驾驶分心的影响因素：车载触摸屏交互视觉需求的可解释预测

Patrick Ebel[a], *, Christoph Lingenfelder[b] 和AndreasVogelsang[a]

[a]科隆大学，德国科隆[b]*MBition GmbH*，德国柏林

arXiv:2301.02065v1 [cs.HC] 5 Jan 2023

文章信息

关键词：驾驶分心 视觉需求 车载信息系统 自然驾驶数据 车载触摸屏交互

摘 要

随着现代信息娱乐系统的发展，驾驶员在驾驶时越来越容易被诱惑去执行次要任务。分心驾驶已是造成致命事故的主要原因之一，因此车载触摸屏人机界面（HMI）必须尽可能减少对驾驶员的干扰。为确保这些系统使用安全，需要对其进行复杂且昂贵的实证测试，而这要求先制作功能完备的原型。因此，能够在早期阶段告知设计师其设计可能对驾驶分心产生何种影响的方法极具价值。本文提出一种机器学习方法，该方法基于预期使用场景，预测车载触摸屏交互的视觉需求，并对影响驾驶员视觉注意力分配的因素提供局部和全局解释。该方法基于从量产车辆中持续收集的大规模自然驾驶数据，并采用 SHapley Additive exPlanation（SHAP）方法提供解释，以助力制定有依据的设计决策。与相关研究相比，我们的方法准确率更高；它能以 68% 的准确率识别发生长时间注视的交互，并以 2.4 秒的均值误差预测总注视时长。我们的解释重现了近期多项研究的结果，并快速且易于获取地揭示用户界面元素、驾驶自动化和车速对驾驶分心的影响。该系统不仅能帮助设计师评估当前设计，还能帮助他们更好地预判和理解其设计决策可能对未来设计产生的影响。

1. 引言

如今，大型中控台触摸屏（例如特斯拉 Model 3[1] 或梅赛德斯-奔驰EQS[2]上的触摸屏）是驾驶员与车载信息系统（IVISs）之间的主要交互界面。在交互过程中，驾驶员需要将视线从道路上移开，以查看屏幕上显示的信息。因此，他们需要在主要驾驶任务和次要触摸屏任务之间分配视觉注意力。这会显著增加碰撞事故风险（Dingus等人，2016；Green，1999），尤其是在视线离开道路超过两秒时（Klauer等人，2006）。随着车载信息娱乐系统变得日益复杂、集成的功能不断增加，驾驶员在驾驶过程中与其交互的选择也比以往更多。驾驶自动化功能不断改进，也进一步增加了驾驶员从事非驾驶相关任务的诱因。在部分自动化驾驶期间，驾驶员往往更频繁地从事非驾驶相关任务，尽管他们仍应监控车辆（Carsten 等，2012；de Winter等人，2014；Ebel 等，2022）。为确保车载信息娱乐系统使用安全，这些系统须遵守严格的法规和详尽的测试规程。汽车原始设备制造商（OEMs）会在人工环境（如驾驶模拟器）中开展成本高昂的实证用户研究，以测试系统的安全性。然而，驾驶模拟器研究只能在一定程度上重现真实世界中的驾驶行为（Riener，2011），且往往缺乏绝对效度（Kaptein等人，1996；Fisher，2011）。此外，要在用户研究中评估新的车载信息系统设计，就必须在功能原型中实现所有相关特征。尽管这类措施仍不可或缺，汽车用户体验专家仍需要可解释的评估方法（Ebel 等，2020），以便在设计早期阶段就识别出可能造成分心的交互模式。这类自动化方法

*通讯作者ebel@cs.uni-koeln.de (P. Ebel); christoph.lingenfelder@mercedes-benz.com(C. Lingenfelder);vogelsang@cs.uni-koeln.de (A.Vogelsang)开放研究者与贡献者身份识别码(： 0000-0001-7511-2910 (P. Ebel); 0000-0001-9417-5116 (C. Lingenfelder);0000-0003-1041-0815 (A. Vogelsang)

[1] https://www.tesla.com/model3

[2] https://www.mercedes-benz.com/en/innovation/future-mobility/eqs-with-unique-mbux-hyperscreen/

