

以形生行——基于图像深度学习的商业空间行为预测方法

金衍孜¹, 沈彦婷², 谢雪莹³, 张琪波⁴, 邵帅⁵, 王煦天⁶, 郑豪⁷, 闫超^{8*}

摘要: 人类行为与建筑空间形态之间存在着本质关联。目前针对建筑空间中的人流行为预测, 主要运用行人运动仿真模型, 但这种方法耗时且难以反映真实环境的行为复杂性。本研究探索了一种基于图像深度学习的建筑空间人流预测方法, 并以商场中庭空间为例, 对其有效性进行实验验证。首先, 通过智能行为感知技术, 实地采集商场中庭空间的人群行为数据, 并通过图像识别技术对人群数量分布进行可视化分析; 进一步, 将人群数量分布热力图和商场平面图进行预处理, 并通过配对构建深度学习数据集; 最后, 通过生成对抗网络 (GANs) 训练模型, 并通过图像生成技术实现在短时间内预测出人流热力图。该方法验证了深度学习在应对人流行为这种高度复杂现象的效果, 可以在建筑设计早期阶段为空间行为推演与优化提供决策依据。

Abstract: There is an intrinsic relationship between human behavior and the morphology of architectural space. Currently, pedestrian behavior prediction in architectural spaces primarily relies on movement simulation models, yet this approach is time-consuming and struggles to capture the complexity of real-world behavior. This study explores a pedestrian behavior prediction method based on image deep learning and validates its effectiveness using a case study of mall atrium spaces. First, intelligent behavior sensing technology was employed to collect on-site crowd behavior data in the mall atrium space, followed by visual analysis of the crowd distribution through image recognition technology. Next, the crowd density heatmaps and mall floor plans were preprocessed and paired to construct a deep learning dataset. Finally, the model was trained using Generative Adversarial Networks (GANs), and pedestrian flow heatmaps were predicted within a short time frame using image generation technology. This method demonstrates the effectiveness of deep learning in addressing the highly complex phenomenon of pedestrian behavior, providing decision-making support for spatial behavior simulation and optimization during the early stages of architectural design.

关键词: 行为计算; 深度学习; 人流预测; 图像识别; 商场中庭

Keywords: Behavioral computation; Deep learning; People flow prediction; Image recognition;; Mall atriums

资助项目情况: 中央高校基本科研业务费专项资金资助 22120240290

引言

1 金衍孜, 同济大学, 硕士生

2 沈彦婷, 同济大学, 硕士生

3 谢雪莹, 同济大学, 本科生

4 张琪波, 同济大学, 本科生

5 邵帅, 同济大学, 本科生

6 王煦天, 同济大学, 本科生

7 郑豪, 香港城市大学, 助理教授

*闫超(✉), 同济大学, 助理教授, yanchao@tongji.edu.cn

人类行为与建筑空间形态具有天然的耦合关系。建筑空间形态会影响其中人群行为的空间分布规律，而人群行为规律的可视化图像结果也会反过来反应建筑空间的形态特征。在这样一种形态与行为互耦的底层逻辑之上，催生出多元化的空间行为研究。其中，在设计初期基于建筑空间形态直接预测出其中的人群行为规律，是面向实践环节的一个关键研究方向。本研究以大型商场建筑的中庭空间为例，探索空间行为的预测方法，成果不仅可以实现更精准地商业空间价值预判，而且可以在设计阶段支撑人因导向的空间形态优化决策。

目前对于人流行为的预测方法主要基于行人运动仿真模型，现有的人群模拟方法一般可以分为基于规则的方法和基于数据的方法[1]。基于规则的方法需要在制定规则方面做大量工作，例如 Helbing 和 Molnár 在 1995 年提出的社会动力学模型[2]，假定智能体在模型环境中行动时前后具有连续性，行为主体的目标和期望速度、空间的物理边界以及其他物体、空间、主体在整体环境中对于行动主体的排斥力和吸引力都会影响主体在环境中的行动模式。而 Lee KH 等人则提出了基于数据的人群模拟方法，使用摄像机从鸟瞰角度记录人群运动，从轨迹中学习智能体模型，并生成一群虚拟智能体进行轨迹模拟。[3]但无论是基于规则的方法还是基于数据的模拟方法，本质上均是依托于行为的抽象机制，难以反映空间行为的真实复杂规律。另外模拟方法在进行人流预测时都较为耗时，难以在设计实践中迅速将预判的行为规律反馈给建筑师。

