

# 基于机器视觉识别的户外环境情绪感受测度研究

## Affective Experience Measurement in Outdoor Environment via Machine Visual Cognition

付而康

四川农业大学风景园林学院讲师

周佳玟

四川农业大学风景园林学院在读硕士研究生

姚智

华南师范大学华南先进光电子研究院彩色动态电子纸显示技术研究所在读硕士研究生

李西\*

四川农业大学风景园林学院教授、博士生导师

FU Erkang

Lecturer at College of Landscape Architecture, Sichuan Agricultural University

ZHOU Jiawen

Graduate Student at College of Landscape Architecture, Sichuan Agricultural University

YAO Zhi

Graduate Student at Institute of Electronic Paper Displays, South China Academy of Advanced Optoelectronics, South China Normal University

LI Xi

Professor and PhD Supervisor at College of Landscape Architecture, Sichuan Agricultural University

\*通讯作者

地址：四川省成都市温江区惠民路211号四川农业大学7教609

邮编：611134

邮箱：lixisicau.edu.cn

编辑 | 冉玲于、田乐

翻译 | 冉玲于、田乐、肖杰

EDITED BY | RAN Lingyu, Tina TIAN

TRANSLATED BY | RAN Lingyu, Tina TIAN, XIAO Jie

摘要

良好的户外环境有利于改善人群身心健康，环境中个体情绪感受的实时测度能客观反映人群对户外环境品质的满意程度，但目前应用于户外环境中的情绪感受测度方法还较少，且无法在场地面积大、样本量较多的情境下实现大范围的高效测度。人工智能领域的机器视觉识别可通过视频数据实现对动态面部表情特征的准确识别，使得开展高精度、长周期的户外环境情绪感受测度成为可能。本文以城市社区广场环境为实证实验场地，基于卷积神经网络算法模型，同步收集人群在户外环境体验时的面部视频数据和皮电数据，训练生成和检验测试可判别户外环境中个体面部情绪的深度学习算法模型——编解码器-SVM优化模型，并在街头绿地环境中展开拓展实验检验模型可靠性。研究表明：1) 实证实验和拓展实验中的人群情绪感受测度准确度分别达到82.01%、65.08%；2) 拓展实验验证该算法模型具备推广应用潜力；3) 模型对于人群行为丰富、视野开阔的空间更具适用性。因此，基于机器视觉识别的户外环境情绪感受测度将有助于在场地面面积较大、样本量较多的情境下揭示人群在环境体验中的心理状态，提升景观品质优化策略的有效性，同时也将为存在沟通或阅读障碍的特殊人群的情绪感受识别提供技术支撑，有望推动智慧城市建设的精细化转型。

关键词

机器视觉识别；户外环境体验；情绪感受测度；卷积神经网络算法模型；健康促进；设计研究

ABSTRACT

Outdoor environments with quality landscapes can benefit people's physical and mental health. Real-time assessment on individuals' environmental affective experience can improve the scientificity in measuring the quality of outdoor environments. Existing measurement methods are often insufficient for the cases of a larger site area or sample size. The machine visual cognition of Artificial Intelligence can realize the recognition of facial expressions and the changes in video images, which supports high-precision and long-cycle measurements on individuals' affective experience in outdoor environments. Taking an urban community square as the study site, this research simultaneously collects participants' facial data from video images and their electrodermal activity data, wherein Convolutional Neural Network algorithm model is trained with a deep learning algorithm, i.e. codec-SVM optimized model, whose reliability is tested through an additional experiment. The research reveals that: 1) The accuracy rate of the main and additional experiments in measuring individuals' affective experience is 82.01% and 65.08%, respectively; 2) The additional experiment verifies the application potential of the codec-SVM optimized model; And 3) the model works more effective for outdoor scenarios with varying usage behaviors and open views. Therefore, machine visual cognition can be used for emotion measurement in a larger site area or sample size and contributes to the effectiveness of landscape optimization efforts, especially as an instrumental tool to study the affective experience of the ones who have communication or reading disability. The findings also demonstrate the model's great potential in building Smart Cities with refined public services.

KEYWORDS

Machine Visual Cognition; Outdoor Environments Experience; Affective Experience Measurement; Convolutional Neural Network Algorithm Model; Health Promotion; Design Research

<https://doi.org/10.15302/J-LAF-1-020055>

收稿时间 RECEIVED DATE | 2021-05-31

中图分类号 | TU-856, B845.6, TU986, TU984.11+1

文献标识码 | A

## 1 研究背景

情绪是人脑基于主体需要对客观环境与事件刺激的反应<sup>[1]</sup>，可以通过生理、心理和社会适应等多种途径作用于健康<sup>[2]</sup>。快节奏、高强度的现代城市生活方式容易诱发焦虑、抑郁、不安等不良情绪，极大地损害了人群身心健康<sup>[3][4]</sup>，而景观品质较好的户外环境能够缓解人群不良情绪，对于改善人群身心健康具有积极作用<sup>[5][6]</sup>。近期，环境心理学相关研究证实，针对情绪感受的实时测量能更客观地反映人群对户外环境景观品质的满意程度<sup>[7]</sup>，有利于增强环境改造/提升措施的有效性。然而，既有的情绪感受测度方法多应用于室内环境，户外环境应用较少，且不适用于在场地面积大、样本量较多的情境下对使用者情绪感受展开大范围测度。因此，如何提高户外环境情绪感受测度的实时性和推广性还有待深入探讨。

## 2 户外环境情绪感受测度方法

### 2.1 既有户外环境情绪感受测度方法及其局限性

户外环境情绪感受是个体根据户外环境对自身需求的满足程度而产生的综合状态，包括了主观体验、生理反应及外部表现三个方面<sup>[8]</sup>，其测度方法通常可分为自我报告、生理测量及行为测量三种类型<sup>[9][28][46]</sup>。自我报告一般为被试者利用情绪评定量表或问卷——常用的情绪评定量表有积极消极情绪量表（PANAS）<sup>[10]</sup>、心境状态量表（BPOMS）<sup>[11]</sup>、状态特质焦虑量表（STAI）<sup>[12]</sup>等——对自身的情绪感受进行描述或等级评价，以此测度其在户外环境中的主观体验。生理测量可通过对人体皮电（EDA）、心率、血压、肌电、脑电波等生理指标<sup>[13][14]</sup>的监测，反映伴随情绪发生的复杂神经过程和生理变化，常用于测度被试者的瞬时情绪变化。行为测量通常对被试者的声音特性、面部表情、躯体行为等外显行为进行直接观察和测度来衡量个体情绪感受<sup>[15]</sup>。

为厘清上述方法在户外环境情绪感受测度过程中的优劣势，本研究参考相关文献<sup>[16]</sup>从适用范围、干扰程度、时间成本、设备成本和人力成本五个方面对三类方法进行比较（图1，表1）。从比对结果看，自我报告适用范围广、测度干扰程度低、设备及人力成本较低的情况，具有简单易行的特点，但其测度周期往往较长，不适用于以提升使用者满意度为目标、需在实践过程中对使用者情绪感受进行实时测度的设计<sup>[17]-[20]</sup>。而生理测量虽然测度周期较短，但存在单人次的测度耗时较

## 1 Research Background

Emotion is the affect and reaction of human mind, based on the needs of the subject, to certain environmental or event stimuli<sup>[1]</sup>. They affect human health physiologically and psychologically, and on social adaptation as well<sup>[2]</sup>. Fast-paced, busy modern lifestyle often make citizens stressed, anxiety, depression, and uneasiness that is harmful for their physical and mental health<sup>[3][4]</sup>. Existing studies prove that outdoor environments with quality landscapes can help alleviate such negative moods and benefit people's physical and mental health<sup>[5][6]</sup>. Recent research in Environmental Psychology suggests that real-time assessment on individuals' environmental affective experience can improve the scientism in measuring the quality of outdoor environmental landscapes<sup>[7]</sup>, and effectively contribute to the enhancement of landscape design and environmental performance. Existing assessment methods are mostly applied in scenarios of indoor environments, while those for outdoor environments are often insufficient, especially for the cases of a larger site area or sample size. It thus needs to innovate measurement methods that can generally applied for studying people's real-time affective experience for outdoor environments.

## 2 Measurement Methods on Affective Experience in Outdoor Environments

### 2.1 Existing Measurement Methods on Affective Experience in Outdoor Environments and the Limitations

Individuals' affective experience in outdoor environments is an overall emotional state determined by their affected level to the environment, which includes subjective experience, physiological reactions, and external performance<sup>[8]</sup> and can be assessed with self-report, physiological measurement, and behavioral measurement, respectively<sup>[9][28][46]</sup>. Self-report often in forms of mood scales or questionnaires—mostly Positive and Negative Affect Scale (PANAS)<sup>[10]</sup>, Brief Profile of Mood States (BPOMS)<sup>[11]</sup>, and State-Trait Anxiety Inventory (STAI)<sup>[12]</sup>—is a means that individuals subjectively describe or rate their affective experience in outdoor environments. Physiological measurement is to reflect individuals' instantaneous emotional changes, which monitors complex neurological processes and physiological changes caused by mood fluctuations and mainly indicated with human electrodermal activity (EDA), heart rate, blood pressure, electromyography, and electroencephalogram<sup>[13][14]</sup>. Behavioral measurement is to assess individuals' affective experience through direct observation of the changes in external behaviors such as voice, facial expression, and body posture<sup>[15]</sup>.

To better understand the pros and cons of each measurement method above, this study reviews them at five aspects, namely application scope, degree of intervention, time cost, device investment, and manpower requirement<sup>[16]</sup> (Fig. 1, Table 1). Self-report is found with a broader application, low degree of intervention, and less device investment and manpower requirement. While, due to its high time cost, self-report is not an ideal choice for real-time measurement of individual's affective

长，不适用于大样本量的情绪测度<sup>[21]-[24]</sup>。此外，生理测量还需提供相应的仪器设备，测度成本较高，同时穿戴设备本身也会对被试者被测度时的情绪感受造成一定干扰<sup>[25]</sup>。行为测量的常用方法有观察者评价<sup>[26]</sup>、叙事陈述<sup>[27]</sup>等，虽然可以对被试者的情绪感受进行直观测量，但往往测度耗时较长，且需要大量劳动和培训<sup>[28]</sup>。鉴于此，随着城市居民对建成环境健康促进功能需求的不断提高，需要引入新的技术手段以开展更高效、更广域、更实时的环境健康情绪测度。