能够促进安全性内置的交互概念开发，从而降低其未能通过最终评估的可能性。

当前基于用户交互信息预测驾驶分心的方法，源自菲茨定律（Fitts，1954年）或击键层级模型（KLM）（Card 等，1980年）等方法：这些方法将特定操作所需的时间相加，以预测任务的总耗时，再根据这一指标推导注视行为。尽管这些模型受限于累积式线性结构，且需要专家手动指定每项任务，但其可解释性很强。更复杂的模型通常会以牺牲可解释性为代价来提高准确率，这凸显了二者之间固有的权衡（Lundberg 和 Lee，2017年）。然而，缺少解释可能会使科学发现不为人知，降低用户接受度，并限制学习效果（Molnar，2020年；Doshi-Velez 和 Kim，2017年）。为了帮助制定以数据为依据的设计决策，不仅要预测可能存在危险的交互序列，还要了解哪些交互会迫使驾驶员将视线从道路上移开。

2. 背景与相关工作

本节将介绍视觉需求的概念，说明如何对其进行测量，并介绍视觉需求的计算模型。最后，我们将介绍 SHAP，这是一种用于解释机器学习预测结果的方法。

2.1. 次要任务参与的视觉需求

视觉需求是指“为执行特定任务而从某个物体中提取信息所需的视觉活动程度或数量”（ISO 15007）。驾驶汽车时，参与次要任务会使驾驶员的注意力从主要驾驶任务上转移，造成视觉分心，从而损害驾驶表现和安全（Engström等人，2005年；Donmez等人，2010年；Liang和Lee，2010年；Green，1999年；Klauer等人，2006年；Horrey和Wickens，2007年）。这同样适用于更高自动化等级的驾驶自动化（SAE J3016）。近期研究表明，经历一段自动驾驶后，次要任务的视觉认知负荷会显著影响接管表现（Wintersberger等人，2021年），而分心也会产生显著影响（Merlhiot和Bueno，2021年）。ISO 15007:2020（ISO 15007）描述了多种视觉需求测量指标，其中应用广泛的两项指标是总注视时长（TGD）和平均注视时长。总注视时长是指“在一项条件任务、子任务或子子任务期间，对某个兴趣区域（或一组相关兴趣区域）的所有注视时长之和”。平均注视时长是指“在一项条件任务、子任务或子子任务期间，对某个兴趣区域（或一组相关兴趣区域）的所有注视时长的均值”。进一步的研究（Horrey和Wickens，2007年）表明，单次注视时间长于正常水平，尤其是超过两秒的注视（Klauer等人，2006年），与驾驶安全性降低高度相关。“NHTSA车载电子设备驾驶员视觉-手动分心指南”（美国国家公路交通安全管理局，2012年）也提到了这一点。然而，根据Victor等人（2014年）的研究，没有任何单一指标能够全面描述注视行为与风险之间的关系，因此需要综合使用多种指标。Burns等人（2010年）还指出，任何单一测量都只能对分心情况作出不完整的评估；因此，在设计流程的早期阶段，就应尽可能纳入上述这类有实证支持的测量指标，以辅助决策。

2.2. 视觉需求预测

多种方法旨在预测驾驶过程中的视觉-手动分心（Kanaan 等人，2019；Li 等人，2018；Wollmer 等人，2011；Li 等人，2020；Risteska 等人，2021）。其中大多数方法侧重于驾驶员分心检测，以便在检测到潜在危险情况时向驾驶员发出警告。这些方法通常基于自然驾驶数据，并采用各种机器学习方法。它们利用驾驶表现指标（例如速度或方向盘转角）（Kanaan 等人，2019；Li 等人，2018；Wollmer 等人，2011）、环境数据（例如交通状况）（Risteska 等人，2021）或驾驶员视频数据（Li 等人，2020；Kutila 等人，2007）。尽管这些方法展现出良好的前景，但并未纳入驾驶员如何与手机或车载信息娱乐系统等辅助设备进行交互的信息。因此，它们无法深入揭示特定用户界面元素或交互的视觉需求。

然而，存在多种基于用户与特定用户界面元素的交互来建模车载人机界面视觉需求的方法。这些方法解释了特定交互对驾驶员视觉分心的影响。这些方法着眼于理解交互行为，并旨在识别车载人机界面中容易导致分心的特征。其目的是在开发过程的早期阶段，向设计师和研究人员说明设计可能对驾驶分心产生的影响。本文聚焦于后者，并概述该领域当前的最新进展。