人工智能、机器学习等技术的发展为空间模拟与预测提供了新的可能性，生成对抗网络 (Generative adversarial networks, GAN) 是一种生成模型和判别模型共同工作训练的模型，该模型可以分析图像中隐含的高维分布数据，并依据这些数据生成图像，实现图像到图像之间有条件的翻译工作。[4]机器学习在城市微观层面的应用主要集中于对于城市空间环境与互联网数据的分析，Sun 等人利用卫星地图和城市 POI 数据训练 pix2pix 模型生成城市中的骑行活动热力图，并研究城市空间环境与 POI 数据对于城市内骑行活动的影响。[5]而在建筑内部层面，机器学习主要应用于物理环境的分析，He 等人提出了一种基于 CNN 和 GAN 的通用平面图采光性能预测模型，用于将平面图直接翻译为光照分布图[6]；Zeng 等人研究了一种基于 GAN 的平面图火灾烟雾与温度分布预测模型，在给定平面图与火源地情况下可以较为精准的预测室内烟雾分布图与温度图[7]。这些研究都表明 GAN 模型可以有效的实现基于图深度学习的空间规律预测工作。

前文提到，现有的人流行为模拟方法不仅难以快速预测空间行为，更难以反映空间行为的真实复杂性。而基于深度学习的方法，虽然在光照分布或烟雾扩散等场景下，已能够替代传统的模拟技术，但其是否能够准确预测人流行为这一复杂现象，目前仍缺乏相关研究。因此，本文探索

了一种基于图深度学习的空间行为预测方法，核心包含两个部分：首先，基于图像识别模型的技术，用于在实际商场中庭空间内对行人的空间分布规律进行可视化分析，从而构建相应的数据集；其次，基于图像生成模型，在该数据集的基础上进行训练，并对空间行为进行预测。通过这一方法，研究能够根据商场中庭的平面图，快速生成商场内部的人流热力图，预测人流活动的分布情况。

1 基于图深度学习的空间行为预测技术框架

1.1 从实地行为观测到虚拟空间预测的技术路径

本研究希望运用图像深度学习的方法对设计的空间形态进行行为预测，因此前提是必须构建一个涵盖空间形态与行为模式的匹配数据集，以支撑预测模型的训练。针对本次研究的大型商场中庭空间，数据集的构建需要从实地商场中庭的行为数据采集开始。具体而言，利用基于图像识别模型的行人追踪技术，对中庭空间中的复杂行人行为进行精准捕捉。通过可视化分析，将行为数据以热力图的形式呈现，从而在图像层面建立起空间形态与行为规律之间的关联数据集，作为后续行为预测模型训练的基础。接下来，在行为预测阶段，基于已构建的数据集，训练一个图像生成模型，以预测设计空间中的人群行为分布。为了确保预测结果的可靠性，最后引入一个验证机制，通过验证集对模型生成的准确性进行验证（图1）。

