

# 基于街景图像的城市街道观山视觉感知研究 ——以泰安市中心城区为例

崔旭辉<sup>1</sup>, 李莎<sup>1\*</sup>, 董昭舍<sup>2</sup>

## 摘要:

随着城市的快速发展与扩张, 建设项目体量增大, 建筑高度不断提升等因素致使观山视廊受到遮挡, 人群观山体验不佳。山体景观的视觉感知应基于人在公共空间的实际需求, 然而对于自然山体与人群视觉关注机制之间, 尚未建立基于真实视野的观山视景空间体系。故本文从人本角度出发满足个体的视觉感知需求对于优化人群观山体验, 完善空间视觉体系, 以泰安市中心城区为例, 使用城市景观街景数据集, 训练高准确率 SegFormer-B2 图像语义分割模型计算城市街道景观的要素指标, 结合批量问卷数据分析得出城市观山满意度与街景山体占比、绿视率呈显著正相关性, 且与水平街道要素有关, 与部分城市基础设施、可移动的车辆等呈明显的负相关。此外, 研究还基于梯度加权类激活映射图来直观地解释神经网络关注图像的特征层, 进而更好地发掘神经网络对街道山景图像的分类特征识别。本文通过深度学习和可解释性分析算法, 深入探讨了街道要素指标与观山满意度之间的关联性。研究结果有助于进一步了解神经网络在街景图像中的视觉感知特征, 为彰显城市特色、提升视觉品质提供策略启示。

**关键词:** 街景图像; 城市街道; 山景; 视觉感知; 梯度加权类激活映射; 深度学习

**Keywords:** Street View Images; Urban Street; Mountain View; Visual Perception; Grad-CAM; Deep Learning

**资助项目情况:** 本项目得到江苏省研究生科研与实践创新计划资助 (KYCX24\_2978); 中国矿业大学研究生创新计划项目资助 (2024WLJRCZL309)

## 引言

近几十年来, 城市的迅速发展扩张导致建设项目规模的扩大和建筑高度的增加, 这些因素使得观山视廊遭到遮挡, 人群的观山体验因此降低。作为城市空间架构的重要生态载体与自然要素, 城市山体也在一定程度上能够决定城市空间结构、布局及城市景观风貌[1]。然而, 随着城市进程的加快, 导致部分景色优美适宜居住的山地景观城市面临着同质化的风险[2]。对于一些丘陵观山城市, 对城市进行宏观建模观察其视域范围是一种较为科学的研究方法[3], 然而过往对于自然山体与人群视觉关注机制之间, 并未建立基于真实视野的观山视景的空间体系, 山体景观的视觉感知应是基于人在公共空间对其的实际需求, 这也强调了需要充分满足人群对山体景观资源的多重视觉诉求[4]。

视觉感知是公共感知的最重要的方式之一[5], 视觉的评价与感知现今也成为了人们了解街道景观

<sup>1</sup> 崔旭辉, 中国矿业大学建筑与设计学院, 硕士研究生, ts22190075a31tm@cumt.edu.cn

<sup>1</sup> 李莎 (✉), 中国矿业大学建筑与设计学院, 副教授, 5979@cumt.edu.cn

<sup>2</sup> 董昭舍, 中国市政工程中南设计研究总院有限公司, 硕士, 2298228689@qq.com

质量的支柱[6]。街景图像具有多数据的来源，能呈现人本视角的城市自然景观和人工场景[7]，为我们提供一种以人为中心的研究视角。如今，已经有丰富的街景数据运用于城市研究之中[8]，例如感知城市可玩性及感知犯罪活动[9, 10]，利用街景图像测度街道空间品质水平[11]，识别环境安全内的视觉因素等[12]。此外，关于街景图像的研究之中，语义分割模型是提取视觉要素的手段[13]，然而大多数研究之中均未涉及到观山这种特殊的场景[14]，借助街景数据开展城市景观研究在我国仍有很大的发展空间。

本文以泰安市中心城区为例，利用区域内丰富的街景数据，通过深度学习语义分割模型等技术探讨观山视觉感知与客观街景要素之间的关系，利用主观预测模型进一步了解神经网络观察图像的视觉特征，为提升人群观山满意程度，改善山景街道风貌提供借鉴和参考。