大多数基于用户交互预测视觉需求的方法，都是源自 KLM 建模技术的自下而上方法（Card 等人，1980年；Card，1983）。在这类方法中，整个任务被分解为一系列特定的基本操作符（例如，按下按钮或在列表中搜索）。随后通过实证确定每个操作符对应的交互时长（Schneegaß 等人，2011年）。由此预测的任务总耗时，等于任务中各个操作符对应交互时长之和。KLM 技术最初用于预测计算机辅助办公任务的处理时间，但后来经过多次调整，以评估车载信息娱乐系统（Schneegaß 等人，2011年；Manes，1997；Lee 等人，2019）。然而，这些方法大多关注任务完成时间，而非视觉需求。Pettitt 等人 （2007年）率先提出一种基于 KLM 的视觉需求预测方法。他们发现预测值与遮挡实验测量结果高度相关。首个直接预测视觉需求的 KLM 方法由 Purucker等人 （2017年）提出，他们构建了一种特定任务的KLM模型。他们认为，使用固定操作符对创新型新硬件建模，效果有限。他们的方法只能预测 TGD，而 Large 等人（2017b）提出的方法还能预测凝视次数和平均凝视时长。他们的信息论方法以用于决策/搜索时间的希克-海曼定律和用于指向时间的菲茨定律为基础。

尽管本文介绍的基于 KLM 的方法取得了令人鼓舞的成果，但它们都存在一些缺点。首先，由于这些模型具有累积性和线性特征，因此不适合模拟不同用户交互或驾驶情境之间可能存在的（非线性）依赖关系。例如，在较低速度下，从列表中选择一个元素与轻触按钮之间的视觉需求差异可能微不足道，但在较高速度下则可能显著。此外，交互序列的长度与特定交互相结合，也可能以非线性、非加和的方式影响视觉需求（Purucker等人，2017年）。例如，如果驾驶员按下两个彼此距离很近的按钮，TGD 不太可能因此翻倍，因为驾驶员可能会在一次注视期间完成这两次交互。其次，所介绍方法的模型参数是通过在受限驾驶场景中使用不同保真度的驾驶模拟器进行实证测试，并招募数量相对较少的参与者得出的。这很可能导致预测高度依赖具体情境，正如Large 等人(2017b)所指出的，并且在一项评估菲茨定律适用性的真实道路驾驶实验中得到了证明（Pampel 等人，2019）。第三，当前方法没有考虑驾驶情境对视觉需求的影响。研究表明，驾驶员会根据驾驶需求（Risteska 等人，2021年）和辅助驾驶程度（Morando 等人，2021年；Tsimhoni 和 Green，2001；Gaspar 和 Carney，2019；Large 等人，2017a）调整任务投入和视觉注意力，因此纳入此类参数至关重要。

另一种方法由Kujala和Salvucci（2015年）提出，他们基于ACT-R认知模型架构构建了一个模型（Anderson等人，2004）。他们的方法旨在呈现驾驶员的视觉采样策略。他们认为，驾驶员会根据当前驾驶表现所决定的时间限制调整注视。该模型能够预测视觉需求的多个方面，但仅考虑了网格和列表布局。此外，驾驶场景相当简单，评估结果显示其预测准确度存在明显缺陷，尤其是在检测长时间注视方面。

我们认为，利用大量真实驾驶和交互数据，并结合能够建模非线性关系的机器学习方法，是迈向更准确且适用范围更广的解决方案的有益一步。正如上文所述，机器学习方法已在分心检测方法领域实现了高质量预测。

然而，有两个主要因素阻碍这些方法深入揭示特定设计元素如何影响视觉需求。首先，交互数据的数量尚无法与驾驶数据相提并论。其次，上述所有机器学习方法都缺乏可解释性。虽然报告了某些性能指标，但这些模型仍是黑箱，无法揭示哪些特征对预测起决定作用。

2.3. 使用 SHAP 进行可解释预测

可解释人工智能（XAI）旨在通过提供人类可理解（可解释）的信息，解释机器学习模型的行为与过程，从而提升机器学习模型的透明度（Barredo Arrieta 等人，2020年；Liao 等人，2020年）。解释充当人类与模型之间的接口（Barredo Arrieta 等人，2020年），在各种应用中都具有价值（Wiegand 等人，2020年；Wang 和 An，2021年）。解释能够增进科学理解（Doshi-Velez 和 Kim，2017年），增强用户信任（Shin，2021年；Lipton，2018），还能够帮助推断数据中的因果关系（Verma 等人，2020年）。对于当前任务，可解释预测尤其值得关注，因为目标不仅是预测视觉需求，还要得出有关