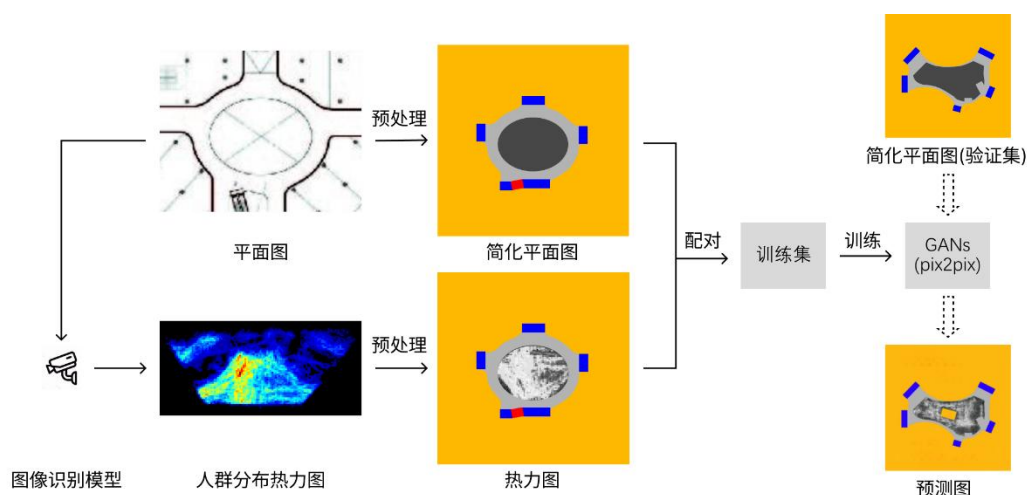


图1 总体框架图（图片来源：作者自绘）

1.2 基于 BodyCAM 自研软件的空间行为数据采集和分析技术

基于图像识别模型的视频追踪技术是行为研究领域的一个重要工具，它能够为研究分析建筑中人的动态行为提供实际数据。同时，相比于佩戴式设备，视频跟踪由于其非侵入性，能够更好地

地观察人在自然状态下的行为行动。基于图像识别技术，本研究运用自主研发的 BodyCAM 软件，通过视频输入开展人群行为的智能分析。

BodyCAM 的基础计算机视觉技术建立在 YOLOX 和 ByteTrack 开源算法的基础上，这两种算法可以根据画面进行物体检测和多物体跟踪。在这两个开源算法的基础上，BodyCAM 能够实现透视视角下将人群行为分析输出为热力图或者轨迹图（图 2（a））。但仅依靠透视视角下的热力分析图或轨迹图，建筑师无法将建筑空间和人群行为紧密联系在一起，并分析建筑空间与人群行为之间的关系，故而 BodyCAM 引入了逆透视变换技术，通过在视频上绘制参考点建立转换矩阵，该参考点建立的转换矩阵为实际空间中工作平面上的矩形，通过该参考矩阵的透视变形关系，将透视视角下的热力分析图或轨迹图转换成正俯视视角（图 2(b)、（c））。正俯视视角的分析图能更好地与平面产生对应关系，将人群行为分布与建筑平面联系在一起，构建一组具有对应关系的数据集，用以支持后续行为预测模型的训练。

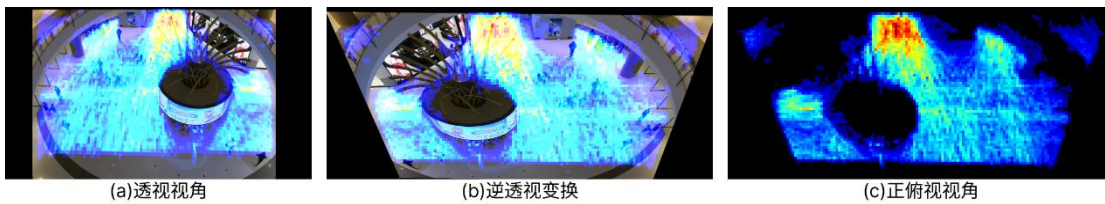


图 2 人群行为分析图(图片来源：作者自绘)

BodyCAM 涵盖多项功能，可以针对多个或单独目标个体进行一系列分析，可以输出人群距离分布热力图、人群占有率热力图、人群密度分布热力图、人群轨迹分布热力图等一系列热力分析图，也可以将同一空间划分为不同区域，分区域进行人群行为分析，输出热力分析图，并进行分区域时空数据采集，计量每个区域内的人群活动数量与时长。本研究希望探索商场中庭中的空间形态对于中庭内部人流活动的影响，因此选择人群分布数量热力图作为热力分析图像，人群密度分布图是以人数为计量的热力分布图，BodyCAM 可以将镜头视野内空间划分为多个区域，识别镜头视野内的多个智能体并标记，每标记一个智能体即在对应区域内计数，在逐帧分析后，将分析数据可视化输出为热力图，停留人数越多的区域颜色越红，停留人数越少的区域则越蓝（图 3）。