## 1 研究方法

### 1.1 研究流程框架

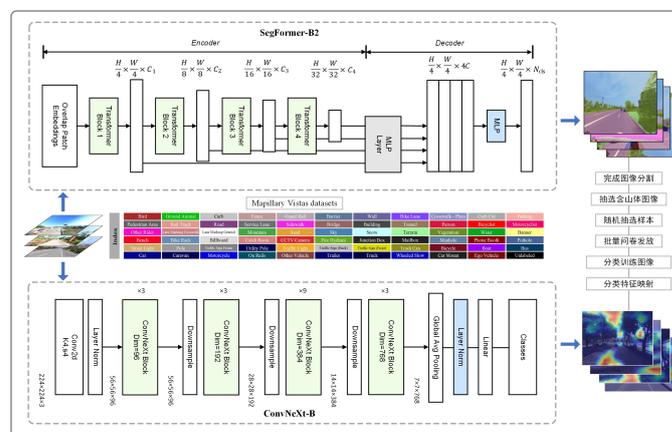


图1 研究框架（图片来源：作者自绘）

本文利用 SegFormer-B2 语义分割模型基于 The Mapillary Vistas Datasets 数据集进行了模型训练和性能评估，并完成了对于研究区域内百度街景图像的分割过程，获取了客观的城市街景要素比例，利用调查问卷收集了人群对于不同山景图像的满意程度数据。随后，对于调查打分后的问卷进行整理和分类，并基于 ConvNeXt-B 模型进行主观预测模型训练，最终通过 Grad-CAM 分类特征映射可视化帮助理解模型观察图像的决定性因素。

### 1.2 研究区域

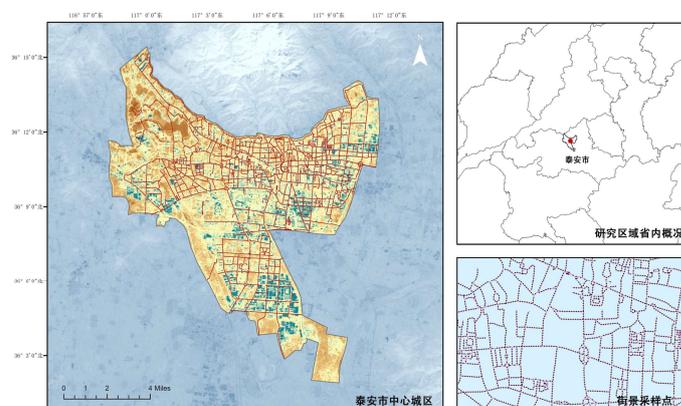


图2 研究区域（图片来源：作者自绘）

本文的研究区域覆盖了泰安市中心城区(面积为 207.7km<sup>2</sup>), 人口约为 135 万, 北部毗邻泰山风景名胜, 如(图 2)所示。泰山作为中国名山, 于 1987 年被列入世界自然文化遗产名录, 被称为“五岳之首”, 享有“天下第一山”的称号, 研究学者一直关注于其景观美学的价值意义[2]。该研究区域不仅邻近名胜风景区, 具有良好的城市自然山体景观, 吸引了国内外大量游客, 并且承担了大部分城市居民的生活和生产活动, 具有完善的公共设施和充足的城市图像数据, 因此我们选择泰安市中心城区作为山景视觉感知的研究区域。

### 1.3 街景要素的提取及指标选取

本文选择 SegFormer-B2 语义分割模型用于针对山景要素的分割提取[15]。文中的街景语义分割模型的训练借助了 The Mapillary Vistas Datasets, 包含了 25000 张高分辨率图像, 其中含有来自不同天气条件、不同成像设备的街道风景图像, 共有 66 个目标注释标签, 例如山、建筑、车辆[16]。模型使用了 Pascal VOC 数据集的预训练权重, 输入图像尺寸为 512\*512 像素, 共 300 个训练轮次。模型性能的评估过程使用了四个评价指标, 采用了平均交并比, 是计算预测像素与真实像素交集与并集的比例的平均值, 二是召回率, 评估模型正确识别的实例占有相关实例的比例, 以及平均精确度和准确率, 分别代表着计算预测为正的样本中实际为正的的比例和预测正确数据占总数据的比例。街景要素的视觉指数是关于一个语义对象在图像总像素中的面积占比, 每个街景要素的占比量计算如下:

