

项目编号：T430PRP37014

上海交通大学

本科生研究计划（PRP）研究论文 (第 37 期)

论文题目：上海城市景观与风貌调查：基于人工智能与大数据

项目负责人： 潘浩之 学院（系）： 设计学院

指导教师： 潘浩之 学院（系）： 设计学院

参与学生： 林意朋 董一尔 朱普申 叶思文 马瑞鹏

项目执行时间： 2020 年 2 月 至 2020 年 8 月

上海城市景观与风貌调查：基于人工智能与大数据

密西根学院：林意朋 518370910060

指导教师：设计学院 潘浩之

摘要

本研究通过发放问卷调查大众对于城市街景的美学认知，并利用问卷调查数据，与 PSPnet 要素提取神经网络，共同建立美学评估模型，从而为街景设计提供参考。同时，对所得数据进行数据分析，通过库克距离、皮尔森相关系数矩阵、特征重要性等手段，获取与大众对城市街景美学认知紧密相连的景观要素。经过验证，该美学评估模型被认为是可行、有效的。在数据分析中，研究者发现大众对街景的美学认知与其年龄、性别、学历并无关联，同时，植物层次丰富度、植物覆盖程度、植物数量丰富度与景观要素丰富度、建筑密度、景观要素丰富度、人工与自然比例对街景美观程度影响较大。

关键词：城市景观，美学，数学模型，数据可视化，机器学习

Abstract

This research uses questionnaires to investigate the public's aesthetic perception of urban streetscapes. By using those acquired data, together with image segmentation neural network pspnet, researchers establish an aesthetic evaluation model on urban streetscapes, which will further provide references for streetscape designs. At the same time, with analysis tools like Cook's distance, Pearson's coefficient, and feature importance, data analysis is performed on the obtained data to obtain landscape elements that are closely connected with the public's aesthetic perception of urban streetscape. After verifying the model, the aesthetic evaluation model is then considered feasible and effective. Researchers find that the public's aesthetic perception on street scenes is not related to their age, gender, and educational background. At the same time, plant layers, plant coverage, plant quantity, building density, landscape elements, ratio of artificial to nature have a greater impact on the aesthetics of the streetscape.

Keywords: Cityscape, Aesthetics, Model, Statistic, Data Visualization, Machine Learning

1. 绪论

1.1 选题意义

在中国城市化进程不断加快的今日，城市景观不断趋于一致化，各个城市的特色与本土文化屈服于大统一的建设进程，而淹没在城市肌理松散的新区中，失去了街道本身具有的社交性与文化性。固然，得益于模版化的城市建设，确实可以在短时间内发展出高度现代化的城市，从而获得高效的经济收益，却忽略了城市中历史风貌、人文环境的重要性。随着人们更多地参与在城市社会活动中，为了满足人们日益增长对幸福生活的需要，城市中的公共景观需要探索在保证建设速度的同时，保留其本身的美观度，进而需要有一个模型以反映城市景观建设的完备性。

本研究中通过大众对街景图片打分的问卷数据，并将街景图片数据化，从而进行美学评价模型的建立。经过数据分析，并结合美学评估，给出城市街景风貌的改进见解。

1.2 文献综述

在利用大数据、人工智能以进行街景美观程度评估的课题上，已有很多学者进行了深耕。Mohamed R. Ibrahim 等人^[1]介绍并总结了城市的分层，并列举了可用于进行街景美学元素识别、量化的卷积神经网络的算法。Fan Zhang 等人^[2]利用 MIT Place Pulse 数据集中 110988 张图片及其六个维度：令人沮丧程度 (Depressing)，无趣度 (Boring)，美丽度 (Beautiful)，安全度 (Safe)，活力度 (Lively)，富裕度 (Wealthy)，建立了美学评估体系，由街景图片经过要素提取后，进行线性回归与六个维度进行连接，达到以街景图片由机器自动评估的效果。并在最后利用上海及北京的城市街景数据，生成六维在城市道路上的分布。吴健生等人^[3]则从自然性、开阔性、多样性、奇特性、协调性方面入手，利用数据统计及计算，得到区域建筑密度、植被密度等值，并将计算数值与问卷所得数值做比对，从而验证数据计算数值的准确性。本研究以这些研究为基础进行调查。