表1: 既有户外环境情绪感受测度方法比较

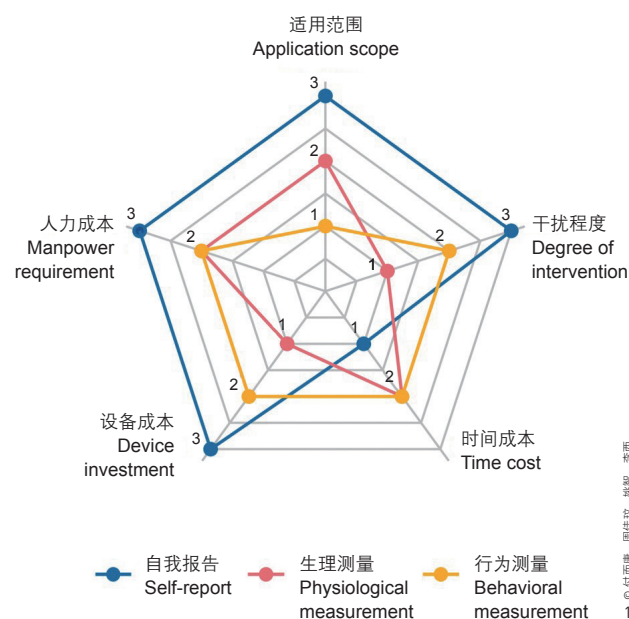
Table 1: The comparison of existing measurement methods on affective experience in outdoor environments

评价标准 Evaluation standard		
评价维度 Aspect	评价指标 Indicator	评价得分 Score
适用范围 Application scope	大多需在实验室环境下进行 Mostly in laboratory environment	1
	可以在实景环境 / 实验室环境中进行 Authentic and laboratory environments	2
	被试者没有区域限制 No limitations	3
干扰程度 Degree of intervention	需要穿戴设备或侵入式采集 Wearing devices or invasive collection	1
	需要简单设备 Wearing simple devices	2
	不需要穿戴设备或侵入式采集 No device or invasive collection	3
时间成本 Time cost	测度周期15天以上 Over 15 days of measurement	1
	测度周期5-15天 5 ~ 15 days of measurement	2
	测度周期5天以下 Less than 5 days of measurement	3
设备成本 Device investment	需要生理测量设备 Physiological measurement device	1
	需要简单设备 Simple device	2
	不需要设备 No device	3
人力成本 Manpower requirement	研究人员10人以上 More than 10 researchers	1
	研究人员3-10人 3 ~ 10 researchers	2
	研究人员1-3人 1 ~ 3 researchers	3

experience in the design practice aiming at improving user satisfaction<sup>[17]-[20]</sup>. Physiological measurement is advantageous in measurement time, but it takes a longer time in individual measurement that makes it not applicable to large sample-size research<sup>[21]-[24]</sup>. Besides, it often requires high device investment, and the wearing of device would more or less intervene individuals' affective experience during the measurement<sup>[25]</sup>. Although behavioral measurement (observers' ratings<sup>[26]</sup> and narrative statements<sup>[27]</sup> mainly) can directly assess individuals' affective experience, but is often time-consuming and require much labor and training<sup>[28]</sup>. As the increase of citizens' health demands in built environment, new technologies and approaches need to be introduced to promote the effectiveness and application scenarios of real-time measurement of affective experience in urban environments.

## 2.2 The Measurement of Affective Experience in Outdoor Environments via Machine Visual Cognition

Machine vision is a technology that simulates human vision by machines. The computer extracts valid information from images and recognizes the image contents through data analysis and processing<sup>[29]</sup>. This technology sees advantages in high precision, supporting real-time application, and advanced intelligence<sup>[29]</sup>,



© 付国顺, 周桂秋, 姚智, 李西

1. 既有户外环境情绪感受测度方法比较雷达图

1. The radar chart of existing measurement methods on affective experience in outdoor environments

### 注

1. 评价得分说明：1分表示优势较弱，2分表示优势较强，3分表示优势最强；
2. 实验周期由单人次实验时间及样本量估算得出。

### NOTES

1. Score: 1 means a slim superiority, 2 means a strong superiority, and 3 means an outstanding superiority;
2. The measurement time is estimated upon the time spent on each participant and the whole sample size.

## 2.2 机器视觉识别在户外环境情绪测度中的应用

机器视觉是机器对人视觉功能的模拟，即通过计算机从图像中提取有效信息进行分析处理，实现对图像内容的识别<sup>[29]</sup>。该技术具有精度高、实时性强、自动化与智能化程度高<sup>[29]</sup>等优点，已在公共安全<sup>[30][31]</sup>、智能驾驶<sup>[32][33]</sup>、医疗健康<sup>[34][35]</sup>等领域广泛应用。

在环境情绪感受测度方面，由于面部表情是人们传达情绪最直接的方式<sup>[36]</sup>，基于机器视觉识别的情绪分析也多集中在面部表情上，主要包括人脸检测、面部特征提取、情绪分类三个步骤<sup>[37]</sup>。其中，面部特征提取和情绪分类是决定情绪识别结果的关键<sup>[37]</sup>。目前，具有强大特征学习能力的卷积神经网络（Convolutional Neural Network，简称CNN）算法模型已能通过计算机自动分析出视频中面部表情及其变化，实现对动态图像的面部情绪识别<sup>[38]</sup>，极大地提升了机器视觉面部情绪识别的精确度和智能化程度<sup>[39]</sup>。该技术不同于对静止图像的识别，基于视频数据的面部情绪识别可以追踪和提取个体具有时空连续性的多幅图像信息<sup>[40]</sup>，从而对个体情绪感受进行更加客观全面的识别和测度。在现有的环境情绪感受测度研究中，机器视觉识别主要应用于室内和室外环境。在室内环境方面，维尔达·巴纳·普特拉等人基于机器视觉识别构建了实时面部表情识别系统，用于实时测度课堂中学生的情绪变化，以便及早发现学生的不良情绪<sup>[41]</sup>；平金元将机器视觉技术和皮电数据相结合，建立了可测度学生心理状态的算法模型，用以实时反馈学生的课堂参与度<sup>[42]</sup>。在户外环境方面，康宇豪等人从社交媒体网站中收集了80个带有地理坐标的旅游景点照片，利用机器视觉识别分析照片中的面部情绪，探究人群环境情绪感受与环境因素之间的关系<sup>[43]</sup>。

因此，相较于既有的户外环境情绪感受测度方法，机器视觉识别具有独特优势。从适用范围看，机器视觉识别对于测度场地的限制较小，既能应用于实验室环境<sup>[44]</sup>，也能应用于不同尺度及景观特征的户外环境<sup>[42][43]</sup>；从干扰程度看，机器视觉识别通过非接触的方式获取个体面部图像信息<sup>[41][45]</sup>，可最大限度减少对被试者的干扰；从实验成本看，模型的训练阶段需要技术支持，而模型训练完成后的数据采集过程可以通过摄像头采集等手段实现<sup>[46]</sup>，设备成本、时间成本及人力成本均较低，有较好的推广应用价值。

但是，机器视觉识别在户外环境的情绪感受测度方面应用还较少；而且，鉴于户外环境具有环境要素复杂、人群姿态多样、光照条件多变等特点，当前的机器视觉面部情绪识别所使用的算法模型的初始识别率通常较低<sup>[47]</sup>。基于此，本研究最终采用了对于复杂户外环境中的面部情绪识别具有较强适用性的CNN算法模型<sup>[48][49]</sup>，相较于传统聚类模型、贝叶斯分类模型为代表的机器学习模型<sup>[50][51]</sup>，该模型具有较强特征提取和分类能力，可以通过卷积层自动完成特征提取，并借助深度学习训练优化模型参数，从而更为精确地实现对个体户外环境情绪感受的识别。

and has been widely used in fields like public safety<sup>[30][31]</sup>, intelligent driving<sup>[32][33]</sup>, and medical health<sup>[34][35]</sup>.

As facial expressions are people's the most direct way to convey feelings<sup>[36]</sup>, emotion analysis supported by machine visual cognition focuses on facial expression dissecting, which includes three steps, namely face detection, facial feature extraction, and emotion labeling<sup>[37]</sup>—the latter two are critical to the emotion recognition results<sup>[37]</sup>. The Convolutional Neural Network (CNN) algorithm model with advanced machine learning capability has greatly elevated the accuracy and intelligence of facial emotion recognition of machine vision<sup>[38]</sup>. This technology can track and extract facial information from images of both time and spatial continuity, which supports a more comprehensive and accurate analysis of human facial expressions and their changes via video monitoring<sup>[39][40]</sup>. At present, emotion recognition by machine vision has been used in both indoor and outdoor environments. Werda Buana Putra et al. built a real-time facial expression recognition system via machine visual cognition which realizes the real-time monitoring of students' emotional changes at class so as to find out their unhealthy emotions as soon as possible<sup>[41]</sup>. By combining machine vision technology with EDA data, Pyoung Won Kim established an algorithm model that evaluates students' real-time mental states, thereby learning about their engagement level at class<sup>[42]</sup>. Yuhao Kang et al. used machine visual cognition to recognize people's facial emotions in 80 geotagged photos about tourist attractions collected from social media platforms, aiming to explore the relationship between the environments and individuals' affective experience<sup>[43]</sup>.

Relatively, machine visual cognition is superior over other existing methods in the measurement of affective experience in outdoor environments. First, machine visual cognition is less limited by the site's conditions and can be applied into both laboratory<sup>[44]</sup> and outdoor environments at varied scales and with different landscape characteristics<sup>[42][43]</sup>. Secondly, machine visual cognition collects individuals' facial image data in a non-contact manner<sup>[41][45]</sup>, minimizing the intervention over experiment participants. Thirdly, as for the experimental cost, although technical support is required for the model training of machine vision, after that data can be collected via cameras at low cost<sup>[46]</sup>, saving device, manpower, and time cost, which makes the technology widely applicable.

However, machine visual cognition is little applied in measuring affective experience in outdoor environments, and the accuracy of facial emotion recognition by existing algorithm model is often far from satisfactory, due to the complexity and variety of environmental elements, body postures, and lighting conditions in outdoor scenarios<sup>[47]</sup>. Therefore, this research adopts CNN algorithm model<sup>[48][49]</sup> for its high applicability to complex outdoor environments—Compared with traditional machine learning models (such as traditional clustering models and Bayesian Classifier), CNN algorithm model shows a stronger capability in facial feature extraction and emotion labeling<sup>[50][51]</sup> via convolutional layers, and optimize the model parameters through deep learning training, recognizing individuals' affective experience in outdoor environments with a higher accuracy.