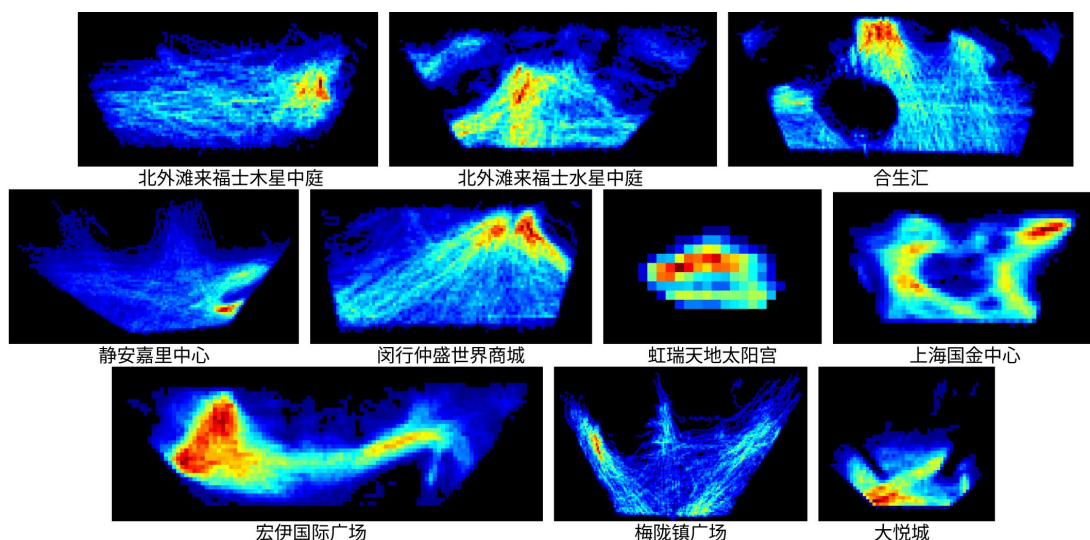


图3 人群密度分布热力图（图片来源：作者自绘）

2.3 基于 pix2pixHD 框架的空间行为热力图生成预测技术

基于商场中庭平面图预测内部人群分布是一个图对图的翻译任务，GAN 模型能很好应对该类输入端与输出端都为图像的翻译任务，在现有的空间相关的预测中也较为常用。而 Pix2pixHD 是基于条件对抗生成网络（conditional GANs, CGAN）的一种用于图像到图像翻译问题的通用框架，能够很好的解决有条件下的图像翻译任务[8]，而根据商场中庭的平面几何信息预测商场中庭内部的人群分布数量热力图就可以理解为一种有条件的图像翻译问题。

Pix2pixHD 模型包含生成器（Generator, G）和判别器（Discriminator, D）两个神经网络，输入图像后，生成器会使用卷积层、残差层和反卷积层将输入图像转换为相同大小的输出图像[5]，判别器则将生成器生成的图像与真实图像比较，判断真假，并将损失和优化梯度反馈给生成器，优化生成器，学习数据集之间的配对关系，使生成图像更加接近于真实情况。也就是说，pix2pix 可以通过训练，建立起输入图像与输出图像之间的隐含联系，当再给训练后的模型输入同一类型的图像时，能够直接生成预测出与之匹配的输出图像。在本研究中，将真实商场中庭平面的转译图作为输入数据集，对应商场中庭的人群分布数量热力图为输出数据集，训练模型，建立数据对之间的配对规律，使新的商场平面图输入模型后，能够立即预测出对应的人群分布数量热力图。

2 基于图像识别模型的空间行为数据集构建

针对空间行为预测所需的“空间-行为”数据集，空间部分需要基于真实平面图进行抽象化处理，行为部分则需要进行实地行为数据采集，并进行可视化热力图分析。

2.1 实地行为采集安排

本研究主要使用 BodyCAM 软件进行人群活动识别并转换为人群分布数量热力图。其中，需要先对实地商场中庭内部的人流活动进行固定视角的视频采集，再使用 BodyCAM 进行分析转换输出为人群分布数量热力图。

商场中庭内部与周边的各类要素，如商场中庭公共空间的边界形状、中庭上方架空层的形状、中庭周边的通道数量和连接中庭的扶梯数量与位置等，都会对中庭内部的人流活动强度与密度产生影响。本研究在尺度体量尽量接近的情况下选择了上海市内十个形状各不相同的商场采集数据。商场内人流活动的强度和密度也会受到时间因素的影响，周末商场内的人流强度会显著强于工作日，为减少时间因素影响，故而本次研究将数据采集的时间确定为周末同一时间段内，并将该短时间内的人流活动全部通过 BodyCAM 分析输出为人群分布数量热力图。

