



深圳技术大学

SHENZHEN TECHNOLOGY UNIVERSITY

本科毕业论文（设计）

基于卷积神经网络的人群计数问题研究

姓 名：周冠霖

学 院：大数据与互联网学院

专 业：计算机科学与技术

学 号：202002020230

指导教师：董丽

职 称：助理教授

提交日期：2024 年 1 月 13 日

深圳技术大学本科毕业论文（设计）诚信声明

本人郑重声明：所呈交的毕业论文（设计），题目《基于卷积神经网络的人群计数问题研究》是本人在指导教师的指导下，独立进行研究工作所取得的成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式注明。除此之外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。本人完全意识到本声明的法律结果。

毕业论文（设计）作者签名：

日期： 年 月 日

目 录

摘要.....	I
Abstract.....	II
1 引言	1
1.1 研究背景与意义.....	1
1.2 主要研究工作	2
1.3 论文组织结构	2
2 相关工作.....	3
2.1 基于密度图的方法	3
2.2 基于点的方法	3
2.3 自注意力模型	4
2.4 竖排多图横排布局	5
2.5 横排多图竖排布局	5
2.6 本章小结.....	5
3 卷积神经网络用于人群计数	7
3.1 ATRM.....	7
4 公式插入示例	8
5 参考文献插入示例	9

6 总结与展望	10
6.1 工作展望.....	11
参考文献	12
致谢.....	13

基于卷积神经网络的人群计数问题研究

【摘 要】 人群计数，又叫人群密度估计，直白的讲就是一种通过计算机计算或估计图像中人数的技术，可用于安防等场景。本文章使用了一种基于混合真值的 CNN 框架——top-k 关系网络用于人群计数。同时，使用了一个自适应的 top-k 关系模块 ATRM，通过自适应滤波机制利用像素点之间的 top-k 依赖增强特征表示。这里先计算两个像素点间的相似度，得到每个像素位置的 top-k 关系。之后，使用一种自适应滤波机制的权值归一化方法，使 ATRM 自适应消除 top-k 关系中低相关位置的影响。最终，使用了一种权重关注机制，使 ATRM 更侧重于 top-k 关系中权重更高的位置。

【关键词】 人群计数；CNN 框架；ATRM 关系模块；top-k 关系

Research on Crowd Counting Problem based on convolutional Neural Network

【Abstract】 Crowd counting or crowd density estimation Estimation is a technique for computing or estimating the number of people in an image via a computer and can be used in security or other scenarios. In this paper, we perform crowd counting using top-k relational network, a CNN framework based on mixed truth values. At the same time, feature representation is enhanced by using top-k dependency between pixels via adaptive filtering mechanism using the adaptive top-k relation module atrm. The top-k relationship of each pixel position is obtained by calculating the similarity between two pixels. Next, a weight normalization method of the adaptive filtering mechanism is used to adaptively eliminate the influence of low-correlation positions in the top-k relation. Finally, the weight attention mechanism is used to ensure that the atrm pays attention to the position with the higher weight in relation to the top-k.

【Key words】 Crowd counting; CNN framework; ATRM relation module; top-k relationships

1 引言

1.1 研究背景与意义

人群计数是智能化监控系统的重要内容,旨在利用计算机视觉技术准确估计图像中的人数 [1],在安防预警、城市规划、智能商业、交通调度 [2] 等领域有着重要应用。随着深度学习技术的发展,基于卷积神经网络的人群计数算法取得了优异的性能,显著地降低了计数误差。因此,本研究拟研究基于卷积神经网络的人群计数问题,重点研究如何利用卷积神经网络提升人群计数的精度和鲁棒性。

现行的主要人群计数方法有基于行人检测的方法、基于人数回归的方法和基于人群密度估计图 [3] 的方法等。对于基于行人检测的方法,比较早的研究可以追溯到 2004 年基于 HOG 算子的行人检测,方法提出之后,有学者就基于该算法构建了行人技术算法 [5],但仅对人数较少的场景有效,因为人数较多的场景存在大量拥挤造成该算法的漏检测严重。