## 2.3 研究对象与假设

基于以上认识,研究选取了城市居民开展日常户外活动较为频繁的空间载体——城市社区广场为样本,通过开展机器视觉情绪感受测度实验,在CNN算法模型基础上,训练生成可判别面部情绪的深度学习算法模型,实现并验证了基于机器视觉识别的户外环境情绪感受测度的可行性,为引入该技术方法辅助景观设计实践提供了实证支持。

## 3 实证研究

### 3.1 实验设计与数据获取

本文选取成都市成华区锦绣社区睦邻广场为研究场地,开展基于机器视觉识别的户外环境情绪感受测度实验。如表2所示,锦绣社区睦邻广场北临城市次干道、南临居民区、东侧为锦绣社区服务中心、西侧为社区街头绿地,面积约为1.8hm<sup>2</sup>。广场内视野开阔,拥有植被、廊架、座椅、铺装等景观要素,是典型的城市开敞空间。为保证全方位获取视频人脸数据,研究者在广场中布置了9个摄像机位。为减少天气波动对数据采集的影响,实验于天气晴朗、温度适宜时段开展(2019年9月19—21日,08:00~12:00)。

参考既有情绪感受测度相关研究,一般实验被试人数为30人以上时即认为数据可靠<sup>[52]</sup>;为提高情绪识别模型的适用性,还应避免被试者的人口特征过于单一,因此本次实验共招募42位不同性别和年龄的锦绣社区居民作为被试者<sup>①</sup>。单人次实验步骤包括(图2):

1) 准备阶段:被试者食指和中指佩戴便携式PPG-EDA生物反馈装置<sup>②</sup>并行走适应,其间研究人员向被试者介绍实验流程及注意事项;

2) 初始阶段:被试者在等候区静坐休息10分钟<sup>③[53][54]</sup>,使被试者情绪恢复平静,以测量其初始皮电数据;

3) 应激阶段:被试者被要求在封闭的车内边听噪音边做算术题3分钟<sup>④</sup>(此应激阶段是为了诱发被试者的焦虑情绪,使所有被试者处于同一情绪状态以减少实验误差并增加结果可比性);

4) 体验阶段:被试者进入研究场地按日常习惯进行环境体验5分钟<sup>⑤[55][56]</sup>;

① 本次实验经作者所在单位实验伦理委员会批准,被试者均自愿参与实验,文中肖像使用均已征得被试者同意。

② 本研究所使用的设备为Biopac MP150便携穿戴式生理记录分析系统,便携性较高,几乎不会对被试者的空间体验造成影响。

③ 参考相关既有研究中的相关测量时间设定,本研究将初始阶段的静坐休息时间设定为10分钟(参见参考文献[53][54])。

④ 在预实验阶段,作者已对算术题、恐怖影片、噪音等应激材料的效果进行了对比,得出在噪音条件下做算术题所诱发的焦虑情绪的效果最佳。

⑤ 参考相关既有研究,3~5分钟的环境体验已经能够引起生理变化(参见参考文献[55][56])。

## 2.3 Sample Selection and Research Hypothesis

Through an experiment on urban community squares, one of the most frequently used place types for citizens' daily outdoor activities, this research tests the accuracy of machine visual cognition in affective experience measurement, wherein CNN algorithm model is trained with a deep learning algorithm that can discriminate facial emotions. The research demonstrates the effectiveness of measuring affective experience in outdoor environments via machine visual cognition, hoping to provide empirical references for landscape design practices.

## 3 Empirical Research

### 3.1 Experiment Design and Data Collection

This research selects the Jinxiu Community Square in Chenghua District, Chengdu City as the study site for the experiment of measuring affective experience via machine visual cognition. Showing as Table 2, the 1.8 hm<sup>2</sup> study site neighbors a secondary city road, a residential area, and a Community Service Center, and connects with other community green spaces. With an open view, the site is covered with vegetation and pavements and has corridors, seats, and other landscape elements. Nine cameras were placed on the site to ensure an all-round vision for facial image collation. The experiment was conducted on sunny days with a comfort weather (08:00 ~ 12:00, from September 19 to 21, 2019).

Referring to previous studies on affective experience measurement, the number of experiment participants is reliable when it exceeds 30<sup>[52]</sup>. A total of 42 males and females from the neighboring community were recruited as participants<sup>①</sup> that ensured the diversity of demographics. Each participant was asked to follow the experiment steps as below (Fig. 2):

1) Preparation: The participant puts on the portable PPG-EDA biofeedback device<sup>②</sup> on forefinger and middle finger, and walks around while the researchers introducing them the experimental procedures and precautions;

2) Beginning: The participant sits quietly in the waiting area for 10 minutes<sup>③[53][54]</sup> to get calm down, then their initial EDA data are collected;

3) Stressor: The participant exposes to a noise in a closed car and does maths for 3 minutes<sup>④</sup>, which ensures that all participants would be in a similar anxiety status.

4) Experiencing: The participant walks around the site as usual for 5 minutes<sup>⑤[55][56]</sup>.

① The experiment was approved by the Experiment Ethic Committee of the authors' university. All participants joined the experiment voluntarily, and permitted the use of their portraits in the article.

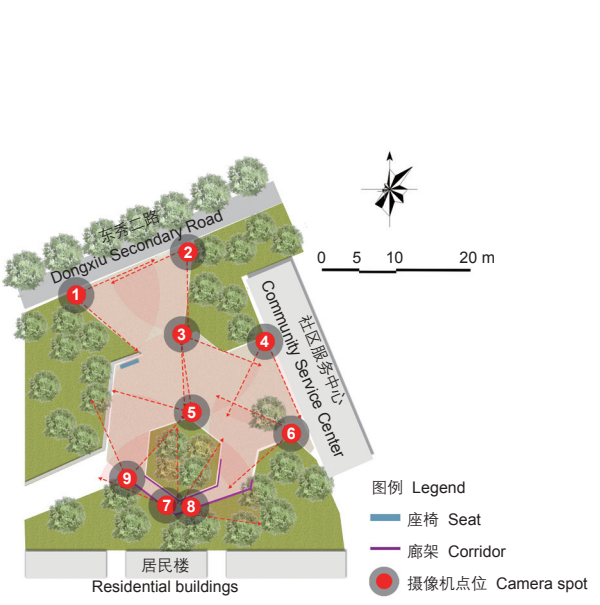
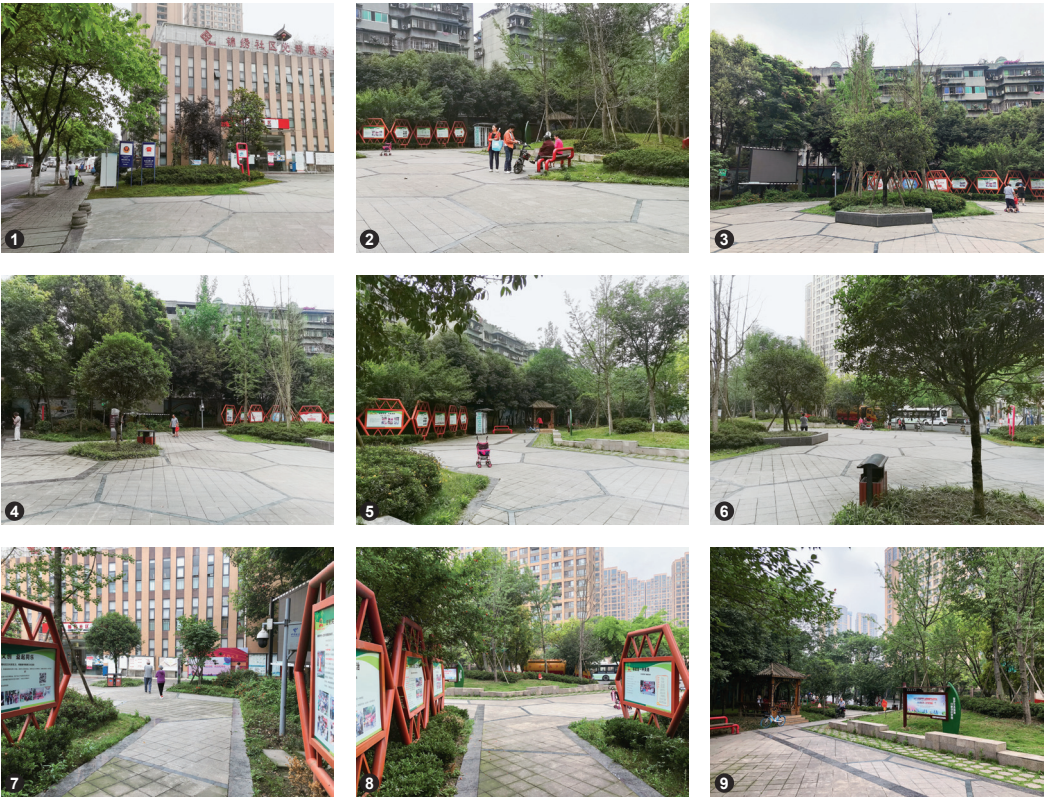
② The devices used in the research is Biopac MP150 Data Acquisition and Analysis System. It little impacts participants' spatial experience during the experiment.

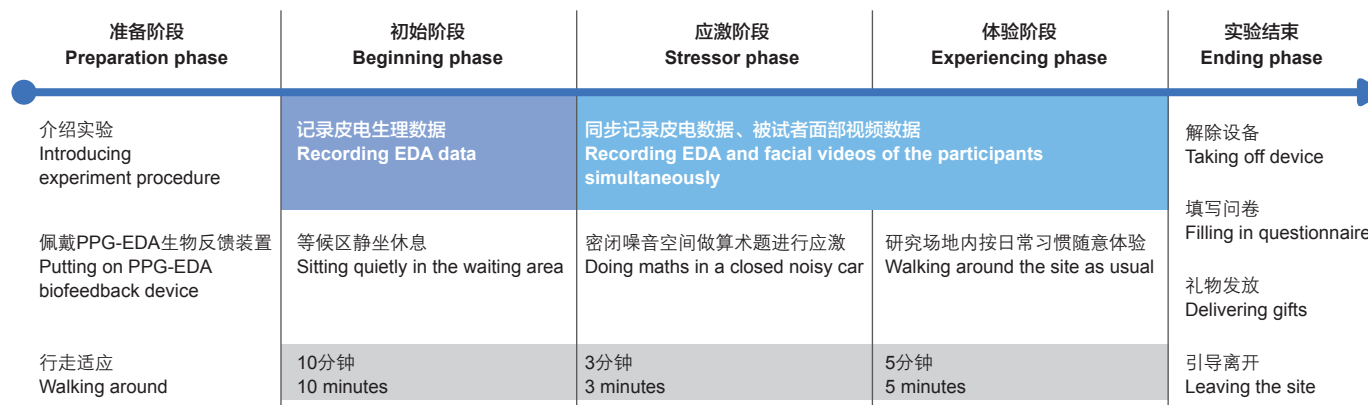
③ The setting of the 10-minutes resting time in the beginning step is referred from previous studies [Sources: Refs. [53][54]].

④ Comparing the effect of stimuli including maths, horror films, and noise in the pre-experiments, the authors found that doing maths in a noisy environment is the most effective stressor.

⑤ Recent findings showed that an environmental experience for 3 ~ 5 minutes is enough to cause physiology changes [Sources: Refs. [55][56]].