2.2 空间平面图抽象处理

本研究收集了需要实地采集人流活动数据的十个商场中庭的平面图数据，但各个商场的尺度规模、平面图形式等存在较大的差异，为了统一数据集的形式与尺寸，需要对平面图进行预处理（图 4）。

首先，确定全部数据集图像大小为 1024*1024，将所有数据集中实际尺寸最大的中庭裁剪为 1024*1024，并将该图像的比例确定为参考比例，缩放所有数据集为该比例，并放置到 1024*1024 的图像中间。其次，上文提到商场中庭公共空间的边界形状、中庭上方架空层的形状、中庭周边的通道数量和连接中庭的扶梯数量与位置等元素都会有影响中庭中的人流活动，在简化处理平面图时，将这些会影响中庭人流活动的元素提取出来，并使用不同颜色的色块表示。其中，橙色表示商场中庭周围的商铺边界和墙等人流无法进入的区域，红色表示电梯等竖向交通连接的人流出入口，蓝色表示中庭周边的通道。同时，为了在二维图像上表达三维的中庭上方架空高度信息，将中庭上方的架空高度定义为不同的灰度值，颜色越深则中庭上方的架空高度越高，反之则越浅，即图上的浅灰色为上方有覆盖的区域，深灰色为上方无覆盖的区域。需要注意的是，通道和交通空间出入口实际上在空间中并没有厚度，但为了方便理解与表达，在平面图中将这两个元素抽象为色块，且在所有数据集中，通道和交通空间出入口所代表的色块的厚度一致。

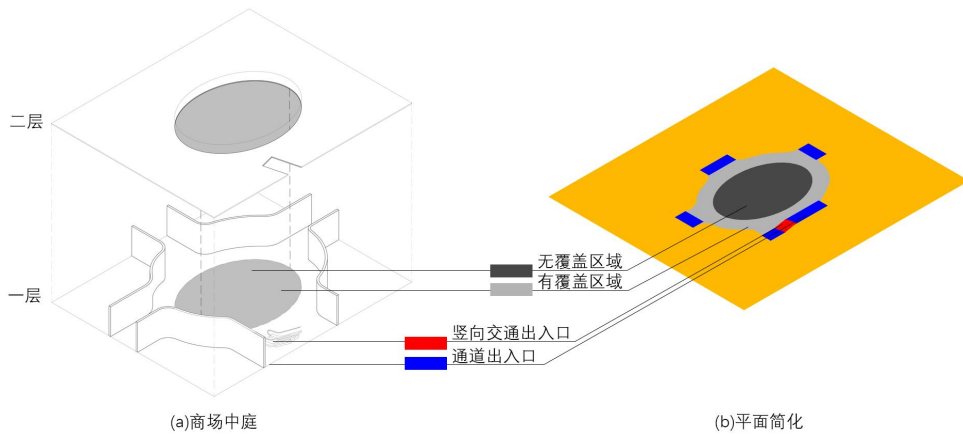


图 4 平面图处理（图片来源：作者自绘）

2.3 空间与行为的图像映射匹配

基于 pix2pix 的训练特点，两个训练集之间需要一一对应，且两个训练集尺寸需要一致，故而将人流热力图的尺寸也确定为 1024×1024 （图 5）。本研究将预测对象限定为中庭通高空间中的人群行为规律，因此需要将 BodyCAM 生成的人流行为热力图映射到商场的空间布局中，并根据中庭通高空间的轮廓，对这些热力图进行裁剪，仅保留内部的人流分布图像。