对于基于人数回归的方法,在 2008 到 2014 年间,基于人数回归的方法解决了行人检测方法的缺陷,能够对人数较多的场景 [7] 有效。其主要思路在于对输入图像或视频采用各种特征描述子提取特征,然后利用各种机器学习模型对人数进行回归。无论是行人检测还是人数回归,其能够输出的信息相对有限。行人检测虽然能输出行人边界框,但是存在大量漏检;人数回归虽然不会漏检,但根本不知道每个行人在哪里。因此,为了克服上面的缺点,寻找更加有效的方法,研究者们提出了人群密度图估计的方式 [4],给每个像素赋予密度值,总和记为场景中的人数 [6]。这样不仅能提高模型的学习能力,也能给管理者带来更多的场景行人分布信息,一举两得。也正是这样的好处,目前人群计数领域中的方法都是以此作为基础进行研究的。

鉴于以上背景,本研究拟使用了一个基于混合真值的 CNN 框架用于人群计数。考虑到高斯核生成的密度图可以提供粗略的人的位置,拟利用高斯人群密度图辅助网络生成最终的人群密度点图来进行最终的人群计数。

1.2 主要研究工作

(1) 复现实践一种混合真值 CNN 人群统计框架。此架构采用 U 形结构生成只有一点监督的点图，将经过高斯平滑之后的真值监督下的密度图作为粗略的人群位置，辅助网络生成点图。

(2) 复现实践一种 ATRM 算法，通过自适应滤波机制捕获像素的 top-k 相互依赖关系，解决人群分布不均匀和规模变化引起的问题。然后提出一种权重注意力机制，使 ATRM 聚焦于 top-k 关系中权重较大的部分。

(3) 将上述算法实践在 3 个大规模公共数据集上。

1.3 论文组织结构

全文内容共六章，具体内容组织如下：

第一章为绪论。

第二章为相关工作。

第三章为 top-k 网络用于人群计数。

第四章为实验。

第五章为参考文献。

第六章总结与展望，总结了本文的主要工作，展望了下一阶段的研究方向。

2 相关工作

本节主要回顾了前期基于密度图、基于点和自注意力机制等进行人群计数的方法。

2.1 基于密度图的方法

基于密度图的人群计数方法使用高斯核平滑后生成的密度图来回归每个像素点的人数。生成密度图真实值最简单的方法是将每个与固定宽度的高斯核进行卷积处理 [8]。后面的研究中, 有人 [9] 根据透视值将每个像素标签的高斯核宽度进行了缩放。但是, 实际运用中透视图经常无法运用, 因此, 有人再次提出了一种几何自适应高斯核的办法, 这种办法的核宽度取决于到最近的带注释邻居的距离, 通过高斯核对每个高斯核标签进行卷积。部分运用实例 [10] 尝试使用高斯核自适应宽度而不是使用手动设置的方式生成高斯核宽度。许多使用密度图的方法 [20] 注重于对人群头部的尺寸变化进行处理, 比如多尺度机制、不同感受出度的多列结构 [22] 等。综上所述, 大多数基于密度图的方法使用了注意力机制进行背景图中的降噪处理。

单图布局如图 2.1 所示。

2.2 基于点的方法

基于点的方法可直接利用标注点作为一种监督信息, 以减轻以往基于密度图的方法中由于高斯平滑而产生的噪声。近年来有人提出 [30] 基于点标注的人群计数贝叶斯函数, 但是这种方法依赖于高斯核。同样, 有学者认为 [12] 可以将人群计数看作是一种分布匹配问题, 其中, 归一化后预测密度图与归一化真值密度图之间的相似性被 OT 来衡量, 但是 OT 在进行计算时是十分耗时的。同样, 有文章提出用排序分布处理不同场景下人头部分布不平衡问题。但是, 以上基于点的方法都忽略了表达人群分布时高斯核平滑密度图的优势和表现多尺度结构在人群计数方面的优势。