表2: 研究场地基本环境情况  
Table 2: The conditions of the Jinxiu Community Square

平面及摄像机位布局 Site plan and the layout of cameras	摄像机位拍摄实景照片示例 Sample photos shot by the cameras		
 <p>图例 Legend</p> <ul style="list-style-type: none"> <li>座椅 Seat</li> <li>廊架 Corridor</li> <li>摄像机点位 Camera spot</li> </ul> <p>居民楼 Residential buildings</p> <p>社区服务中心 Community Service Center</p> <p>东秀二路 Dongxiu Secondary Road</p> <p>0 5 10 20 m</p>			
绿化情况 Greenery	植被: 乔木、灌木、草本、地被 Vegetation: trees, shrubs, grass, and ground cover	主要环境设施 Facilities	座椅、路灯、垃圾箱、廊架、电子显示屏、建筑、墙绘 Seats, lights, bins, corridor, LED screen, buildings, and mural
	种植设施: 微地形、种植池 Planting facilities: micro-topography and planting beds	铺装情况 Paving	硬质铺装、草坪、植草砖 Pavement, lawn, and grass-planting brick



2  
© 付而康, 周佳政, 姚智, 李西

2. 单人次实验流程图  
2. Experiment procedure for individual participants

5) 实验结束: 被试者需填写问卷, 并在研究人员协助下解除设备、离开场地。

单人次实验流程需20分钟左右, 摄像机从实验初始阶段开始录制视频, 直至实验结束。本实验获取被试者在研究场地中进行环境体验的同步皮电数据 (Biopac MP150便携式穿戴生理记录分析系统中的PPG-EDA模块检测皮肤电SCR变化值)、视频人脸数据和问卷调查数据共42组。

### 3.2 实验过程及测度结果

#### 3.2.1 情绪定义与数据集准备

机器视觉识别需要对数据有较为准确的情绪定义才能进行视频情绪分析。既有研究中对情绪定义的方式主要有三种: 1) 直接激发被试者产生某种特定情绪并以此定义; 2) 在实验后记录被试者的情绪自我描述完成定义; 3) 根据被试者在实验中的生理数据反映完成定义<sup>[57]</sup>。前两种定义方法中分别引入了理想化判断及主观判断, 因为被试者在指定实验条件刺激下不一定能激发期望情绪, 而人们在情绪自我描述时由于具有先验条件又可能会产生错误描述; 且前两种方法只有在持续受激的条件下才能完成情绪定义。因此, 本研究以实验获取的视频人脸数据为对象, 用同步采集的被试者皮电生理数据 (第三种方法) 对其进行情绪定义。

And 5) Ending: The participant fills in a questionnaire and leaves the site before removing the device with help of the researchers.

Each individual experiment took about 20 minutes, and the cameras recorded videos from the beginning to the ending. A total of 42 sets of real-time EDA data (the SCR value changes detected by PPG-EDA module in Biopac MP150 Data Acquisition and Analysis System), facial image data, and questionnaire data were collected.

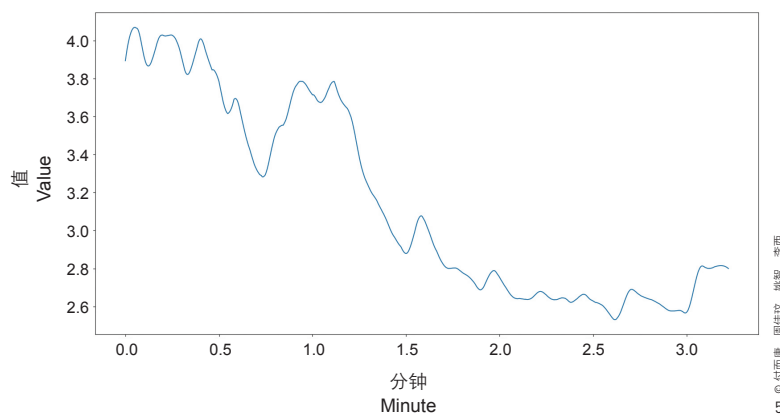
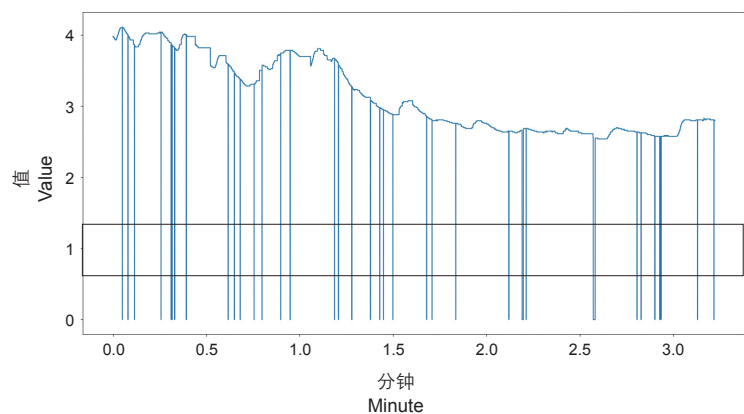
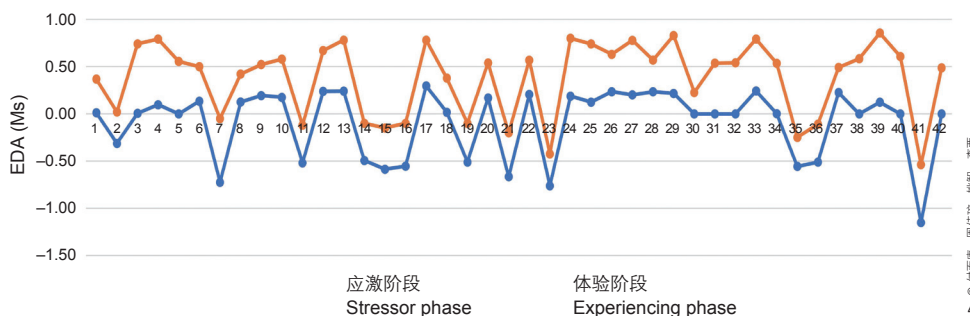
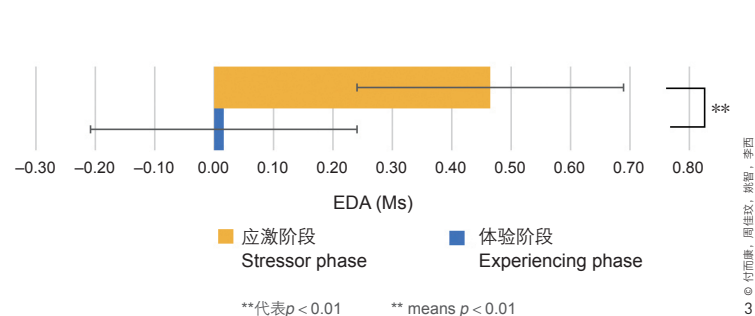
### 3.2 Experiment Process and Measurement Results

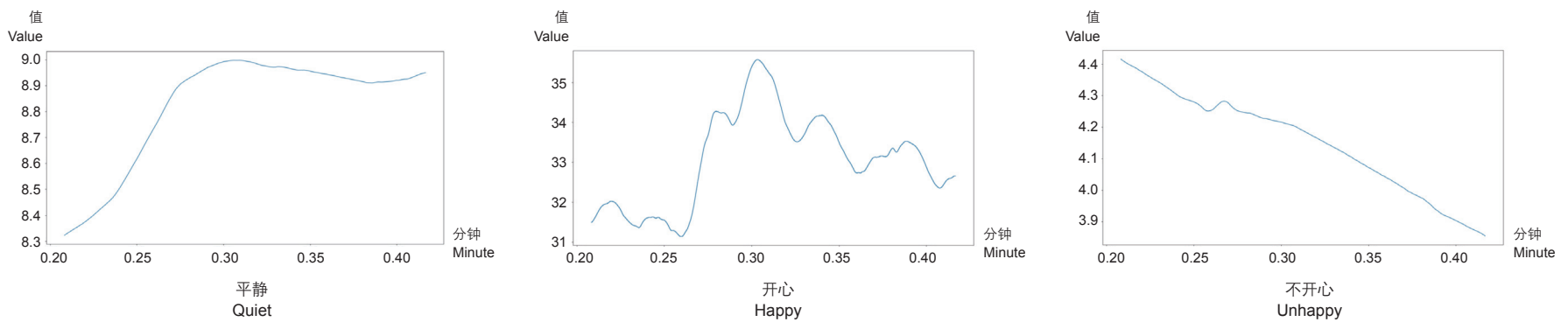
#### 3.2.1 Emotion Labeling and Data Set Preparation

Emotion labeling is required for the emotion analysis on the videos via machine visual cognition. There are three emotion labeling ways demonstrated in existing studies: 1) directly stimulate participants to generate a specific emotion and label it; 2) self-report by participants after experiments; and 3) label participants' emotion according to physiological data during the experiment<sup>[57]</sup>. The first is too ideal since the expected emotion maybe cannot generate under experimental conditions, and the second is poor in credibility because the participants may misrepresent due to priori knowledge. Also, for the both, only constant stressors could emotions be successfully labeled. This research thus adopts the third method, i.e. labeling emotion by the collected real-time EDA data.

3. 被试者应激阶段和体验阶段皮电数据对比
4. 被试者应激阶段和体验阶段皮电数据配对样本t检验
5. 去噪平滑处理前 (左图) 后 (右图) 被试者皮电数据波动图示例

3. EDA data comparison between the Stressor and Experiencing phases
4. The matched samples t-test of EDA data at the Stressor and Experiencing phases
5. A sample of EDA data fluctuations before (left) and after (right) denoise and smooth





6. 本研究情绪定义模式图样  
6. The samples of emotion labels in this study

首先，对皮电数据在空间体验前后的均值进行配对样本 $t$ 检验（应激阶段 $0.02 \pm 0.38$ ，体验阶段 $0.47 \pm 0.13$ ， $p < 0.01$ ，小于 $\alpha = 0.05$ ），证明本研究选取的样本空间具有显著的情绪缓释效用（图3，4）。然后，使用Savitzky-Golay滤波器，采用就近原则将每位被试者的皮电数据进行去噪平滑处理（图5），并参考相关文献中监督实验所得情绪与皮电数据模型的峰值匹配方法<sup>[57]</sup>将本研究中的情绪分为三类：将皮电数据几乎无波动的数据模式定义为“平静”情绪；将15秒区间内呈现密集且多段波动幅度适中的数据模式定义为表积极特征的“开心”情绪；而呈现突然性波动的数据模式则被定义为“不开心”情绪（图6）。

最后，设定每15秒长度的视频为一个面部表情，每个面部表情之间有5秒间隔。实验模型训练使用的数据集通过OpenCV软件库中的CascadeClassifier级联分类器获得，该分类器可以获取图像的哈尔特征，而后对图像中的人脸区域进行标定裁剪（在人脸图像角度不超过 $90^\circ$ 的情况下亦可被裁剪，不存在图像丢失问题）。由此共获得具有情绪定义的面部图片57 600张，作为后续模型训练的数据集。

### 3.2.2 情绪特征模型训练

#### (1) 模型定义

最初的人脸图像由固定大小的像素矩阵构成，当该矩阵展开为一维矩阵时，会导致矩阵数据量较大，很难直接找到同类与不同类图像之间的关联与区别。本研究选择以CNN为基础的编解码器模型结构——编解码器—SVM优化模型（图7），通过卷积与反卷积配合完成特征的提取与拟合。其中，卷积可利用不同的卷积核提取图像不同的纹理特征进而实现图像的分类，而反卷积则是以恢复原图像为目的进行卷积的逆过程。编码器通过卷积提取的特征向量再输入反卷积构成的解码器，而后通过网络训练，直至输出图像与输入图像之间的交叉熵不断减小至平稳，可认为此时编解码器中间的特征向量最能体现图像特征。该图像特征向量不仅具有代表性还具有生成性，而普通CNN训练得到的图像特征向量则缺少了生成特性。因此，本实验选用编解码器—SVM优化模型进