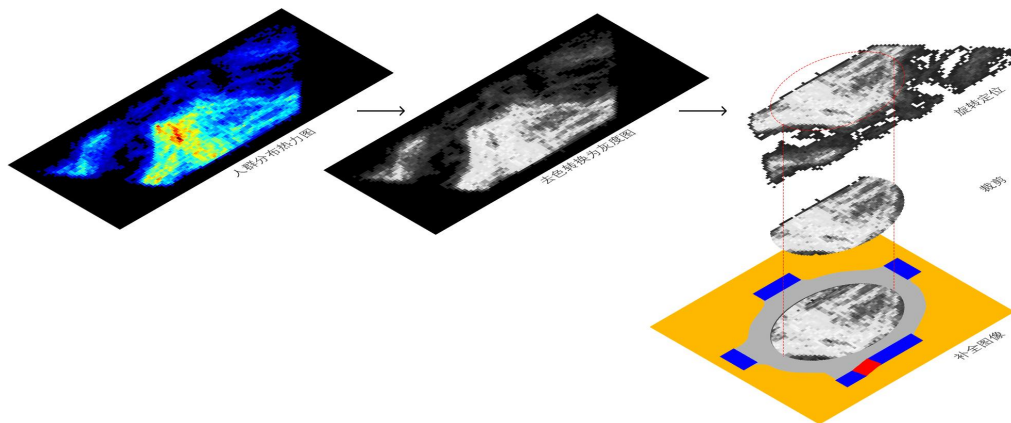


图 5 人流热力图处理（图片来源：作者自绘）

将实际采集的两小时商场内部视频输入 BodyCAM 软件，通过软件分析导出以时间为计量的人群占有率热力分析图。将 BodyCAM 分析生成的热力图受到输入的视频视角和画幅的限制，无法覆盖整个商场中庭区域，只能生成视频拍摄镜头内区域的热力图，即生成的人流热力图无法覆盖整个中庭平面，部分区域没有实际热力图，需要将该部分区域根据实际规律进行规则化填充。

3 基于图像识别模型的空间行为数据集构建

由于在实地空间中进行视频数据采集和行为规律分析会耗费大量的时间，为了确保本研究提出的方法具有实践可行性和可推广性，因此重点探索了基于小样本模型的训练和预测实验。研究共构建了九组数据集，其中八组用于模型训练，剩余一组则用于测试和验证训练模型的预测准确性（图6）。为了可视化模型训练的过程，每当模型训练2000代，就由生成器预测生成一张热力图，并将生成的图片绘制在图中。在模型训练初期，热力图和周边图像的边界都较为模糊，且在内部的预测热力图人流分布较为不规律，在6000代之后，随着训练误差的减小，周边简化平面图部分边缘变得清晰，内部热力图部分也开始呈现出规律，到20000代以后，整体预测热力图图像几乎保持不变，随着训练整体变得更加稳定（图7）。