1.3 研究范围

本研究主要从四个层面同时推进，分为现场调研、美学实验、人工智能、数据分析。

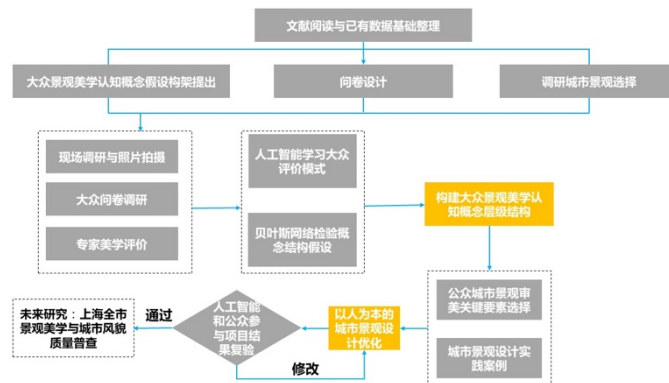


图 1 研究流程

现场调研部分负责给予 24 张街景图片专家打分，得到最终打分成绩。美学实验部分负责设计问卷，并向大众发放问卷并统计结果。人工智能部分负责使用要素提取模型进行街景图片的要素提取，从而作为训练美学评估模型的基础数据。数据分析部分负责对模型及数据的后续处理，评估模型及数据的有效性。

2. 研究内容与方法

2.1 现场调研组及美学实验组

美学实验组设计并发放了问卷。该问卷从年龄，性别，受教育程度，职业四个方面入手，优先取得可能与美学评判标准有关的数据。之后再要求受调研从 15 张街景图片中分 4 次，挑选 5 张分别具有自然性 (Naturalness)、开阔性 (Openness)、多样性 (Diversity)、协调性 (Well-coordinated) 的街景图片。汇总所得将被标记图片记为 1，未标记图片记为 0。搜集共计 215 份问卷。每份问卷对 15 张图片进行评估，共计数据条目 3225 条数据。

现场调研组在美学实验组的基础上，对自然性、开阔性、多样性、协调性这四种美学属性进行细分，得到植物层次丰富度 (plant layers)、植物覆盖程度 (plant coverage)、植物数量丰富度 (plant quantity)、建筑密度 (building density)、街道宽度 (street width)、空间拥挤感 (space congestion)、颜色丰富感 (color)、景观要素丰富度 (landscape elements)、建筑形态 (architectural form)、人工与自然比例 (ratio of artificial to nature)。经由专家的打分，将照片的各个指标通过公式计算得到中间分，之后使用层次分析法 (analytic hierarchy process (AHP)) 方法计算得出个指标的权重，再经过对以上细分美学属性的计算，得到每张照片最终的美学得分。

2.2 人工智能组

人工智能组对卷积神经网络进行学习之后，尝试使用多种要素分割神经网络模型，并选取了用 Cityscape 数据集训练的 PSPnet 模型得到 pspnet_101_cityscapes 模型。

2.2.1 Cityscapes 数据集

该数据集^[4]包含 5000 张详细标注、20000 张粗略标注的街景图片。经标注，照片中的不同元素被赋予不同的色彩，使其便于训练，共 22 种要素：交通信号灯 (traffic light)、小径 (path)、垃圾桶 (ashcan)、卡车 (truck)、小型摩托车 (minibike)、路标 (signboard)、花 (flower)、石块 (rock)、栅栏 (fence)、自行车 (bicycle)、街灯 (street light)、汽车 (car)、行人 (person)、道路 (road)、电线杆 (pole)、植物 (plant)、人行道 (sidewalk)、草 (grass)、树木 (tree)、天空 (sky)、建筑 (building)、墙壁 (wall)。



图 2 Cityscapes 数据集标注样例

数据集中的照片来自于德国的 50 个城市，摄于春、夏、秋不同天气情况下的白天，保证了数据集的多样性。

2.2.2 PSPnet 模型

由 Hengshuang Zhao 等人^[5]提出的 PSPnet 模型，在 FCN 模型的基础下，通过在 CNN 与卷积层之间增加多尺度池化层(Pyramid Pooling Model)，提升了模型对复杂场景的语义分割准确性。本研究使用了由 Image Segmentation Keras^[6]包含的预训练模型 pspnet 101 cityscapes 模型。