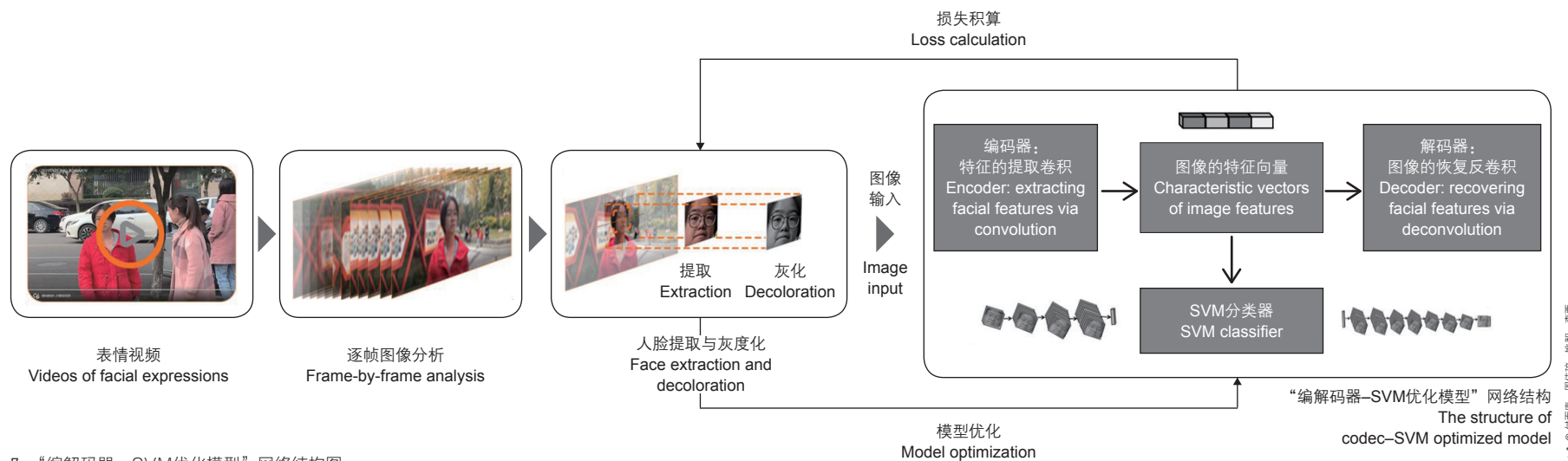
To begin with, a matched samples  $t$ -test of the mean values of EDA data before and after Experiencing phase is conducted ( $0.02 \pm 0.38$  at the Stressor phase and  $0.47 \pm 0.13$  at the Experiencing phase,  $p < 0.01$  less than  $\alpha = 0.05$ ), proving a strong significance of the study site to relieve stress (Fig. 3, 4). Then, the Savitzky-Golay Filter is employed to denoise and smooth each participant's EDA data under the proximity principle (Fig. 5). The peak values are paired between the emotional and EDA data models referring from previous experiments<sup>[57]</sup>. Finally, the emotions are divided into three types in this study: 1) Being almost no fluctuations in EDA data, the emotions are labeled as “calm”; 2) Being intensive moderate fluctuations in multiple segments within 15 seconds, the emotions are labeled as “happy”; And 3) being an abrupt fluctuation, the emotions are labeled as “unhappy” (Fig. 6).

Each 15-second segment is viewed as a facial expression, with a 5-second interval. The dataset used for the model training is obtained with the CascadeClassifier of OpenCV, which can acquire the harr-like features of the images and then perform calibrated cropping of the face areas (all images can be cropped as long as the angle of the input face images is less than  $90^\circ$ ). Finally, a total of 57,600 face images with emotional labels are obtained which are used as the dataset for subsequent model training.

### 3.2.2 Emotional Facial Feature Model Training

#### (1) Model definition

The original face image consists of a fixed-size pixel matrix. When the matrix is rendered into a one-dimensional matrix, voluminous matrix data would be generated, which makes it difficult to identify the similarity and disparity among images of the same/different classes. The codec-SVM optimized model based on CNN (Fig. 7) is introduced into this research to extract and fit the facial features via convolution and deconvolution. Convolution classifies the images by extracting different texture features of images using varied convolutional kernels. In turn, deconvolution is a process of recovering the original images. The characteristic vectors extracted by the encoder via convolution are then fed into the decoder developed via deconvolution. Then the network training repeated until the cross entropy between the output and the input images demonstrate a



7. “编解码器 - SVM优化模型”网络结构图  
7. The structure of codec-SVM optimized model

行图像特征提取以及之后的模型训练，以此为基础模型结构可实现人脸<sup>⑥</sup>差异较大情况下的情绪识别，更适用于人群特征多样的户外环境情绪测度<sup>[58]</sup>。

## (2) 模型训练

由于本研究在没有先验标签的情况下进行模型训练，需要利用基于Savitzky-Golay滤波器得到的类别标签与该模型训练得到的标签作验证，以获得最终的识别率。为了保证两项实验者数量一致，模型训练集和测试集的比例为5:5，即从数据集中随机抽取21位被试者的面部图像作为训练集，剩余21位被试者的面部图像为测试集。应用编解码器-SVM优化模型进行图片特征提取和智能识别，期间不断优化算法参数以提高识别率。每次同时输入32张 $112 \times 112$ 维的图像进行同步训练，设定学习率为 $1e-3$ 。由于涉及训练解码器所对应的模型生成特征，模型的训练部分耗时较长，过程如图8所示。

本研究模型训练共使用了28 800张图像、耗费了120个时期 (epoch)，得到具有图像特征的中间向量，之后训练SVM分类器以完成模型的训练过程。SVM分类器利用内积核函数代替向高维空间的非线性映射，可以增强模型的鲁棒性，避免了由于多全连接层带来的训练损失以及过拟合<sup>[59]</sup>。训练过程如下：首先调整核函数 $\gamma$ 的选取，最高识别率0.9771发生在 $\gamma$ 为7~8.5范围时，微调后三种分类核函数（即Gauss函数、Poly函数和Sigmoid函数）对应的最佳 $\gamma$ 值分别为8.935、8.935和2（图9）。在SVM分类过程中不同的核函数对应了不同的分类模型，在输入数据特征未知的情况下，实验尝试对不同的核函数及其取值进行对比，以寻找具有更高分类准确率的编解码器-SVM优化模型。在此基础

⑥ 视频提取人脸时存在正、侧面多种情形，该优化模型可以保证正、侧脸均有对应图像输入，也能自动检测相应的角度纹理，并在图像输出过程中还原正、侧脸图像。

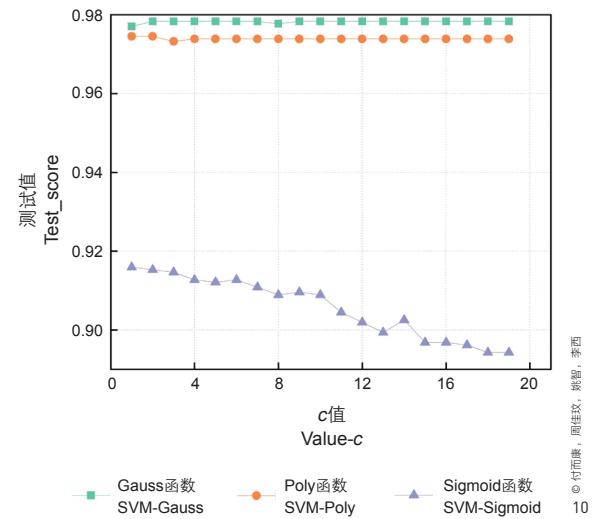
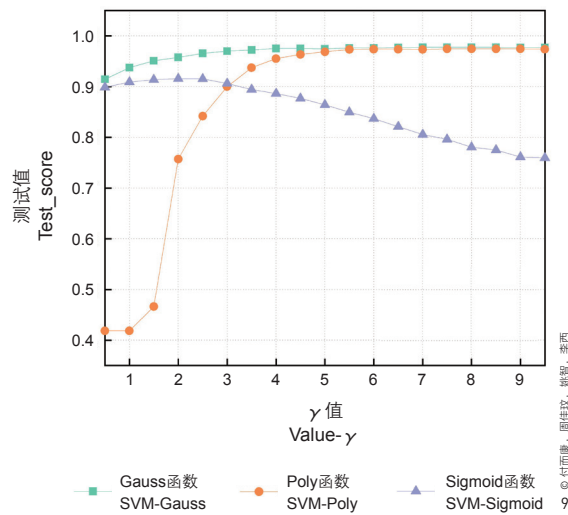
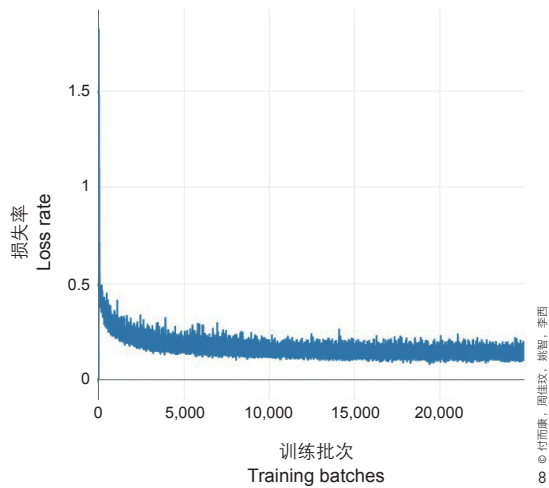
⑥ The optimized model can solve the problem of frontal and side human faces detection via corresponding image input, and automatically detect angle texture so as to recover the face images in the output process.

steady pattern, i.e. when the characteristic vectors in the codec can embody the image characteristic. Combining representation and generation, these characteristic vectors outperform the ones obtained from ordinary CNN training that supports representation only. The codec-SVM optimized model used for image characteristic extraction and subsequent model training<sup>[58]</sup> can realize the recognition of facial expressions<sup>⑥</sup> for the measurement of affective experience in varying outdoor environments with diverse population characteristics.