$$VI_{obj} = \frac{Area_{obj}}{Area_{total}} \times 100\% \quad (1)$$

其中,  $Area_{obj}$ 是在分割过程中获得的标签对象, 比如山体、天空的像素数,  $Area_{total}$ 则为一张图像的像素总量。

文中结合街景数据集的标签索引, 使用十类要素作为关键分析的街道指标。其中, 包含以山体、水体等自然要素, 与绿视率相关的植被要素, 能表现天空开阔程度的天空要素及水平视觉广度的马路等, 此类要素的增加意味着场景中具有较少的视觉障碍并提供更宽阔的视觉感受[17]。以及各类基础设施建设及街道上存在的标识、广告牌等[18], 能表现城市活力与社交环境活跃程度的车辆、行人要素。指标的选取覆盖了从自然景观到人造结构的各种视觉要素, 它们可能直接影响城市居民在街道场景下山景视觉的体验感受。

### 1.4 主观视觉感知的提取

用于主观调查问卷中使用的主要采样数据库为 OpenStreetMap (OSM) 开放街道地图和百度街景。获取 OSM 路网文件后, 使用 ArcMap10.7 彼此相距 50 米的地理距离, 获取每个采样点的经度纬度并存入 CSV 文件中。之后, 使用百度街景的 API 获取图像。由于街景采集车辆的方向不固定, 其图像中含有山体的图像也随之变化, 因此我们在每个采样点分别采集了 (0°、90°、180°、270°) 四个方向的街景图像, 共计获取 43680 张街景图像 (每张图像尺寸为 1024 \* 700 像素)。

文中的主观感知调查利用了在线问卷平台, 将含有山体的街景图像提供给志愿者进行感知印象评估。利用 Python 程序随机选取了含有山体的街景图像, 每组调查问卷包含 25 张不同的山景图像, 共 20 组问卷用于山景感知分数的调查, 参与者使用李克特量表根据对每张山景图像的满意程度对它们进行从 1 (低) 到 5 (高) 的打分。调查中的图像数据根据其平均得分标准差来划分满意度等级, 本研究共收获了 534 份问卷样本。

### 1.5 主观预测模型的可视化

对于主观预测模型的可视化阶段, 问卷中收集到的图像用于模型的数据输入, 基于问卷结果中不

同满意度等级的街景图像，80%划分为训练集，20%划分为验证集，训练轮数为 300，以此用于进行分类模型的训练。其中，主观预测模型选择了以卷积神经网络为主体的模型 ConvNeXt-B[19]，该模型在分类测试性能方面表现良好。

在山景视觉感知的研究中，利用问卷调查中得到的视觉感知图像进行图像满意度主观预测模型的训练，主观预测模型模拟了人对山景图像满意程度的感知重点。在此特别的是，利用 CAM 可以展示可视化深度学习模型关注图像的部分，像我们展示模型关注的区域，帮助我们理解影响模型分类的决定性因素[13]。基于最终的分类模型，我们得到了结果中的特征语义信息，并且使用每个特征映射通道的权值进行 Grad-CAM 可视化展示，通过模型结构中不同的解释信息来探讨模型感知的共性和差异。

## 2 研究结果

### 2.1 街景要素空间分布

最终，SegFormer-B2 模型的平均交并比达到 0.4022、召回率达到 0.4848，平均精确度和准确率分别为 0.6366 和 0.8940，已经拥有了有关于此数据集训练的高性能表现。随后，我们基于 SegFormer-B2 对 43680 张百度街景图像进行语义分割并提取各个街景要素指标，计算了山景感知实验中客观视觉要素的空间分布。如（图 3）所示，利用自然断点法绘制了山景视觉指数 Mountain view index (MVI)、建设视觉指数 Structure view index (StruVI)、天空视觉指数 Sky view index (SVI)、绿色视觉指数 Green view index (GVI)、水平视觉指数 Flat view index (FVI)、障碍视觉指数 Barrier view index (BVI)、水景视觉指数 Water view index (WVI)、车辆视觉指数 Vehicle view index (VVI)、物体视觉指数 Object view index (OVI)和人类视觉指数 Human view index (HVI)在中心城区范围内的空间分布规律，并计算其平均值分别为 0.000857、0.100984、0.329137、0.234099、0.212988、0.044794、0.000313、0.023918、0.040691 和 0.001275。除 MVI 和 WVI 具有较强的空间异质性外，如（图 3-b），其中以 StruVI 为例，在中心城区的核心区域集中度最高，向四周分散降低，这与中心区域建设的集中度有关。相对来说，高密度的建设降低了区域内的天空开阔度，SVI 在居住的集中区域较低，随后向周边扩散增加，如（图 3-c）。除核心区域部分历史城区保护区域非常密集外，GVI 与 SVI 分布情况相似（图 3-d）。如（图 3-e, 3-i），FVI, OVI 等城市基础设施在整个中心城区分布均匀，核心区域与 StruVI 相似，证明高密度的纵向建设也伴随着城市横向基础设施的发展，如（图 3-h, 3-j），对于 VVI 和 HVI 来说，在中心城区的核心居住区域呈现聚集的现象，说明该区域内人类行为活跃程度较高。