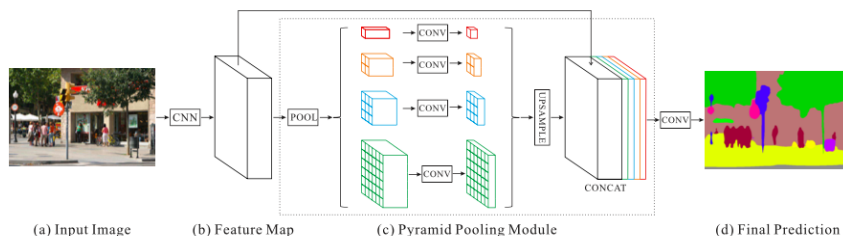


图 3 PSPnet 模型图解

2.2.3 图像处理

由于美学实验组及现场调研组所得照片大部分来源于谷歌街景，故其水印及 UI 无法移除。因此在将图片输入模型前，对图像按照照片尺寸每条边的 20%进行裁剪，最后保留中间 80%的图片。为了减少图片分割时模型的误判，让图片更好地被语义分割，之后还利用了池化 (pool) 将图片进行了模糊化处理。



图 4 图像处理示例

2.2.4 数据处理

使用 pspnet_101_cityscapes 模型对美学实验组放入问卷的 15 张街景样本以及现场调研组使用的 15 张街景照片样本进行语义分割。将经过语义分割的图片重映射到类别，统计各类别占有总图片的比例，将统计结果输出到表格。美学得分预测模型根据每张图片的要素占比得到最终的预测美学分数。此处将所有需要预测的图片的要素占比整合到一张表格，用以美学得分预测模型预测。数据分析组基于所得的美学分数进行可视化研究。

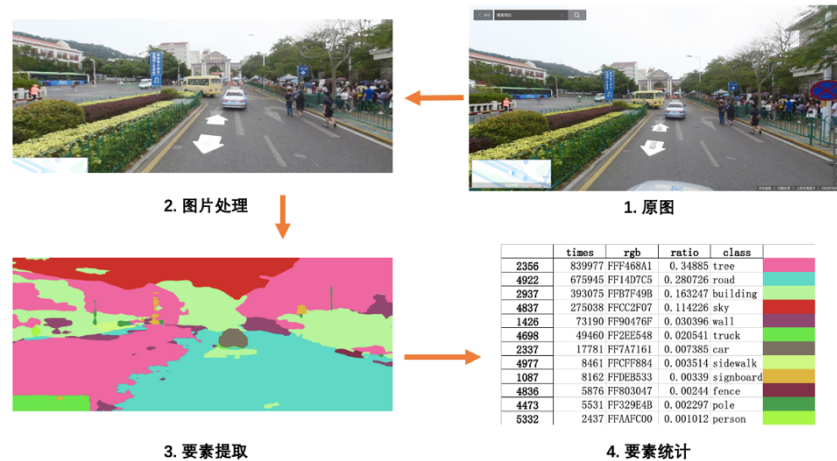


图 5 要素提取流程示例

2.3 数据分析组

数据分析组对问卷数据从以下几方面进行了可视化分析，并建立了回归模型来预测美学打分。

2.3.1 皮尔森相关系数 (Pearson's coefficient)

皮尔森相关系数描述了数据特征与数据特征之间的联系，其总是落在-1 到 1 的区间内。皮尔森系数越远离 0，两特征的相关程度越高。当皮尔森相关系数为正时，两特征为正相关，即当一个特征数值升高时，另一个也倾向于升高。当其为负时，两特征为负相关。当皮尔森相关系数趋近于 0，则两特征倾向于没有联系。皮尔森相关系数是两特征共有的，因而在可视化为热力图时，热力矩阵呈现为延左上右下对角线镜像，故仅保留左下三角部分。

2.3.2 库克距离 (Cook's distance)

库克距离展现了样本对回归模型的贡献程度，即重要程度。库克距离越大，样本对模型的影响越大。与此同时，由于诸样本对于模型的影响程度应为均等的，库克距离也可以用于判断样本中的异常值。

2.3.3 美学评估模型一

针对美学实验组的数据，利用逻辑回归模型，将问卷所得的每一美学特征的值，即被标记图片为 1，未标记图片为 0，作为模型的目标值，将图片要素提取后个元素的占比作为训练值。同时，将问卷数据集分为训练集与检验集，比例为 7.5:2.5，随机抽取。

2.3.4 模型评估

得到逻辑回归模型后，使用测试集为模型预测进行打分，通过分类报告直接生成多个打分成绩：精度 (precision)、召回值 (recall)、f1 值 (f1 score)、支持度 (support)，以此判断模型的有效性。