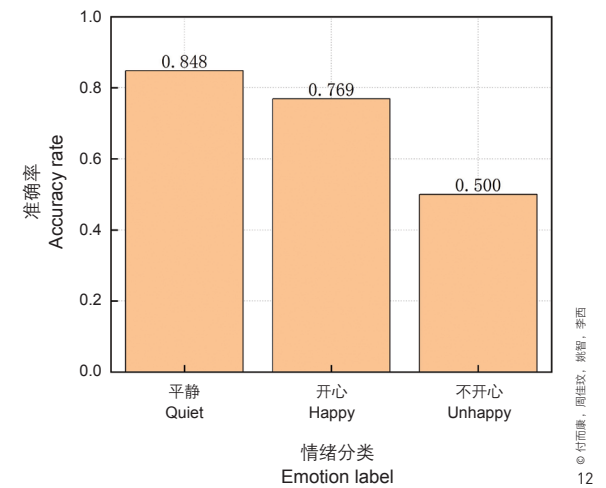
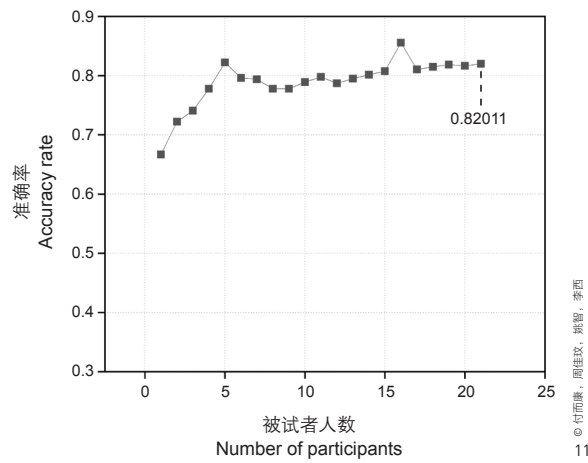
## (2) Model training

As the model training is not trained with prior labels from previous studies, the final recognition accuracy rate is calculated as the labels from emotion labels by Savitzky-Golay Filter to those of the trained model in this research. To ensure the same participants sizes of the model training set and the model testing set, the face images of 21 participants are randomized into the training set and the testing set each. The codec-SVM optimized model extracts and recognizes face images while improving its recognition accuracy by iterating algorithm parameters. The learning rate set as  $1e-3$ , a total of  $32 \ 112 \times 112$ -dimensional images are used for the training each time. The model training takes a relative long time due to the model generation characteristics by the decoder (Fig. 8).

Taking 120 epochs, a total of 28,800 images are used to obtain the intermediate vectors with image characteristics and to train the SVM classifier. In the SVM classifier, the nonlinear mapping to high-dimensional space is replaced with the inner kernel function, so as to enhance the robustness of the model and avoid the training loss and overfitting caused by multiple fully connected layers<sup>[59]</sup>. In the training process, the kernel function  $\gamma$  is adjusted: the recognition accuracy of 0.9771 occurs when  $\gamma$  ranges from 7 to 8.5; and after the fine adjustment, the best  $\gamma$  values of the three kernel functions (i.e. Gaussian Function, Poly Function, and Sigmoid Function) are 8.935, 8.935, and 2, respectively (Fig. 9). During the process of SVM classification, different kernel functions match with varied classification models. As the characteristics of the input data are unknown, the experiment



8. 模型训练过程图
  9. SVM分类器在训练集上的识别准确度随不同核函数类型以及  $\gamma$  值的变化
  10. SVM分类器在确定了  $\gamma$  值后识别准确度随  $c$  值的变化
  11. 测试集上机器视觉情绪识别准确率
  12. 测试集的情绪分类识别准确率
8. Model training
  9. Changes of recognition accuracy of SVM classifier on the training set correlating with kernel function types and  $\gamma$  values
  10. Changes of recognition accuracy of SVM classifier correlating with  $c$  values after the  $\gamma$  value is set
  11. Emotion recognition accuracy rate of the testing set via machine visual cognition
  12. Recognition accuracy rate of each emotion labels of the testing set



上实验调整损失系数  $c$  获得最终的情绪识别模型 (图10), 最终确定在核函数为 Gauss 函数分布模式下  $\gamma$  值为 8.935,  $c$  为 15, 识别准确度为 0.9783。

### 3.2.3 视频情绪识别分析

运用训练好的模型对人脸图像进行表情预测, 将数据集中剩余 21 位被试者的面部图像作为测试集进行视频情绪识别分析<sup>[60][61]</sup> (共 28 800 张, 未添加情绪标签)。其中图像的特征提取与智能识别借助 CascadeClassifier 级联分类器完成 (延展规模设置为  $160 \times 160$  像素, 比例系数为 7, 检测的最小尺寸和最大的尺寸分别  $100 \times 100$  像素和  $1\ 000 \times 1\ 000$  像素), 视频的分辨率为每秒 30 帧, 每一帧图像都会进行一次面部情绪识别。将模型对于视频情绪识别的结果与同步获取的皮电数据匹配, 最终准确率为 82.01% (图 11), 其中, “平静” “开心” “不开心” 三种情绪类别对应的样本数量分别为 105、78、6 (图 12)。

compares different kernel functions and their values to improve the classification accuracy of the codec-SVM optimized model. The loss coefficient  $c$  is then adjusted to obtain the final emotion cognition model (Fig. 10). It is concluded that the  $\gamma$  value is 8.935,  $c$  15, and recognition accuracy 0.9783 under the Gaussian Function.

### 3.2.3 Emotion Recognition and Analysis of Video Images

To predict facial expressions via the trained model, the 28,800 face images of the rest 21 participants with no emotion labels are used as the testing set<sup>[60][61]</sup>. The feature extraction and recognition of the images are realized with the CascadeClassifier (the scale is set as  $160 \times 160$  px, the factor 7, the detected minimum size  $100 \times 100$  px, and  $1,000 \times 1,000$  px for the maximum). The resolution of the video is 30 fps and the facial expressions in each frame are recognized with emotion labels. The matching rate of the recognition results and EDA data is 82.01% (Fig. 11), and the size for the labeled emotion “quiet,” “happy,” and “unhappy” is 105, 78, and 6, respectively (Fig. 12).

### 3.3 拓展实验检验

为了进一步验证此算法模型是否可适应多样的户外环境，2020年9月研究团队选取成都市成华区锦绣社区一处街头绿地进行了拓展实验。该街头绿地面积为 $0.6\text{hm}^2$ ，主要景观要素有植被、健身设施和硬质铺装，其中植被以乔木为主，空间的郁闭度较高。为保证全方位获取视频人脸数据，实验共布置了8个摄像机位（图13，14）。招募21位社区居民作为被试者，按照和图2相同的实验步骤获取21组数据。运用本研究得到的编解码器—SVM优化模型对21位被试者的视频数据进行情绪分析，结果显示和同步获取的皮电数据匹配准确率为65.08%（图15）。其中，“平静”“开心”“不开心”三种情绪类别对应的样本数量分别为122、60、7（图16）。

## 4 讨论

### 4.1 结果分析

本研究提出的编解码器—SVM优化模型在两类不同景观特征的户外环境情绪测度中准确率分别达到82.01%和65.08%，说明机器视觉识别对于测度户外环境中的人群情绪具有较好的应用潜力；相较于既有的情绪感受测度方法，该模型具有实时性强、效率高等优势，可进一步在面积较大、样本量较多的户外环境中测度情绪感受。此外，机器视觉识别提高了情绪测度的智能化程度，未来可考虑将其应用于沟通或阅读障碍人群的情绪识别研究中。

此外，根据实证实验及拓展实验的情绪分类结果（图12，16），可以看出“平静”情绪的识别准确率最高，“不开心”情绪的识别准确率相对较低，这可能和测试集的样本数量有关，也从侧面印证了提高机器视觉识别的可靠性需要进行大量样本的训练与测试。

### 3.3 Additional Experiment Testing

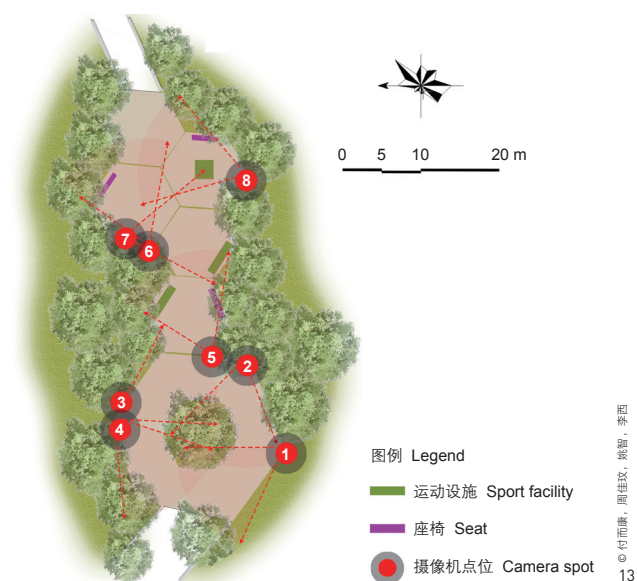
To further test the effectiveness of algorithm model in different outdoor environments, the research team conducted an additional experiment in a street green space in the same district in September 2020. Covering an area of  $0.6\text{hm}^2$ , the study site is covered with vegetation (trees mainly), fitness facilities, and pavements, with a high canopy density. To collect all-round facial data, 8 cameras were placed on the site (Fig. 13, 14). Following the same experiment procedure as the main experiment above, the obtained datasets of 21 participants (community residents) are analyzed by the codec-SVM optimized model. The matching rate between the recognition results and EDA data is 65.08% (Fig. 15). The size for the labeled emotion “quiet,” “happy,” and “unhappy” is 122, 60, and 7, respectively (Fig. 16).

## 4 Discussions

### 4.1 Result Analysis

The accuracy rate of the codec-SVM optimized model developed in this study reaches 82.01% and 65.08% in the measurement of affective experience in two outdoor environments with different landscape characteristics. The results indicate a satisfactory potential of the machine visual cognition in related applications. Compared with existing measurement methods, this technology supports real-time emotion measurement in a larger site area or sample size. In addition, machine visual cognition promotes the intelligence level of affective experience measurement, which would benefit the studies on people facing communication or reading disability.

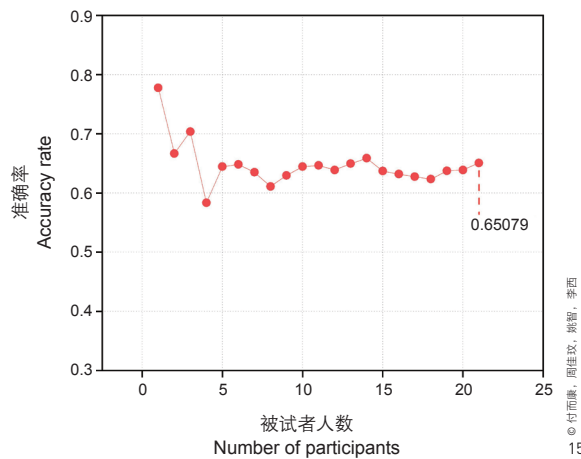
Also, the results of the main and additional experiments (Fig. 12, 16) show that the machine visual cognition has a high recognition accuracy of “quiet” emotion but that of “unhappy” is unsatisfactory—This may be due to the sample size of the testing set, which could somehow explain that the reliability of machine visual cognition depends on intensive training based on a considerable sample size.