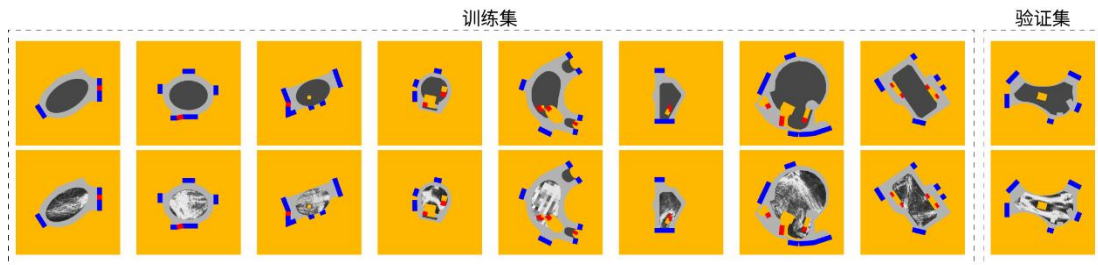


图6 数据集（图片来源：作者自绘）

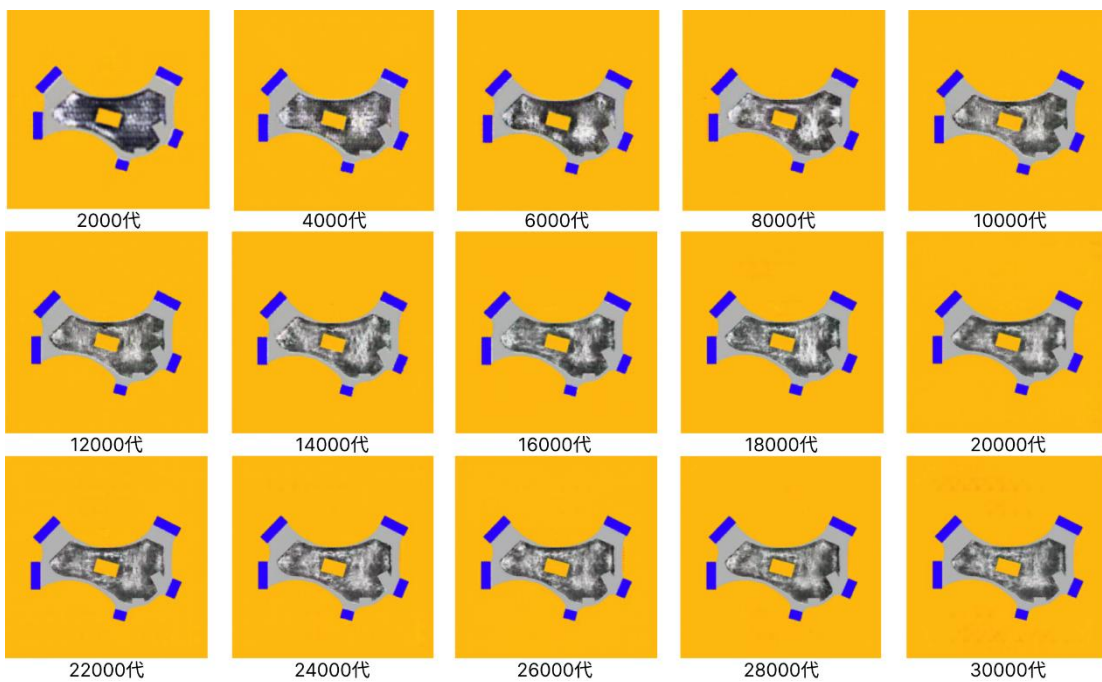


图7 训练历程（作者自绘）

进一步，将验证集数据投入预训练的模型生成预测热力图（图8），将预测热力图与真实图像比较，预测热力图可以判断出中庭内部基本的人流走向与密度分布，人流收到中庭左右两侧通道的影响，人流在两侧接近通道处较为密集。同时，也能够预测出中庭内部陈设的构筑物对于人流走向的影响，预测人流密度图像真实人流密度图一样绕过了中间构筑物，在中构筑物周边颜色较深，两侧颜色较浅。

尽管本研究所使用的数据集规模有限，但在训练之后，模型仍然可以在宏观趋势层面预测出中庭内部的人群分布数量热力图，足以验证利用 pix2pix (GAN) 框架训练一种图对图的人流预测模型方法的可行性，且相较于以规则或数据为基础的仿真预测模型，利用 pix2pix (GAN) 为框架的人流预测模型能够在短时间内预测出复杂的人流分布状况。因此，其作为设计初期的决策支持工具，在理解建筑内部空间的行为性能时更具优势，能够基于其预测结果快速分析判断平面对应的人流分布状况，并调整建筑内部的布局。