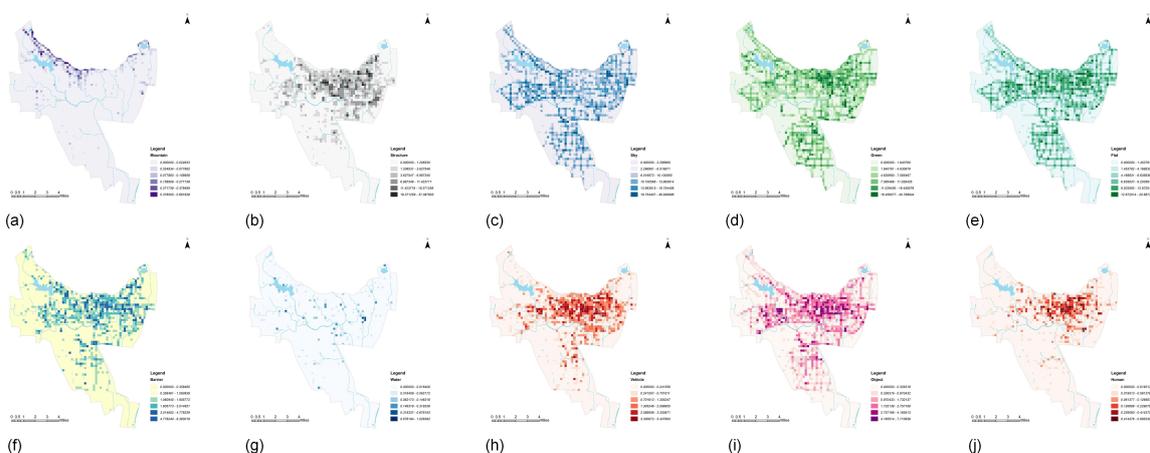


图 3 街景要素空间分布 (a) MVI; (b) StruVI; (c) SVI; (d) GVI; (e) FVI; (f) BVI; (g) WVI; (h) VVI; (i) OVI; (j) HVI (图片来源: 作者自绘)

## 2.2 问卷结果分析

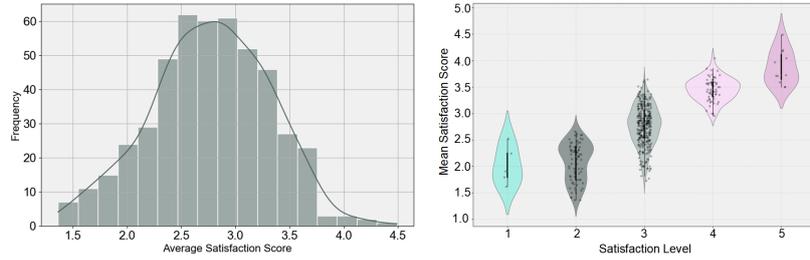


图 4 问卷正态直方图及小提琴分布 (图片来源: 作者自绘)