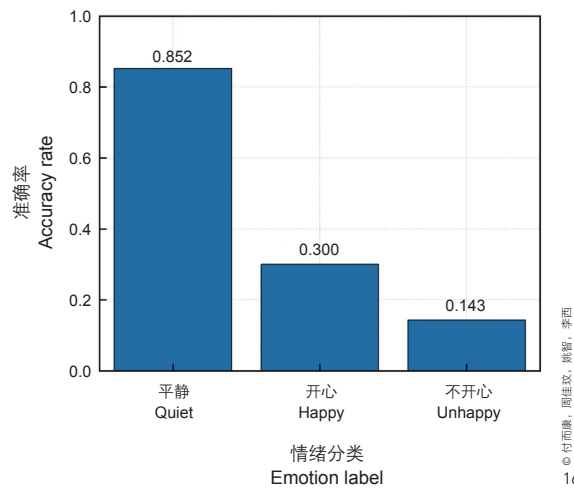
13. 拓展实验平面及摄像机点位布局  
14. 拓展实验空间现状照片

13. Site plan and camera locations on the study site of the additional experiment  
14. Photos of the additional experiment site





15. 拓展实验空间机器视觉识别准确率



16. 拓展实验空间情绪分类识别准确率

15. 拓展实验空间机器视觉识别准确率  
16. 拓展实验空间情绪分类识别准确率

15. Emotion recognition accuracy rate of the additional experiment via machine visual cognition  
16. Recognition accuracy rate of each emotion labels of the additional experiment

而拓展实验结果表明，运用编解码器—SVM优化模型的准确率相较于第一次实验有所下降，造成这一现象的原因可能是户外环境的景观特征差异。一方面，该空间郁闭度较高可能导致光线不足，使数据采集受到一定影响；另一方面，这一街头绿地空间面积较为局促，人群行为活动受限，被试者的平静情绪出现频率较其他情绪高，从而导致其他情绪的识别存在着瞬时性及随机性。不过，通过对比两个实验空间的物质要素构成，可以初步推测该模型在人群行为丰富、视野开阔的空间中更为适用。

#### 4.2 研究不足

研究尚存在诸多不足。主要包括：

1) 模型泛化能力不足，导致拓展实验中环境改变后模型的识别准确率下降。后续研究可以通过对大量样本进行训练、调整模型参数、优化模型结构等方式进行优化，并可针对不同景观特征的户外环境，训练出相应的机器视觉情绪感受识别模型，从而进行不同户外环境下人群面部情绪的分类识别。

2) 本研究将人群情绪分为“平静”“开心”“不开心”三类，情绪分类细粒度水平较低。后续研究可对情绪分类识别进行细化，如增加对厌恶、悲伤、愤怒等情绪的识别。

3) 模型测试时偶尔会出现因被试者低头而无法检测面部表情的情况，然而该部分的匹配结果依旧会被归入测试集的匹配准确度结果中的不准确部分，使准确度可信度降低。后续研究可以通过人脸数据筛选规避该情况的影响。

4) 数据采集可能会受到光线等环境变化的影响，后续研究可以通过优化数据筛选及拍摄方式规避该影响。

5) 模型的实践应用待进一步研讨，虽然该情绪测度模型是基于客观数据构建的，但情绪定义主要依据主观判断，未来研究中可通过与环境感知问卷相结合的方式分析使用者情绪变化的具体原因，从而对环境进行针对性的优化提升。

The recognition accuracy result of the codec-SVM optimized model in the additional experiment is lower than that in the main experiment—This may be down to the disparity of the landscape characteristics of the two sites: the high canopy density of the site may impact data collection to some extent, and the small site area may limit participants’ physical activities that often brings a quiet emotion state, as a result, more transient and random in the cognition results of other emotion states. The finding also suggests that the codec-SVM optimized model is more applicable for scenarios with varying usage behaviors and open views.

#### 4.2 Limitations

This research sees limitations as follows.

1) The fact that the model yields a lower recognition accuracy in the additional experiment suggests that its application capacity needs to be promoted by training the model with a larger number of samples, adjusting model parameters, and optimizing model structure. Also, different machine visual cognition models can be developed for outdoor environments with varied landscape characteristics.

2) The emotion labels in this research are “quiet,” “happy,” and “unhappy.” In the future, more emotion labels (such as disgust, sadness, and anger) need to be studied.

3) During the model test, when the subjects lowered heads, the facial expressions cannot be detected but recorded as the inaccuracy data, which affected the matching accuracy of the testing set. Applying face data screening in future studies to avoid such errors is expected.

4) Outdoor data collection is largely subject to environmental conditions (e.g., sunlight). Such impacts can be minimized by optimizing data screening and camera shooting methods.

And 5) although the codec-SVM optimized model is developed on objective data, the emotion labeling is mostly defined by researchers’ subjective assessment. Other methods such as questionnaire surveys on environmental perceptions can be combined to further analyze the factors that cause people’s emotional changes, which would better inform related environmental design.

## 5 结论

在人工智能的大背景下，机器视觉识别能够辅助景观设计实践、提升景观品质优化的有效性和准确性，在追求精细化服务的智慧城市建设中具有极大应用潜力。本文以城市社区广场环境为研究场地，通过同步采集人群在户外环境体验时的面部视频数据和皮电数据，训练生成了可用于户外环境情绪测度的机器视觉算法模型——编解码器-SVM优化模型，并在街头绿地环境中对模型进行了可靠性检验，实现了机器视觉识别在户外环境情绪感受度中的研究应用，为在场地面积更大、样本量更多的情境下对使用者情绪感受展开大范围测度提供了研究基础。得出如下主要结论：

1) 基于机器视觉识别的情绪测度算法模型（编解码器-SVM优化模型）识别个体面部情绪的结果与同步皮电数据匹配的准确率达82.01%，证明了机器视觉能实现对户外环境中人群情绪感受的识别。

2) 拓展实验的准确率为65.08%，证实了该算法模型的有效性，具备推广应用潜力。

3) 实证实验与拓展实验的户外环境特征差异表明该模型对人群行为丰富、视野开阔的空间具有更好适用性。

4) 机器视觉识别弥补了既有情绪测度方法的不足，同时提高了情绪测度的智能化程度，对特殊人群的情绪测度具有一定优势，为探究环境与人群情绪间的复杂关系提供了良好的技术手段。LAF

## 5 Conclusions

As Artificial Intelligence technology advances, machine visual cognition contributes to landscape design and improves the effectiveness and accuracy of landscape optimization efforts, demonstrating a great potential in the building of Smart Cities with refined public services. Taking an urban community square as the study site, this research simultaneously collects participants' facial data from video images and their EDA data in the outdoor environments, and further develops a machine visual algorithm model, i.e. codec-SVM optimized model, whose reliability is tested through an additional experiment. The study explores the application effectiveness of machine visual cognition in outdoor environments, offering references to other cases of a larger site area or sample size. The main research findings include:

1) The matching rate between the EDA data and the emotion labeling results of the codec-SVM optimized model via machine visual cognition is 82.01%, suggesting that machine vision can be used to recognize affective experience in outdoor environments.

2) The matching rate in the additional experiment is 65.08%, verifying the effectiveness and application potential of the algorithm model.

3) Comparing the environmental differences between the main and additional experiments, the model works more effectively for outdoor scenarios with varying usage behaviors and open views.

And 4) machine visual cognition breaks up the limitations of existing measurement methods and increase the intelligence level of recognition. It could be an instrumental tool to explore the complex relationships between the environments and users' affective experience. LAF

## REFERENCES

- [1] Zhou, W., Nie, X., Chi, M., Zhan, Y., Yan, C., Xiu, X., & Lan, S. (2020). Influence of the Emotional State of Forest Therapy Consumers on the Recovery of Physical and Mental Health. *Forestry Economics*, (09), 53-62. doi:10.13843/j.cnki.lyjj.20201204.001
- [2] Dong, Y., Wang, Q., & Xin, C. (2012). Progress in Research on the Relationship between Positive Emotion and Physical and Mental Health. *Journal of Psychological Science*, (02), 487-493. doi:10.16719/j.cnki.1671-6981.2012.02.043
- [3] Tost, H., Champagne, F. A., & Meyer-Lindenberg, A. (2015). Environmental influence in the brain, human welfare and mental health. *Nature Neuroscience*, 18(10), 1421. doi:10.1038/nn.4108
- [4] Wang, Q., Wang, C., Ren, H., & Li, T. (2013). Urbanized environment and the risk of schizophrenia. *Chinese Journal of Nervous and Mental Diseases*, (12), 758-763. doi:10.3936/j.jissn.1002-0152.2013.12.014
- [5] Peng, H., & Tan, S. (2018). Study on the Influencing Mechanism of Restoration Effect of Urban Park. *Chinese Landscape Architecture*, 034(009), 5-9. doi:10.3969/j.jissn.1000-6664.2018.09.002
- [6] Jiang, B., Zhang, T., William, C. S., & Wu, X. (2015). Healthy Cities: Mechanisms and Research Questions Regarding the Impacts of Urban Green Landscapes on Public Health and Well-being. *Landscape Architecture Frontiers*, 3(1), 24-35. Retrieved from <https://journal.hep.com.cn/laf/EN/article/downloadArticleFile.do?attachType=PDF&id=13393>
- [7] Chen, Z., & Liu, S. (2018). Real-time Environmental Affective Experience Assessment via Wearable Sensors. *Chinese Landscape Architecture*, 34(3), 12-17. doi:10.3969/j.jissn.1000-6664.2018.03.003
- [8] Meng, Z. (2005). *Emotional Psychology*. Beijing, China: Peking University Press.
- [9] Kong, Y. (2014). *The research methods of emotion* (Master's thesis). Retrieved from CNKI database.
- [10] Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070. doi:10.1037/0022-3514.54.6.1063
- [11] McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *EITS Manual for the Profile of Mood States*. San Diego, CA: Educational & Industrial Testing Services.
- [12] Zheng, X., & Li, Y. (1997). State-Trait Anxiety Inventory. *Chinese Mental Health Journal*, 11(4), 219-220.