特定用户界面元素、手势以及不同驾驶情境所产生影响的结论。这种方法旨在实现人工智能辅助决策（Zhang 等人，2020年），根据人类专家的领域知识以及模型预测和相应解释所产生的见解，优化联合决策。

SHAP 由 Lundberg 和 Lee （2017年）提出，是一种基于合作博弈论中 Shapley 值的方法（Shapley，1953）。SHAP 方法为任意预测模型提供局部和全局解释。SHAP 属于加性特征归因方法。其核心思想是使用可解释解释模型$g(z')$，采用线性函数的形式，使模型对某个实例的预测等于其特征贡献之和 $\phi_i \in \mathbb{R}$（Molnar，2020年）：

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i, \tag{1}$$

其中 $z' \in \{0,1\}^M$ 和 z'_i 分别表示特征 i是否存在，ϕ_0 表示不存在任何特征时模型的输出，而 M 表示输入特征的数量（Lundberg 等人，2020年）。

Lundberg 和 Lee （2017年）进一步指出，存在一个唯一解，该解遵循加性特征归因方法的定义（见1），并满足局部准确性、缺失性和一致性。局部准确性是指特征归因之和等于原始模型的预测值。缺失性是指缺失的特征（$z_i = 0$）被赋予零归因；一致性则表示，当改变模型使其更加依赖某个特征时，该特征的归因不应减少。

唯一的 由Lundberg 和 Lee（2017年）所描述的可能解决方案由 SHAP值给出 ：

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|M| - |S| - 1)!}{M!} \left(f_x(S \cup \{i\}) - f_x(S) \right), \tag{2}$$

其中，S 是 z'中非零索引的集合，$f_x(S)$ 是给定输入特征子集 S 条件下函数的期望值，N 是所有输入特征的集合。

存在多种不同的方法，可针对不同类型的机器学习模型近似计算 SHAP值。然而，在本研究中，我们使用 TreeSHAP（Lundberg 等，2020年），该方法能够为基于树的方法计算精确的 SHAP值。

与 LIME（Ribeiro 等，2016年）等方法，或针对基于树的模型的特定方法（如置换重要性和特征不纯度计算）相比，SHAP 具有诸多优势。由于其坚实的博弈论基础（Molnar，2020年），SHAP值在一致性和局部准确性方面具有理论保证。此外，局部和全局解释保持一致，SHAP值能够表明每个特征的贡献是正面还是负面，并且 Lundberg 等 （2018）表明，SHAP值与人类直觉具有更高的重叠度（Lundberg 和 Lee，2017年；Lundberg 等，2020年）。

3. 提出的方法

在这项研究中，我们展示了如何有效利用机器学习方法，基于大规模自然驾驶数据，预测并解释车载触摸屏交互的视觉需求。本文的贡献有两方面：首先，我们提出一种机器学习方法，根据交互类型及相关驾驶参数，预测车载触摸屏交互的视觉需求。其次，我们应用 SHAP 方法（Lundberg 和 Lee，2017年），解释预测结果，并直观呈现用户交互、车速、方向盘转角和自动化等级如何影响驾驶员长时间注视概率和 TGD。下文将介绍本文后续部分会用到的若干定义，并详细描述数据收集、数据处理和建模流程。

3.1. 定义与问题陈述

该方法旨在根据用户交互及相关驾驶参数，预测驾驶员视觉注意力分配。为此，我们结合交互序列、驾驶序列和注视序列，对驾驶员的次要任务参与进行建模。下文将介绍这些概念。[3]

[3]正式定义请参阅A.1

交互序列。 交互序列定义为驾驶员连续执行的一组触摸屏交互。两次连续交互之间的持续时间必须小于 Δt_{max}。

注视序列。 注视序列定义为驾驶员朝向预先指定的兴趣区域（AOI）连续注视所构成的集合。

驾驶序列。 驾驶序列是由一系列驾驶数据观测构成的序列。每个观测包含车速、方向盘转角，以及自适应巡航控制（ACC）和转向辅助（SA）的状态。

次要任务参与。 次要任务参与 S 描述驾驶员与中控台触摸屏交互时的触摸屏交互、驾驶行为和注视行为。我们考察首次交互前 t_b 秒至最后一次交互后 t_b 秒期间的车速和方向盘转角。此外，还会考虑首次与最后一次交互之间的所有注视。