如(图 4)所示,从问卷收集到的主观结果得知,山景感知图像的平均得分为 2.77,数据符合正态分布的趋势(标准差为 0.54),利用标准差将收集到的图像分为了 5 个等级。随后,文中使用 IBM SPSS 27.0 进行了 Pearson 双变量相关性分析,来验证街景客观要素指标与观山满意度 Average Satisfaction Score (ASS) 和满意度等级 Satisfaction Level (SL) 之间的关系。基于双变量显著性检验的相关性系数如(表 1)所示,其中,观山满意度与 5 个视觉指数呈显著的正相关,包含 MVI、SVI、GVI、FVI、WVI,与 StruVI、BVI、VVI、OVI 和 HVI 呈负相关关系,观山满意度等级结果与平均满意度分数相似。由(图 5)可知,StruVI 的增加显著影响了 SVI 和 GVI,表明更多的纵向建筑会减少天空的开阔度,并且会影响绿视率的占比。这个影响也是相互的,但建筑增加对于 ASS 会有显著的负面影响,而当 SVI 和 GVI 上升时,会对 ASS 产生积极的推动作用。对于 FVI 和 BVI 之间的相对关系,BVI 意味着纵向障碍的增加,例如护栏、篱笆等,这与水平要素为主的 FVI 呈负相关关系。而 FVI 和 SVI 增加,表明天空开阔度和路面的开阔度增高,这在一定程度上会提升人们对山景图像的满意程度。VVI、HVI 此类包含移动元素的指数增加,如各种车辆、骑行者会对满意程度产生负向作用。此外,部分城市基础设施例如电线杆、标识等,这些物体的比例上升也会降低满意程度。

表 1. 观山满意度与街景要素相关性系数 (作者自绘)

$r$	MVI	StruVI	SVI	GVI	FVI	BVI	WVI	VVI	OVI	HVI
ASS	0.389**	-0.501**	0.153**	0.324**	0.138**	-0.128**	0.133**	-0.269**	-0.105*	-0.140**
SL	0.381**	-0.460**	0.145**	0.273**	0.118**	-0.119**	0.164**	-0.166**	-0.082	-0.136**

\*\* 在 0.01 级别(双尾),相关性显著, \*\* 在 0.05 级别(双尾),相关性显著。

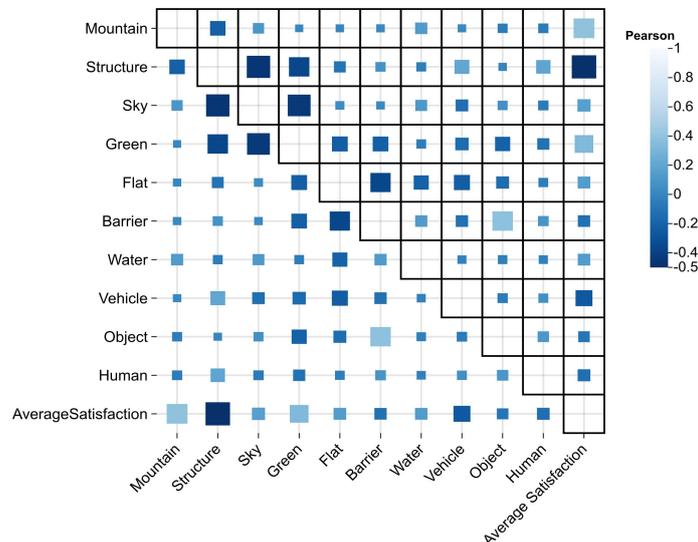


图 5 各要素相关性分析 (图片来源: 作者自绘)

### 2.3 预测模型结果

由（图 4）可知，问卷中表现出的打分情况趋于三分类水平，相对极端分数而言，人们更倾向于选择较为中庸的分数。由于数据集的不均衡，本文将训练图像整合为三分类数据集进行分类模型训练。主观预测模型基于 ConvNeXt-B 神经网络进行分类训练，在模型性能表现中达到较高的水平，其中，准确率达到 0.7045，精确度 0.7522，召回率 0.7045，F1 Score 达到 0.7140。对于模型的预测结果的图像表现而言，如（图 6）所示，我们选择了三个不同等级较有代表性的图像，输入主观模型进行注意力可视化的表达，左侧两列为低等级，中间两列为中等，右侧两列是最高等级。从模型决策分类的映射图之中可以看出，低等级的如 1C、2C，模型更关注与建筑和墙体，而高等级的 5C、6C，模型对远处的山体、天空具有更高的注意力。结果表明，低等级的分类图像中，其中建筑、墙体等占比高，模型判断的注意力更偏向于人工的、生硬的建设类的要素。相反，模型对于高等级的图像，更多关注于广阔的、自然的要素，如天空、植物等等。其次，对于不同模块输出的可视化情况，我们可以纵向地观察到前两层模块更偏向于较低级别特征，比如说点状的纹理、边缘信息，而更深入的表达则是判断需要的高级特征，如 C、D 两行图像，映射图的热点明显且突出，更关注有助于满意度提升的自然要素。