精度展现了预测当前类别中同一值中正确的比例，召回值展现了归于当前类别占归于正确类别的比例。两者结合展现了模型对数据的拟合程度。f1 值则是这两者的加权调和平均值，因而 f1 的值能很好地展现拟合程度。其分布在 0 到 1 之间，越接近 1，模型拟合程度越好。该值也是本研究中对模型评判的重要标准。

2.3.5 美学评估模型二

针对现场调研组的数据，利用岭回归模型，将现场调研组提供的美学打分进行拟合，并进行了特征重要性（Feature Importance）、残差图（Residuals）和皮尔森相关系数分析。

2.3.6 特征重要性

在 yellow brick 库中，特征重要性展现了数据中某一要素对模型某一特征的相对重要性。相关系数为正，则该代表街景中的该要素与某一美学特征具有正相关的联系，数值越远离 0，相关性越高。负则为负相关。通过展现特征重要性，可以了解模型内部对要素的权重。从而总结出城市街景要素与美学特征之间的联系。

2.3.7 残差分析

残差为模型的预测值与真实值之间的差值。残差的绝对值越大，预测的点与真实点之间的差别越大。通常由残差图来展现模型的残差。残差图以模型的预测值为横坐标，残差为纵坐标，每个点表示每条数据在某一特征上的残差。如残差图上的散点具有一定特征，则该模型在预测值时可能具有一定的偏差。好的模型的残差图，残差点应该随机分布在水平坐标轴附近。

3. 研究结果及讨论

3.1 美学实验组数据

3.1.1 皮尔森相关系数

在美学实验组的问卷数据中，共有 7 个目标属性，分别为性别（gender）、年龄（age）、受教育程度（Education）、自然性（Naturalness）、开阔性（Openness）、多样性（Diversity）、协调性（Well-coordinated），下图展现了各属性之间的皮尔森系数：

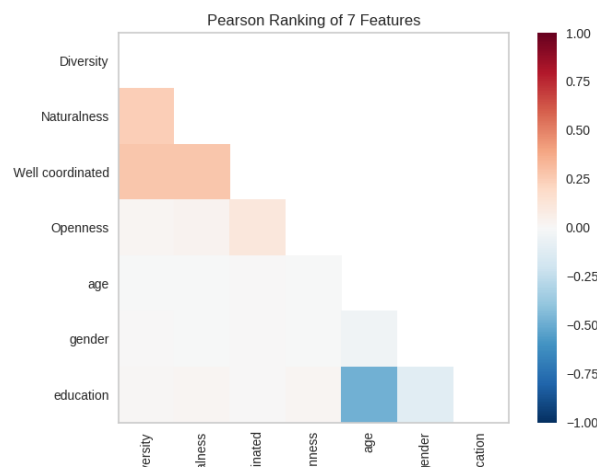


图 6 7 项特征的皮尔森相关系数矩阵

从数据上来看，自然性与多样性之间、协调性与多样性之间、自然性与协调性之间，相关系数位于 0.25 到 0.5 之间，可以认为三者互相具有一定的相关性。可以总结出该三点属性人们对街景美观程度的认识是相似的。

同时，从其余皮尔森系数中也可以看出，年龄、性别、学历对受调研者对街景美观程度的感知影响并不大，其相关系数均趋近于 0。因而可以得出结论对于街景的美观程度，大众的评判标准较为统一。