横排布局如图 2.2 所示。



图 2.1 单图布局示例

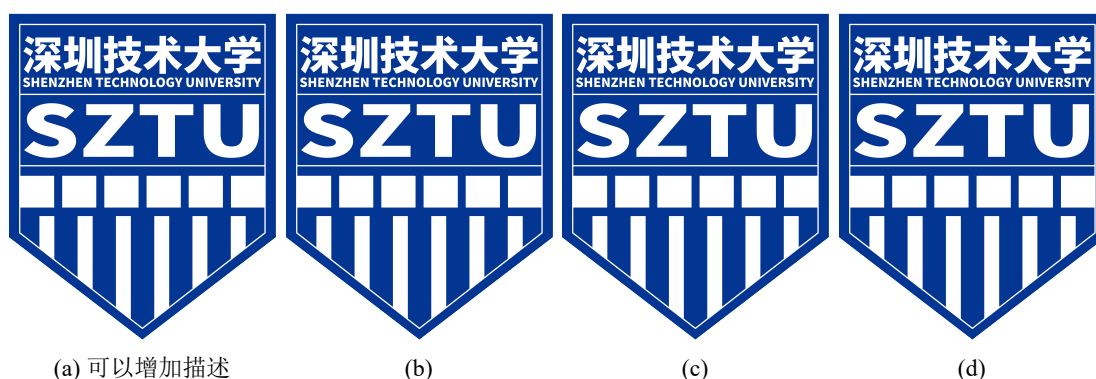


图 2.2 横排布局示例

2.3 自注意力模型

竖排布局如图2.3所示。

神经网络所接收的不同向量之间有不同的关系，自注意力机制实际上是想让机器注意到整个输入中不同部分之间的相关性。对于人群计数，有人提出关系型网络，旨在发现全局和局部自关注像素之间的依赖关系。但是，局部自注意力只对空间上与中间位置高相关的邻居像素起作用。对于全局自注意力机制来说，只对图像经过池化处理得到依赖关系特征图有作用。可见，以上两种机制都会存在噪声问题。



(a)



(b)

图 2.3 竖排布局示例

2.4 竖排多图横排布局



(a)

(b)

图 2.4 竖排多图横排布局

竖排多图横排布局如图 2.4 所示。注意看 (a)、(b) 编号与图关系。

2.5 横排多图竖排布局

横排多图竖排布局如图 2.5 所示。注意看 (a)、(b) 编号与图关系。

2.6 本章小结

本章主要回顾了前期关于人群计数的不同方法。

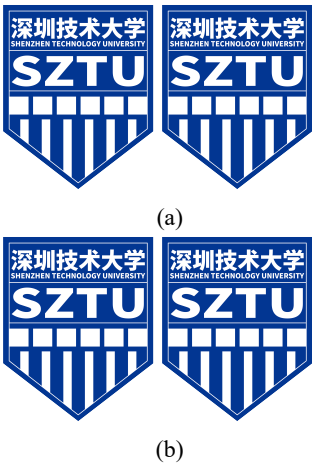


图 2.5 横排多图竖排布局

这里再测试一下不同章节的公式编号

$$p_i = \frac{e^{-\varepsilon_i/kT}}{\sum_{j=1}^M e^{-\varepsilon_j/kT}} \tag{2.1}$$

3 卷积神经网络用于人群计数

这篇文章复现了一种 CNN 框架，框架使用了以下三种模块：ATRM、MDME 混合密度估计器以及 FME（特征提取器）。框架先将输入的图像放在特征提取器 FME 中将图像特征图提取出来。这里的 FME 为归一化后的 VGG16[33] 中的前 13 层。特征图提取完成之后，在 3 个卷积层中提取出多尺度特征。文章运用了一种 ATRM 机制，使用与查询位置相关的前 k 个位置通过自适应过滤机制进行特征增强，即这里利用像素之间的 top-k 关系增强 ATRM 的特征表示。文章中还运用了 MDME 框架，此框架采用 U 形架构并负责将不同层次的多尺度特征进行融合，最终生成估计密度图与最终预测点图。经过以上过程，最终输出包括三个图像：人群注意力图、预测点图和人群估计密度图。