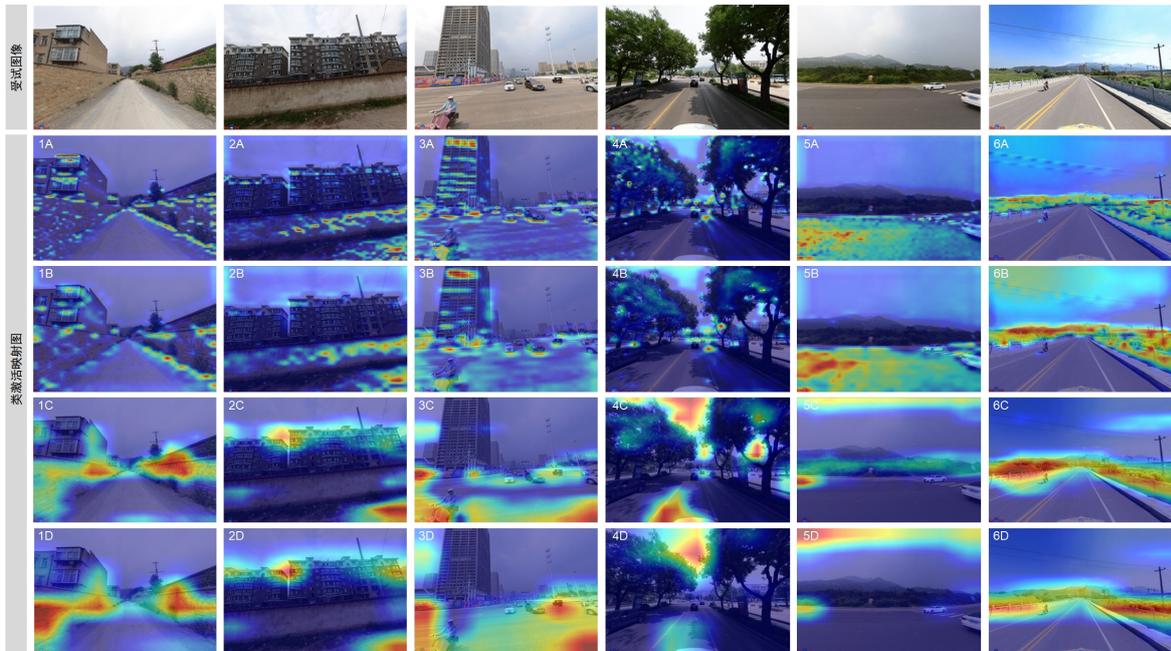


图 6 主观预测模型不同模块可视化结果（图片来源：作者自绘）

### 结语

本文以泰安市中心城区为例，通过对语义分割模型的训练和街景图像要素指标的提取，获取了各街道视觉要素的空间分布信息，结合主观问卷信息，总结了观山满意程度与各街道视觉要素的相关性关系，最后通过主观预测模型的训练和类激活映射图像的可视化，进一步探讨了神经网络观察山景图像的规律。研究表明，观山满意度与山体占比（0.389）、天空开阔度（0.153）、绿视率（0.324）、马路的水平开阔度（0.138）呈正相关，这些指数的提升对观山的满意程度有积极影响。关于纵向的建设要素（-0.501）、城市基础设施等则会起到负面影响，场景内可移动的、出现频次不固定的车辆（-0.269）、人群（-0.140）等存在一定的负相关。最后，本文尝试利用问卷信

息提取分类数据进行主观预测模型的训练，模型准确度达到了 0.7045，类激活梯度加权可视化的结果也提供了研究中分类权重的决策意向。

通过对研究结果及指标数据的分析，本文揭示了不同要素对居民观山满意度产生的影响，旨在实践中为城市规划和设计师提供参考，为街道的改造策略提供新思路。其中，山景指数和绿视率是提升视觉感知体验的关键，建议在后期开发建设过程中限制部分区域的建筑高度，调整结构布局，进一步保护观山视廊。对于建设密集街道区域，通过调整修缮绿化带，增加一定的公共绿地及绿色基础设施，在提升视觉美感的同时增加居民的幸福感和亲近自然的机会。同时，提升道路的开阔程度是改善山景视觉感受的有效途径，减少街道的视觉障碍物，如废弃的广告牌和围栏等等，增加视觉的通透性。此外，针对部分观山游线的街道区域，实施步行优先的设计策略，缓解车辆密度，减轻交通流对视觉的干扰，进一步提升观山体验感受。