3.1.2 库克距离

根据问卷所得数据，可以绘制如下的库克距离。

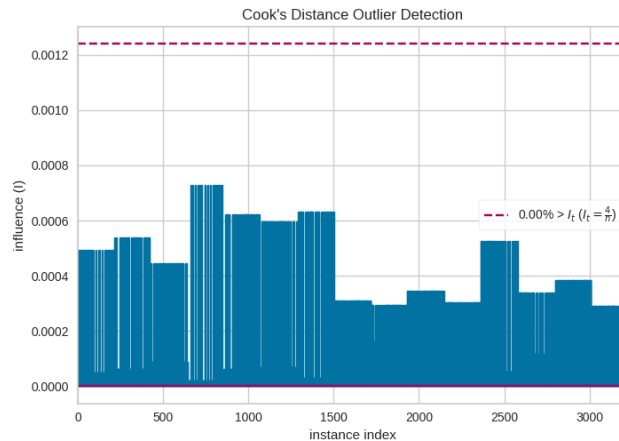


图 7 库克距离

从所绘制的图像可以看出，库克距离并没有出现过大的样本值，说明所有样本均符合规律，不存在异常样本。

3.1.3 特征重要性

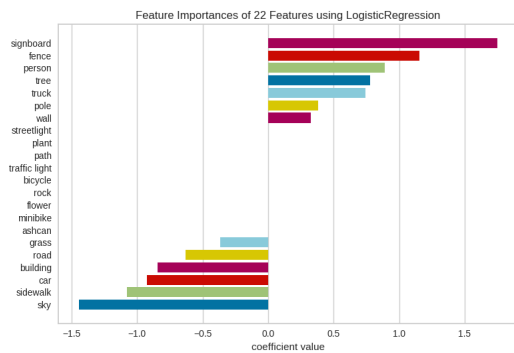


图 8 22 种要素与多样性特征重要性

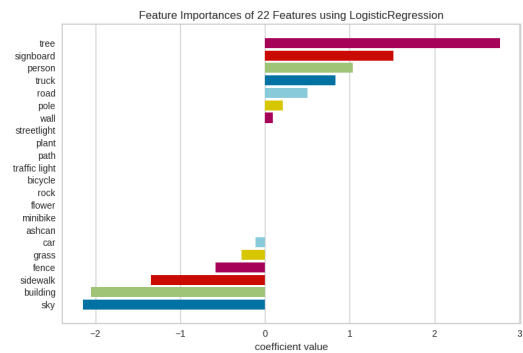


图 9 22 种要素和自然性的特征重要性

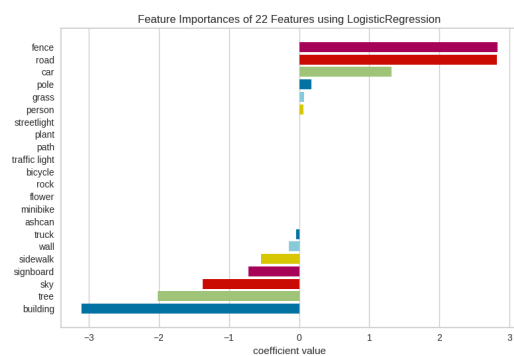


图 10 22 种要素和开阔性的特征重要性

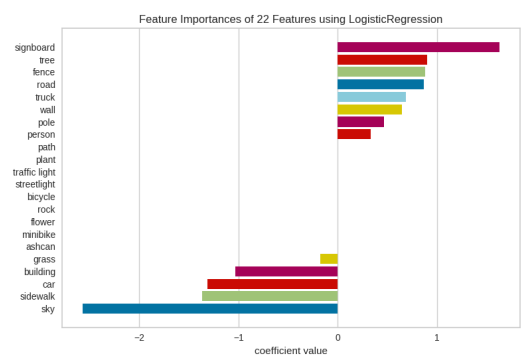


图 11 22 种要素和协调性的特征重要性

就多样性而言，路标、栅栏、行人、树木、卡车与多样性正相关，道路、建筑、汽车、人行道、天空与之成负相关，原因在于道路、建筑等是街景图片的要内容，而路标、栅栏、行人、树木则可以为街景增加不一样的元素。从这里可以总结出街景多样性的主要影响内容是路标、栅栏、行人、树木。

就自然性而言，可见树木与自然性成强烈正相关，这也符合人们对自然要素的定义。从中可以出的，绿化程度直接影响了人们对街景自然度的判断。

就开阔性而言，栅栏、道路与之成强烈正相关，建筑成强烈负相关，其原因在于道路占比大、建筑占比小通常意味着街道更加宽阔，栅栏占比多也意味着图中视觉阻挡物少，因而图像中道路、栅栏、建筑的占比极大影响了人们对街景图片开阔性的判断，这也符合常识。

就协调性而言，多数要素都与其有部分联系，值得注意的是天空与之成强烈负相关，可能原因是天空占比过高导致街景图片有一大片空白，引起受调研者认为街景图片不协调。

另外由于图片中不包含，或者无法从图片中分离小径、植物、交通信号灯、街灯、自行车、小型摩托车、石块、花、垃圾桶这些元素，因而其与四个属性的相关性均为 0

3.1.4 模型评估

美学评估模型使用了逻辑回归模型。其完成了对测试集的预测，数值上其对自然度预测的精确度达到了 0.847 与 0.711。

但在高精确度的背后，可能存在的问题是数据泄漏。由于只有 15 张图片，虽然有 215 人参与了评分，但在分离数据集的时候，由于 7.5:2.5 的分离比，共有 806 条数据划入测试集，这意味着模型训练基于的要素统计比例，模型在训练的时候就已经遇到过，所以对其的评判准确度会非常高。正确的做法应该是将 15 张图片分为训练集和验证集，但由于图片样本数量稀少，不足以支撑模型的训练，因而此处可以认为该模型是当前情况下能训练出的最优模型。