- [13] Chen, Z., Schulz, S., Qiu, M., Yang, W., He, X., Wang, Z., & Yang, L. (2018). Assessing affective experience of in-situ environmental walk via wearable biosensors for evidence-based design. *Cognitive Systems Research*, (52), 970-977. doi:10.1016/j.cogsys.2018.09.003
- [14] Rad, P. N., Shahroudi, A. A., Shabani, H., Ajami, S., & Lashgari, R. (2019). Encoding Pleasant and Unpleasant Expression of the Architectural Window Shapes: An ERP Study. *Frontiers in Behavioral Neuroscience*, (13), 186. doi:10.3389/fnbeh.2019.00186
- [15] Xie, J., Fang, P., & Jiang, Y. (2011). Advances in Emotion Measuring Methods. *Journal of Psychological Science*, (02), 488-493. doi:10.16719/j.cnki.1671-6981.2011.02.048
- [16] Xu, X., Zhao, W., & Liu, H. (2018). Research on Application and Model of Emotional Analysis in Blended Learning Environment: From Perspective of Meta-analysis. *E-education Research*, (08), 70-77. doi:10.13811/j.cnki.eer.2018.08.011
- [17] Chen, C., & Zhang, H. (2018). Using emotion to evaluate our community: Exploring the relationship between the affective appraisal of community residents and the community environment. *Architectural Engineering and Design Management*, 14(3-4), 1-16. doi:10.1080/17452007.2018.1457942
- [18] Li, D., Deal, B., Zhou, X., Slavenas, M., & Sullivan, W. C. (2018). Moving beyond the neighborhood: Daily exposure to nature and adolescents' mood. *Landscape and Urban Planning*, (173), 33-43. https://doi.org/10.1016/j.landurbplan.2018.01.009
- [19] Chen, Z., Dong, N., Liu, S., Zhang, Y., & Ding, Q. (2017). Study on the Impacts of Urban Park Use on Public Health in Shanghai. *Landscape Architecture*, (09), 99-105. doi:10.14085/j.fjyl.2017.09.0099.08
- [20] Liu, C., Li, S., & Chen, S. (2018). Study on Behavior of Visiting Campus Green Space' Role in Emotion Regulation under the Influence of Multi-factors: Take 3 Universities in Beijing for Example. *Landscape Architecture*, (03), 46-52. doi:10.14085/j.fjyl.2018.03.0046.07
- [21] Lin, W., Chen, Q., Jiang, M., Zhang, X., & Zeng, Q. (2019). The effect of green space behaviour and per capita area in small urban green spaces on psychophysiological responses. *Landscape and Urban Planning*, (192), 103637. https://doi.org/10.1016/j.landurbplan.2019.103637
- [22] Lin, W., Chen, Q., Jiang, M., Tao, J., & Zeng, Q. (2020). Sitting or Walking? Analyzing the Neural Emotional Indicators of Urban Green Space Behavior with Mobile EEG. *Journal of Urban Health*, 97(4), 191-203. https://doi.org/10.1007/s11524-019-00407-8
- [23] Liu, B., & Xu, Y. (2016). Study on the Effects of Different Landscapes on Elderly People's Body-Mind Health. *Landscape Architecture*, (7), 113-120. doi:10.14085/j.fjyl.2016.07.0113.08
- [24] Weng, Y., Zhu, Y., Dong, J., Wang, M., & Dong, J. (2021). Effects of Soundscape on Emotion and Attention on Campus Green Space—A Case Study of Fujian Agriculture and Forestry University. *Chinese Landscape Architecture*, 37(2), 88-93. doi:10.19775/j.cla.2021.02.0088
- [25] Ge, Y., Chen, Y., Liu, Y., Li, W., & Sun, X. (2014). Electrophysiological Measures Applied in User Experience Studies. *Advances in Psychological Science*, 22(6), 959-967. doi:10.3724/SP.J.1042.2014.00959
- [26] Ekman, P., Freisen, W. V., & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology*, 39(6), 1125-1134.
- [27] Lengen, C. (2015). The effects of colours, shapes and boundaries of landscapes on perception, emotion and mentalising processes promoting health and well-being. *Health & Place*, (35), 166-177. https://doi.org/10.1016/j.healthplace.2015.05.016
- [28] Sinclair, R. R., Wang, M., & Tetrick, L. E. (2014). *Research Methods in Occupational Health Psychology: Measurement, Design and Data Analysis*. Nanjing, China: Southeast University Press.
- [29] Zhu, Y., Lin, Z., & Zhang, Y. (2020). Research progress and prospect of machine vision technology. *Journal of Graphics*, 41(6), 871-890. doi:10.11996/JG.j.2095-302X.2020060871
- [30] Hu, W., & Hu, H. (2020). Disentangled Spectrum Variations Networks for NIR-VIS Face Recognition. *IEEE Transactions on Multimedia*, 22(5), 1234-1248. doi:10.1109/TMM.2019.2938685
- [31] Grant, J. M., & Flynn, P. J. (2017). Crowd Scene Understanding from Video: A Survey. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 13(2), 1-23. https://doi.org/10.1145/3052930
- [32] Gerónimo, D., López, A. M., Sappa, A. D., & Graf, T. (2010). Survey of Pedestrian Detection for Advanced Driver Assistance Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(7), 1239-1258. doi:10.1109/TPAMI.2009.122.
- [33] Wei, J., He, J., Zhou, Y., Chen, K., Tang, Z., & Xiong, Z. (2019). Enhanced Object Detection with Deep Convolutional Neural Networks for Advanced Driving Assistance. *IEEE Transactions on Intelligent Transportation Systems*, 21(4), 1572-1583. doi:10.1109/ITITS.2019.2910643
- [34] Parmar, C., Barry, J. D., Hosny, A., Quackenbush, J., & Aerts, H. J. (2018). Data Analysis Strategies in Medical Imaging. *Clinical Cancer Research*, 24(15), 3492-3499. doi:10.1158/1078-0432.CCR-18-0385
- [35] Yu, Z., Jiang, X., Zhou, F., Qin, J., Ni, D., Chen, S., ... Wang, T. (2019). Melanoma Recognition in Dermoscopy Images via Aggregated Deep Convolutional Features. *IEEE Transactions on Biomedical Engineering*, 66(4), 1006-1016. doi:10.1109/TBME.2018.2866166
- [36] Mehrabian, A. (1968). Communication Without Words. In C. D. Mortensen (Ed.), *Communication Theory* (p. 8). New York, NY: Routledge.
- [37] Li, G., Liu, B., Li, K., & Huang, S. (2019). Research on Face Emotion Recognition Based on Machine Learning. *Computer Technology and Development*, 29(5), 27-31. doi:10.3969/j.issn.1673-629X.2019.05.006
- [38] Schmidt, T., Newcombe, R., & Fox, D. (2017). Self-Supervised Visual Descriptor Learning for Dense Correspondence. *IEEE Robotics and Automation Letters*, 2(2), 420-427. doi:10.1109/LRA.2016.2634089
- [39] Xu, L., Zhang, S., & Zhao, J. (2017). Summary of facial expression recognition methods based on image. *Journal of Computer Applications*, 37(12), 3509-3516, 3546. doi:10.11772/j.issn.1001-9081.2017.12.3509
- [40] Yan, Y., & Zhang, Y. (2009). State-of-the-Art on Video-Based Face Recognition. *Chinese Journal of Computers*, 32(5), 878-886. doi:10.3724/SP.J.1018.2009.00878
- [41] Putra, W. B., & Arifin, F. (2019). Real-Time Emotion Recognition System to Monitor Student's Mood in a Classroom. *Journal of Physics: Conference Series*. Advance online publication. doi:10.1088/1742-6596/1413/1/012021
- [42] Kim, P. W. (2018). Real-time bio-signal-processing of students based on an intelligent algorithm for internet of things to assess engagement levels in a classroom. *Future Generation Computer Systems*, (86), 716-722. doi:10.1016/j.future.2018.04.093
- [43] Kang, Y., Jia, Q., Gao, S., Zeng, X., Wang, Y., Angsuesser, S., ... Fei, T. (2019). Extracting human emotions at different places based on facial expressions and spatial clustering analysis. *Transactions in GIS*, 23(3), 450-480. doi:10.1111/tgis.12552
- [44] Qiao, L., Zhuang, J., Zhang, X., Su, Y., & Xia, Y. (2021). Assessing Emotional Responses to the Spatial Quality of Urban Green Spaces through Self-Report and Face Recognition Measures. *International Journal of Environmental Research and Public Health*, 18(16), 8526. doi:10.3390/ijerph18168526
- [45] Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1), 39-58. doi:10.1109/TPAMI.2008.52
- [46] Chen, Z., & Zhu, X. (2021). Research on Automatic Emotion Recognition for Learners based on Facial Expression: Relevance, Research Situation, Existing Problems and Development Paths. *Journal of Distance Education*, 37(4), 64-72. doi:10.15881/j.cnki.cn33-1304/g4.2019.04.008
- [47] Shen, J. (2019). *Research on Driver's Emotion Recognition Algorithm Based on Machine Vision in Complex Illumination Environment* (Doctoral dissertation). Retrieved from CNKI database.
- [48] Kumar, N., & Sukavanam, N. (2019). An improved CNN framework for detecting and tracking human body in unconstrained environment. *Knowledge-Based Systems*, (193), 105198. doi:10.1016/j.knosys.2019.105198
- [49] Tang, L., Gao, C., Chen, X., & Zhao, Y. (2019). Pose detection in complex classroom environment based on improved faster R-CNN. *IET Image Processing*, 13(3), 451-457. doi:10.1049/iet-ipr.2018.5905
- [50] Chen, T., Li, X., Grosse, R., & Duvenaud, D. (2018). Isolating Sources of Disentanglement in Variational Autoencoders. *CoRR*. Advance online publication. Retrieved from https://arxiv.org/pdf/1802.04942.pdf
- [51] Burgess, C. P., Higgins, I., Pal, A., Matthey, L., Watters, N., Desjardins, G., & Lerchner, A. (2018). Understanding disentangling in  $\beta$ -VAE. *CoRR*. Advance online publication. Retrieved from https://arxiv.org/pdf/1804.03599.pdf
- [52] Guo, S., Zhao, N., Zhang, J., Xue, T., Liu, P., Xu, S., & Xu, D. (2017). Landscape visual quality assessment based on eye movement: College student eye-tracking experiments on tourism landscape pictures. *Resources Science*, 39(6), 1137-1147. doi:10.18402/resci.2017.06.13
- [53] Tsunetsugu, Y., Lee, J., Park, B. J., Tyrvaainen, L., Kagawa, T., & Miyazaki, Y. (2013). Physiological and psychological effects of viewing urban forest landscapes assessed by multiple measurements. *Landscape and Urban Planning*, (113), 90-93. https://doi.org/10.1016/j.landurbplan.2013.01.014
- [54] Joung, D., Kim, G., Choi, Y., Lim, H., Park, S., Woo, J. -M., & Park, B. -J. (2015). The Prefrontal Cortex Activity and Psychological Effects of Viewing Forest Landscapes in Autumn Season. *International Journal of Environmental Research and Public Health*, 12(7), 7235-7243. doi:10.3390/ijerph120707235
- [55] Barton, J. O., & Pretty, J. (2010). What is the best dose of nature and green exercise for improving mental health? A multi-study analysis. *Environmental Science & Technology*, 44(10), 3947. doi:10.1021/es903183r
- [56] Ulrich, R. S., Simons, R. F., Losito, B. D., Fiorito, E., Miles, M. A., & Zelson, M. (1991). Stress recovery during exposure to natural and urban environments. *Journal of Environmental Psychology*, 11(3), 201-230. https://doi.org/10.1016/S0272-4944(05)80184-7
- [57] Zhou, Y. (2012). *The Research of the Emotional Feature Extraction and Recognition Classification of GSR* (Master's thesis). Retrieved from CNKI database.
- [58] Gao, K., Sun, S., Yao, G., & Zhao, H. (2017). Semantic segmentation of night vision images for unmanned vehicles based on deep learning. *Journal of Applied Optics*, 38(3), 421-428. doi:10.5768/JAO201738.0302007
- [59] Gayathri, N., & Mahesh, K. (2020). Improved Fuzzy-Based SVM Classification System Using Feature Extraction for Video Indexing and Retrieval. *International Journal of Fuzzy Systems*, 22(5), 1716-1729. https://doi.org/10.1007/s40815-020-00884-z
- [60] Shan, L., & Deng, W. (2018). Deep Facial Expression Recognition: A Survey. *IEEE Transactions on Affective Computing*. Advance online publication. doi:10.1109/TAFFC.2020.2981446
- [61] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., & Matthews, I. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops* (pp. 94-101). New York, NY: IEEE. doi:10.1109/CVPRW.2010.5543262