最后，在未来我们将会收集更丰富的数据信息，加大主观感知问卷的样本量，增加预测模型的样本数据，以获得更高的准确率和可视化水平。本研究也有助于为观山游线的街道改造策略、基础设施的建设和维护提供部分借鉴意义，更好地捕捉其不同城市下环境视觉体验的多样性与复杂性，制定更加具有个性化、场景化的城市规划策略。

## 参考文献

- [1] An Q, Xiao H, Guo Y, et al. Landscape Visual Aesthetics Measurement, Assessment, and Improvement of Street Pedestrian Spaces in Mountainous Cities - Case Study on the Historic Downtown of Jinan, Shandong Province[J]. *Landscape Architecture Frontiers*, 2022,10(1):9.
- [2] Yu Y, Liu B, Ma L, et al. 'Thousand Years of Charm': Exploring the Aesthetic Characteristics of the Mount Tai Landscape from the Cross-Textual Perspective[J]. *Land (Basel)*, 2023,12(12):2129.
- [3] Zhang G, Cheng S, Gao Y. A visibility-based approach to manage the vertical urban development and maintain visual sustainability of urban mountain landscapes: A case of Mufu Mountain in Nanjing, China[J]. *Environment and Planning B: Urban Analytics and City Science*, 2024,51(2):384-400.
- [4] 成实, 苏子玥, 张冠亭. 以观山体验为导向的山地丘陵型城市公共空间评估与优化策略研究——以南京市为例[J]. *中国园林*, 2023,39(03):60-65.
- [5] Krause C L. Our visual landscape: Managing the landscape under special consideration of visual aspects[J]. *Landscape and urban planning*, 2001,54(1):239-254.
- [6] Qi Y, Chodron Drolma S, Zhang X, et al. An investigation of the visual features of urban street vitality using a convolutional neural network[J]. *Geo-spatial information science*, 2020,23(4):341-351.
- [7] Zhang F, Salazar-Miranda A, Duarte F, et al. Urban Visual Intelligence: Studying Cities with Artificial Intelligence and Street-Level Imagery[J]. *Annals of the American Association of Geographers*, 2024,114(5):876-897.
- [8] Biljecki F, Ito K. Street view imagery in urban analytics and GIS: A review[J]. *Landscape and Urban Planning*, 2021,215:104217.
- [9] Kruse J, Kang Y, Liu Y, et al. Places for play: Understanding human perception of playability in cities using street view images and deep learning[J]. *Computers, Environment and Urban Systems*, 2021,90:101693.
- [10] Zhang F, Fan Z, Kang Y, et al. "Perception bias": Deciphering a mismatch between urban crime and perception of safety[J]. *Landscape and Urban Planning*, 2021,207:104003.
- [11] 唐婧娴, 龙瀛, 翟炜, 等. 街道空间品质的测度、变化评价与影响因素识别——基于大规模多时相街景图片的分析[J]. *新建筑*, 2016(05):110-115.
- [12] 李心雨, 闫浩文, 王卓, 等. 街景图像与机器学习相结合的道路环境安全感知评价与影响因素分析[J]. *地球信息科学学报*, 2023,25(4):852-865.
- [13] Zhao X, Lu Y, Lin G. An integrated deep learning approach for assessing the visual qualities of built environments utilizing street view images[J]. *Engineering Applications of Artificial Intelligence*, 2024,130:107805.
- [14] Zhang A, Song L, Zhang F. Perception of pleasure in the urban running environment with street view images and running routes[J]. *Journal of Geographical Sciences*, 2022,32(12):2624-2640.
- [15] Xie E, Wang W, Yu Z, et al. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers[J]. *arXiv.org*, 2021.
- [16] Neuhold G, Ollmann T, Bulow S R, et al. The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes, 2017[C]. *IEEE*, 2017-01-01.
- [17] Ma X, Ma C, Wu C, et al. Measuring human perceptions of streetscapes to better inform urban renewal: A

- perspective of scene semantic parsing[J]. *Cities*, 2021,110:103086.
- [18] Li J, Zhang Z, Jing F, et al. An evaluation of urban green space in Shanghai, China, using eye tracking[J]. *Urban forestry & urban greening*, 2020,56:126903.
- [19] Liu Z, Mao H, Wu C, et al. A ConvNet for the 2020s[J]. 2022.