同时，从得分可以看出，该模型 recall 的分数比较低，其原因在于问卷评分得到 1 与 0 的比例失调，对每一项美学属性，由于在调研时只允许受试者从 15 张中选择 5 张认为符合条件的图片，因而 1 与 0 的比例较低，使得 recall 值较低。

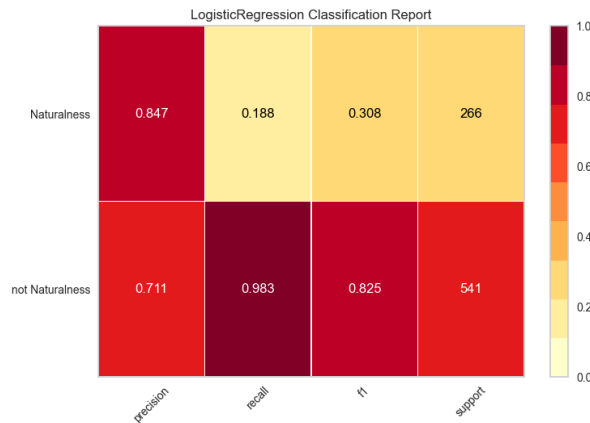


图 12 逻辑回归分类报告

3.2 现场调研组数据

3.2.1 皮尔森相关系数

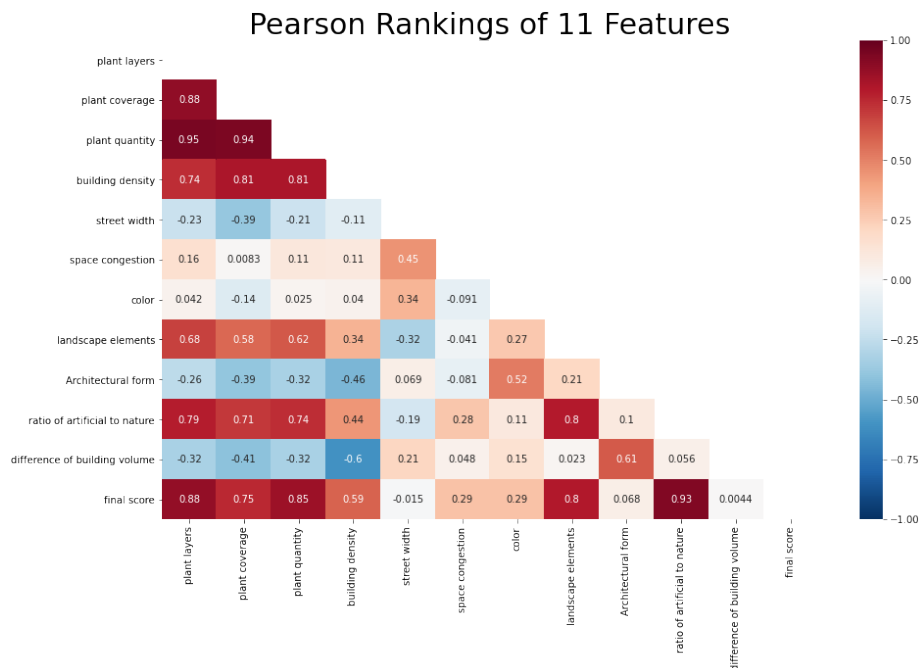


图 13 11 项特征与最终得分的皮尔森相关系数矩阵

从图中可以看出，植物层次丰富度 (plant layers)、植物覆盖程度 (plant coverage)、植物数量丰富度 (plant quantity) 与建筑密度 (building density)、景观要素丰富度 (landscape elements)、人工与自然比例 (ratio of artificial to nature) 高度相关。由于此处建筑密度为密度越低评分越高，故可以得出结论植物层次丰富度、植物覆盖程度、植物数量丰富度，即自然性，与建筑密度成反比，自然度越高，建筑密度越低。同时，自然性中的三要素又与多样性中的景观要素丰富度、协调性中的人工自然比例紧密联系，该联系符合对景观要素丰富度的评判标准：“建筑小品、植物、水体、园路、铺装”五大要素存在越多，人工与自然比例越协调。