长时间注视预测任务。 长时间注视预测任务是指预测驾驶员在参与次要任务时，是否会注视中控台触摸屏超过两秒。

TGD预测任务。 TGD预测任务是指预测驾驶员在参与次要任务期间对中控台触摸屏的总注视时长。

3.2. 数据收集

本研究所用数据通过梅赛德斯-奔驰的远程信息处理数据记录框架，从 100 多辆测试车辆和五种不同车型通过无线方式收集。有关记录机制的更详细说明见 Ebel 等人（2021a）。数据收集期从 2021年10月中旬至 2022年1月中旬。所有配备最新软件架构、眼动追踪器、自适应巡航控制和半自动驾驶功能的公司内部测试车辆均参与了数据收集。这些测试车辆用于各种测试驾驶（其中大多与用户界面无关），也用于员工的休闲驾驶。车辆在欧洲各地行驶，但记录的大多数行程发生在德国。我们利用梅赛德斯-奔驰的数据收集与处理框架，收集触摸屏交互、驾驶数据（车速、方向盘转角和驾驶自动化等级）以及眼动追踪数据。我们未收集任何人口统计或环境数据。

触摸屏交互数据采集自用户界面软件：驾驶员每次触摸中控台触摸屏时，系统都会记录一个数据点。每个数据点包括被触摸的用户界面元素、用于该手势的所有手指的起始位置和结束位置，以及一个时间戳。

注视数据通过安装在方向盘后方仪表盘处的立体摄像头采集。该系统已实现商业化，并应用于量产车辆。与大多数远程眼动追踪器一样，注视检测基于瞳孔-角膜反射技术（Merchant，1967）。在这种方法中，系统会追踪瞳孔中心相对于角膜反射位置的变化（Hutchinson 等人，1989年11月至12月）。驾驶员的视野被划分为不同的感兴趣区域，每当驾驶员的注意焦点在不同感兴趣区域之间切换时，系统便会采集一次新的注视。在所有条件下，各感兴趣区域的平均真阳性率均超过90%。该系统不会采集任何原始视频数据。

所有与驾驶相关的数据均直接从控制器局域网（CAN）总线采集。连续信号（车速和方向盘转角）以 $4\,Hz$ 的频率采集，基于事件的信号（自适应巡航控制、半自动驾驶和安全带信息）则在发生变化时采集。自适应巡航控制会根据限速和前方车辆情况自动调整车速。半自动驾驶会主动辅助横向控制，使车辆保持在车道中央。两种系统的工作速度范围均为 0 千米/小时至 210 千米/小时。

3.3. 数据处理

下文将介绍我们如何处理数据，使相应的序列符合3.1中给出的定义。图1 展示了各个序列数据合成过程的示意图。为清晰起见，我们仅显示车速（黑色实线），以代表其他驾驶数据。

3.3.1. 交互数据

在受控实验中，参与者会按照实验人员指定的预定义任务进行操作。在这类情境中，用户交互与任务之间的映射十分直接。然而，在此处的观察性场景中，并未定义任务边界。我们不知道驾驶员在参与次要任务期间打算完成哪项任务，也不知道任务的开始或结束时间。一种提取交互序列的方法，是考虑一次行程中发生的所有交互。但这样会产生非常长的交互序列，其中交互密集的时段零散地分布在很长一段时间内。为了解决这个问题，我们

