji}^{t,t-1})^2 + \sum_{k \in N(j)} (f_{jk}^{t,t+1} - f_{kj}^{t+1,t})^2 \right]$$

在减少带标签训练集实验中，还引入了扩展损失函数：

$$L_{overall} = L_{combi} + \gamma L_{spatial} + \delta L_{advers}$$

$$L_{spatial} = \sum_j \left(\sum_{i \in P_j^t} m_i^t + \sum_{k \in P_j} \sum_{i \in P_k^t} m_i^t - \sum_{i \in S_j^t} M_i^t \right)^2$$

$$L_{advers} = - \sum_{P \in A} \log(D(m_i^t)) - \sum_{P \notin A} \log(1 - D(m_i^t))$$

$\{P_k^t\}$: 图 I^t 分割出的网格合集; P_j^t : 被标记的网格; P_k^t ($k \in P_j$): 未被标记的

网格; S_k^t : P_j^t 超合集 m_k^t : 每张图像单独输入网络获得的人员计数; M_k^t : 超合集输入网络获得的人员计数; D : 判别器; A : 被标记的网格合集

3.模型表现

使用 CrowdFlow (合成数据集)、FDST、UCSD、Venice 和 WorldExpo10 数据集，以 MAE 和 RMSE 为指标。

对比以往方法 (如 CNN、CAN、MCNN 等) 在这五组数据集中该方法的 MAE 和 RMSE 均为最低，提高了准确率。

实验同样还表明：(1) 采用人流建模的实验效果优于直接进行人群密度估计 (人流法与人群密度法对比实验)；(2) 结合光流法进行人流估计，利用光流和人流的相关性，可以使性能进一步提高；(3) 在 FDST、Venice 和 WorldExpo10 数据集上，对比其他的主动学习方法，随着标注率的提高，该方法有更高的准确率；(4) 在地平面坐标系建模估计人流，能获得更优的性能；(5) 在使用更少的带标签训练集实验里，当标注量达到图像总数的 6.25% 时，其精度已接近全标注效果，表明这种主动学习策略既能大幅降低标注成本，又能获得接近全监督模式的准确率。