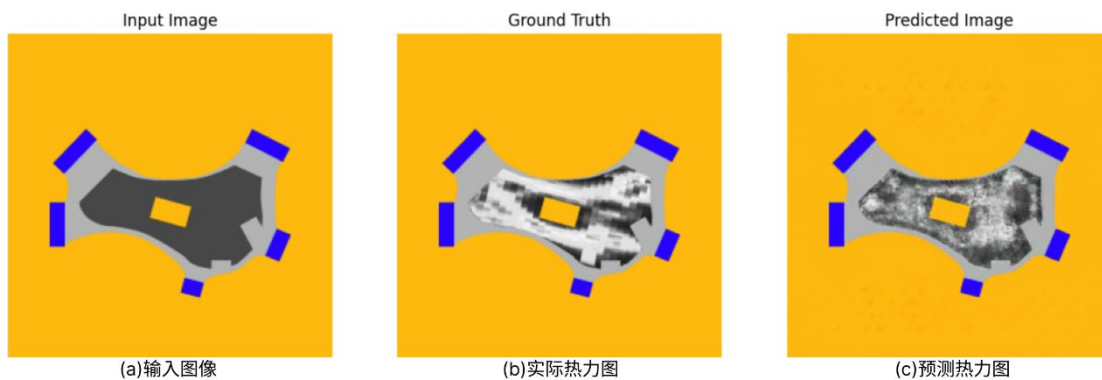


图8 测试集结果（图片来源：作者自绘）

诚然，本研究所提出的技术路径仍存在一些局限性，需要在未来的研究中加以解决。一方面，在实地采集人群行为数据时，受到摄像头视角的限制，单一视角并不能涵盖全部的中庭区域，故而需要补全一部分图像，对于图像的精准度造成了一定影响。因此，需要在进一步研究中探索多摄像头耦合的视频数据采集方法。另一方面，为了确保该方法在实际应用中的可推广性，研究中的数据规模被限制在较小范围内。这种小样本数据集虽然有助于提高训练效率，但也对预测结果的精细度产生了一定影响。目前使用的生成式对抗网络 (GAN) 模型在处理精细化的行为预测细

节时还存在局限。因此，未来的研究可以考虑引入更复杂的算法，比如图卷积神经网络（GCN），结合注意力模型来提升预测精度。

结语

本文通过整合图像识别和图像生成这两种深度学习模型，提出了一种新的空间行为预测方法，并通过商场中庭空间的人流分布预测实验，初步验证了方法的有效性。训练得到的模型可以在不需要复杂的输入数据的情况下，依靠简化的平面图图像生成对应的人流分布热力图。因此，相较于传统的人流模拟方法，本研究提出的基于深度学习的预测方法在预测效率和应对复杂度层面都具有显著优势，可以适用于建筑设计的早期阶段，讨论中庭的边界形态、通道数量与位置、扶梯出入口数量与位置以及中庭上方覆盖的形态对于中庭内部人流分布的影响，以更好地分布商场中庭周边的商业业态和设施，同时也可以从行人的行动视角出发，反向推演商场中庭空间的形态以及内部构筑物的布局。从宏观层面，这一研究验证了机器学习方法在应对人流行为这种高度复杂现象的效果，不仅为人工智能在建筑领域中的应用开辟了新的可能性，也为扩展空间行为研究方法提供了全新的视角。

参考文献

- [1]Lin, X., Liang, Y., Zhang, Y. et al. IE-GAN: a data-driven crowd simulation method via generative adversarial networks. *Multimed Tools Appl* 83, 45207–45240 (2024). <https://doi.org/10.1007/s11042-023-17346-x>
- [2]Helbing D, Molnár P (1995) Social force model for pedestrian dynamics. *Phys Rev E* 51(5):4282–4286. <https://doi.org/10.1103/PhysRevE.51.4282>
- [3]Lee KH, Choi MG, Hong Q, et al (2007) Group behavior from video: a data-driven approach to crowd simulation. In: *Proceedings of the 2007 ACM SIGGRAPH/Eurographics symposium on computer animation*. Eurographics Association, Goslar, DEU, SCA '07, pp 109–118, <https://doi.org/10.1145/1272690.1272706>
- [4]A. Creswell, T. White, V. Dumoulin, et al., Generative adversarial networks: an overview[J], *IEEE Signal Process. Mag.* 35 (1) (2017) 53–65, <https://doi.org/10.1109/MSP.2017.2765202>.
- [5]Sun, Yunjuan, Jonathan A. Scelsa, et al. “A Machine Learning Method of Predicting Behavior Vitality Using Open Source Data.” *Proceedings of the 40th Annual Conference of the Association of Computer Aided Design in Architecture (ACADIA)* (2020): n. pag.
- [6]He, Qiushi, Ziwei Li, Wen Gao, et al. “Predictive models for daylight performance of general floorplans based on CNN and GAN: A proof-of-concept study.” *Building and Environment* (2021): n. pag.
- [7]Zeng, Yanfu, Yizhou Li, Peilun Du, et al “Smart fire detection analysis in complex building floorplans powered by GAN.” *Journal of Building Engineering* (2023): n. pag.
- [8]Isola, P., J. Zhu, T. Zhou, et al. 2017. “Image-to-Image Translation with Conditional Adversarial Networks.” *arXiv preprint*. <https://arxiv.org/abs/1611.07004>.