3.1 ATRM

本文运用了一种 ATRM。这个 ATRM 通过使用自适应机制与查询位置相关的 top-k 位置进行特征增强。对于其余位置则不关心。同时，FME 进行卷积运算之后的最后一层生成的特征作为 ATRM 的输入，以此来获得特征增强。以下为 ATRM 实现过程：

(1) 首先将输入图像进行重塑，

表 3.1 学校文件里对表格的要求不是很高，不过按照学术论文的一般规范，表格为三线表。

	A	B	C	D	E
1	212	414	4	23	fgw
2	212	414	v	23	fgw
3	212	414	vfwe	23	长一些的内容
4	212	414	4fwe	23	嗯
5	af2	4vx	4	23	长一些的内容
6	af2	4vx	4	23	fgw
7	212	414	4	23	fgw

表格如表 3.1 所示， \LaTeX 表格技巧很多，这里不再详细介绍。

4 公式插入示例

公式插入示例如公式 (4.1) 所示。

$$\gamma_x = \begin{cases} 0, & \text{if } |x| \leq \delta \\ x, & \text{otherwise} \end{cases} \quad (4.1)$$

5 参考文献插入示例

\LaTeX ^[1] 插入参考文献最方便的方式是使用 `bibliography`^{ENV}^[2]。

大多数出版商的论文页面都会有导出 **bib** 格式参考文献的链接，把每个文献的 **bib** 放入 `thesis-references.bib`^{BIB}，然后用 `bibkey` 即可插入参考文献。

6 总结与展望

纯数字编号

- 1. XXXXXXXXXXXX
- 2. XXXXXXXXXXXX
- 3. XXXXXXXXXXXX

罗马编号

- (i) XXXXXXXXXXXX
- (ii) XXXXXXXXXXXX
- (iii) XXXXXXXXXXXX

括号编号

- (1) XXXXXXXXXXXX
- (2) XXXXXXXXXXXX
- (3) XXXXXXXXXXXX

半括号编号

- 1) XXXXXXXXXXXX
- 2) XXXXXXXXXXXX
- 3) XXXXXXXXXXXX

小字母编号

- a) XXXXXXXXXXXX
- b) XXXXXXXXXXXX
- c) XXXXXXXXXXXX

引用测试, 正如 1、(i)、(1)、1)、a) 所示

6.1 工作展望

手动编号

本课题针对 XX，鉴于 XXX，对 XX 进行了提高，但是 XXX，所以有如下 XX：

（1）目前 XX 虽然 XX，但是 XX 仍然 XX，所以 XX 仍然是一个值得 XX 的问题。

（2）随着 XX，XX 具有 XX 的问题，仍值得进一步 XX。

（3）本课题在 XX 有了 XX，但是 XX 的 XX 还存在 XX，所以 XX。

参考文献

- [1] LAMPORT L. Latex: a document preparation system: user's guide and reference manual[M]. Addison-wesley, 1994.
- [2] PRITCHARD A, et al. Statistical bibliography or bibliometrics[J]. Journal of documentation, 1969, 25(4): 348-349.

致谢

本文能顺利完成，首先最要感谢的是...

为了能把主要精力放在论文撰写上，许多国际期刊和高校都支持 LaTeX 的撰写与提交，新手不需要关心格式问题，只需要按部就班的使用少数符号标签，即可得到符合要求的文档。且在需要全篇格式修改时，更换或修改模板文件，即可直接重新编译为新的样式文档，这对于 word 新手使用 word 的感受来说是不可思议的。

本项目的目的是为了创建一个符合深圳技术大学学位论文撰写规范的 TeX 模板，解决学位论文撰写时格式调整的痛点。