值得注意的是，对于最终打分，植物层次丰富度、植物覆盖程度、植物数量丰富度与景观要素丰富度、建筑密度、景观要素丰富度、人工与自然比例六项又与之成强正相关，可见该六项在对街景美感产生中的重要作用。

3.2.2 特征重要性

表 1 11 项特征与要素的特征重要性

	正相关			负相关	
植物层次丰富度	tree			signboard	fence
植物覆盖程度	tree			building	signboard
植物数量丰富度	tree			building	fence
建筑密度	building				
街道宽度	signboard	sky		tree	wall
空间拥挤感	road			signboard	fence
颜色丰富度	fence	building	signboard	road	
景观要素丰富度	tree	building		road	signboard
建筑形态	building			road	tree
人工与自然比例	tree			signboard	fence
建筑体量差异比例	building	sky	person	tree	fence

与美学实验组中获取特征重要性的方法相同，通过生成 11 幅特征重要性视图，将其整合到一张表格中。

从表格中可以看出，树木与所有与自然有关的特征中都成正相关关系。表明在人视角中的街景植被占比率，能反映街景中自然与美的程度。

3.2.3 残差分析

对于该美学评估模型，使用了岭回归的方法进行了拟合。为了展现模型的拟合程度，研究者可视化了残差图：

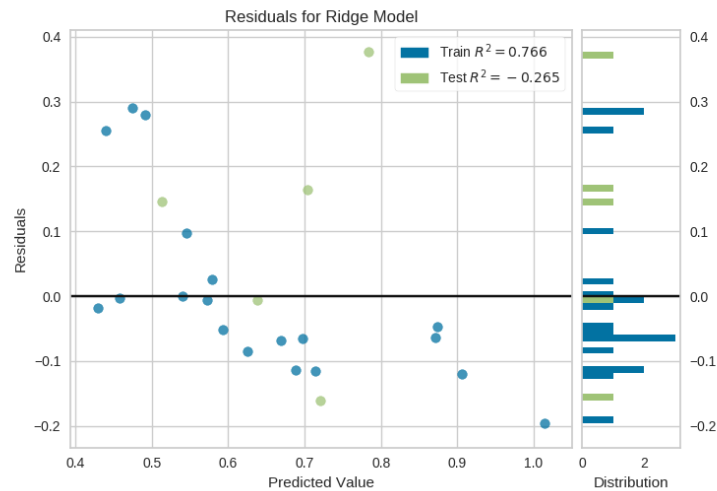


图 14 岭回归模型残差分析

从图中可见，所有数据均落在-0.2 到 0.4 之间，偏离水平坐标轴较小，且数据点分布不具有特定规律，因而可以推断，拟合的岭回归模型有良好的拟合率，能准确预测美学分数数据。

4. 讨论

4.1 模型

在图像要素提取的过程中，本研究使用了由 cityscapes 数据集训练所得的 PSPnet_101_cityscapes 模型，将需要建立美学评价模型的街景照片作为输入，并得到语义分割后的图片。

从直观角度而言，PSPnet_101_cityscapes 模型并没有能很好地对本次研究中使用的街景图片进行很好的分割。

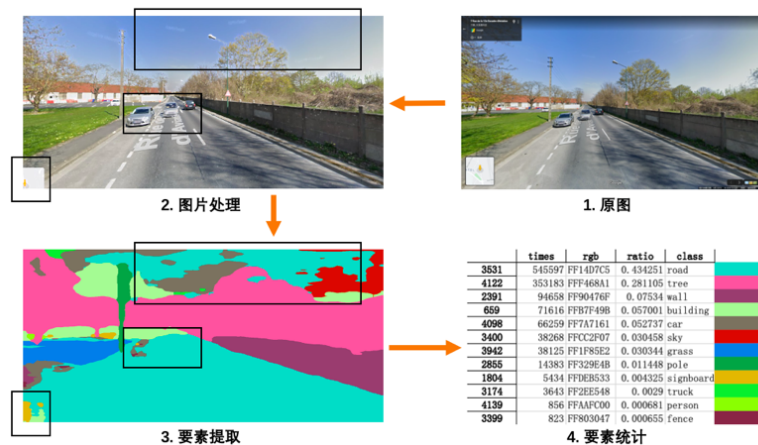


图 15 水印影响示例

以上图为例，由于原图从谷歌街景中直接截取得来，故其 UI 及水印无法移除。经过图像处理，大部分 UI 可以被截取到输入图片外，但水印、小部分 UI 仍然保留在图片上，在进行语义分割的过程中，模型并不能很好地排除水印的干扰。在天空的部分，水印对语义识别的干扰很大，导致天空部分无法准确地分割，对后续开闭性模型的训练产生影响。

4.2 要素统计后的信息损失

街景图片经过语义分割统计后，保留的图片信息仅余留某种要素占据图片的比例，因而重新利用这些数据训练美学评价模型会导致图片得到的某些图片特征不能被评估模型习得。利用要素统计训练自然度的模型具有很高的精准度，其原因在于人们对自然度的评价主要基于图片中植被

的密度,该参数与要素统计中植物占据图片的比例高度相关,因而可以被模型高度准确地预测。但在建立多样性和协调性的模型时,其准确度与召回度趋近于零,可能原因有如下几点:1,打分较为松散,即在 215 人对某一图片的多样性、协调性认可的比例几乎相同,使得模型无法分辨打分的倾向。2,要素统计对图片的某些要素没能准确识别,或判断错误,导致统计的结果出现偏差,与实际的要素占比有差异。3,用以训练的样本量不足,本次研究采用 215 人对 15 张照片的打分,归并后仅有 15 条数据,无法建立可靠的模型。4,要素占比与多样性、协调性没有直接联系,两者之间没有可建模性。

改进层面可以有以下几点:

- 1, 增加样本量,尤其为打分的照片数量,由此可以学习到参数与多样性、准确性的联系。
- 2, 扩充图片特征,如增加道路、植被、人行道的特征,以此减少图片中信息的损失,从而与待学习的多样性、准确性产生联系。

5. 结论

本研究通过发放问卷,利用问卷数据建立评估模型并进行数据分析的模式,建立了大众对城市街景美学认知的模型,用以利用人工智能程式来模拟大众的美学认知。通过数据可视化与模型评估,研究者验证了该美学评估模型的可行性,并得出了以下结论:

街景美观程度通常与植物层次丰富度、植物覆盖程度、植物数量丰富度与景观要素丰富度、建筑密度、景观要素丰富度、人工与自然比例六项紧密联系,在建立对于城市街景的设计中,该六者需要着重考量。这六者中又以自然性分属的植物层次丰富度、植物覆盖程度、植物数量丰富度为主体,相关性最为密切,由此可见自然性在大众对街景的美学认知中的重要作用。

年龄、性别、学历对受调研者对街景美观程度的感知影响并不大。从受调研者对街景图片打分与他们的个人信息的相关系数来看,并没有直接的证据证明这两类之间存在紧密的联系。因而可以推断大众对街景的认知与大众本身的年龄、性别、学历几乎无关。

最后,本研究也有很多需要改进的方面,比如增加样本数量以建立更加准确的模型、通过图片直接训练美学评价模型等。

参考文献

- [1] IBRAHIM M, HAWORTH J, CHENG T. Understanding cities with machine eyes: A review of deep computer vision in urban analytics[J]. Cities, 2020, 96: 102481.
- [2] ZHANG F, ZHOU B, LIU L et al. Measuring human perceptions of a large-scale urban region using machine learning[J]. Landscape and Urban Planning, 2018, 180: 148-160.
- [3] 吴健生,袁甜,王彤.基于三维景观指数的城市景观美学特征定量表达研究——以深圳市为例[J/OL].生态学报,2017(13):1-10[2020-09-15].
- [4] Cityscapes Dataset – Semantic Understanding of Urban Street Scenes[EB/OL]. Cityscapes-dataset.com, 2020. (2020). <https://www.cityscapes-dataset.com/>.
- [5] ZHAO H, SHI J, QI X et al. Pyramid Scene Parsing Network[J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- [6] divamgupta/image-segmentation-keras[EB/OL]. GitHub, 2020. (2020)[2020-09-15]. <https://github.com/divamgupta/image-segmentation-keras>.

谢辞

感谢潘浩之教授为本研究提供全程的指导。感谢数据分析组的董一尔、朱普申、叶思文同学能共同讨论关于数据处理的事宜,并完成了精美的数据可视化。感谢吴恩达教授的机器学习入门课程和阿斯顿·张、李沐、扎卡里 C. 立顿、亚历山大 J. 斯莫拉撰写的《动手深度学习》,为本文研究者提供了窥探机器学习领域的一个机遇。再次感谢所有支持并为本项目提供帮助的人。