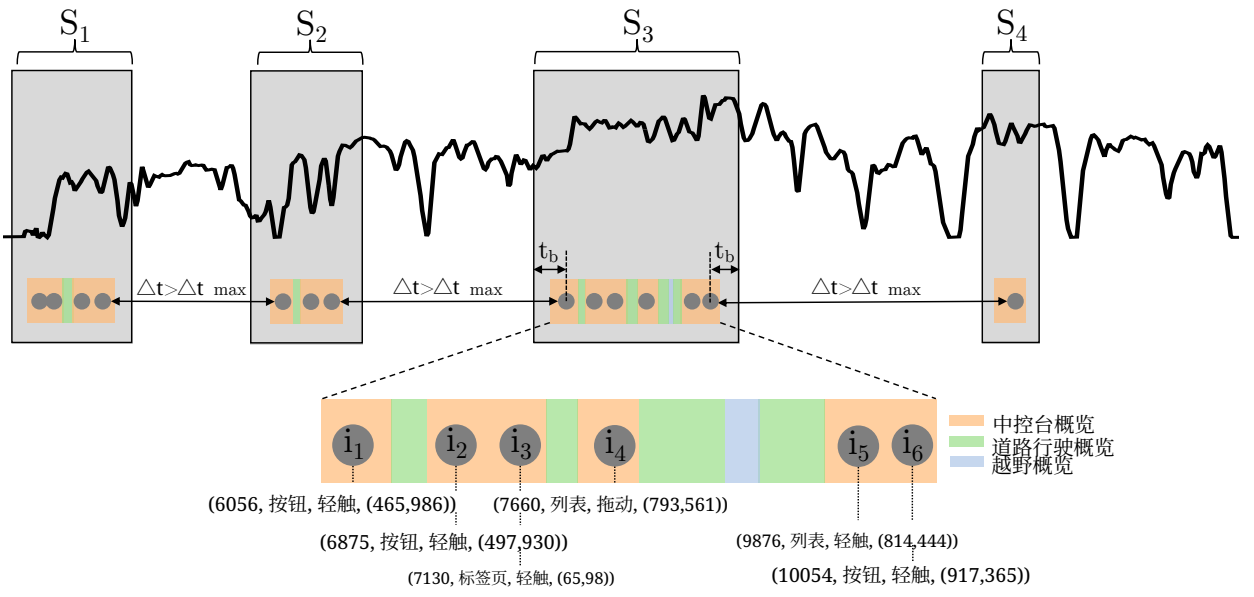


图1： 示意图概述了如何从驾驶序列（黑色实线）、注视序列（彩色矩形）和交互序列（灰色圆点）中提取次要任务参与情况 S_n。

将最大更新间隔设定为（定义见3.1 节），即 $10\,s$，$\Delta_{t_{max}} = 10\,s$。我们假设，如果驾驶员连续10秒没有进行交互，便会停止参与次要任务，并且交互序列在最后一次交互后结束。随后，我们将下一次交互视为新交互序列的起点。我们认为10秒这一假设是合理的，因为交互序列时长的分布以及驾驶员注视中控台触摸屏的总时长分布，都与文献中报告的数值相符（美国国家公路交通安全管理局，2012年；Angell等人，2008）。

对于所有其余交互，我们根据手指位置信息计算手势类型（轻触、拖动和多点触控）以及两次交互之间的距离。最后，我们将每个用户界面元素映射到一个更高层级的元素类型（见表 1）。此步骤可减少数据稀疏性，并确保该方法能够针对特定元素类型得出可泛化的结论。

3.3.2. 注视数据

我们采用多个筛选步骤来提高注视数据的数据质量。筛选方法部分参考了相关研究（Morando 等人，2019年）。第一步，将注视信息汇总为更宽泛的兴趣区域（道路区域、非道路区域、中控台）。根据 ISO 15007-1:2020（ISO15007），我们将所有未直接注视道路的注视（例如，注视后视镜）都视为视线离路凝视。由于我们特别关注朝向中控台触摸屏的注视，因此将这类注视与常规的视线离路凝视区分开来。第二步，如第 3.1 节所述，我们考虑相关交互序列中首次与最后一次触摸屏交互之间的所有注视。对于交互序列开始或结束处不完整的注视，若其开始时间早于首次交互或结束时间晚于最后一次交互，则将其视为一个整体。第三步，如果跟踪丢失时间较短（少于 $300\,ms$），且丢失前后的兴趣区域相同，则对该时段进行插值。光照条件变化、眼镜反光，或驾驶员将手放在方向盘上遮挡摄像机视野，都可能导致跟踪丢失。第四步，根据 ISO 15007-1:2020（ISO15007），对短于 $120\,ms$ 的注视进行插值，因为对某个兴趣区域的注视点持续时间在生理上不可能短于此值。第五步，基于同样的理由，对不同兴趣区域之间短于 $120\,ms$ 的跟踪丢失也进行插值。第六步，对短于 $500\,ms$ 的眼睑闭合进行插值，以按 ISO 15007-1:2020（ISO15007）的建议去除眨眼的影响。

3.3.3. 驾驶数据

驾驶数据由连续数据和基于事件的数据组成。首先，我们提取与相应交互序列相关的数据。为在非常短的交互序列（可能仅包含一次交互）中更稳定地估算驾驶参数，我们会考虑方向盘和车辆

表1描述次要任务参与的最终输入特征概览。

特征	描述
交互数据	
$n_{\text{B钮}}$	# 与常规按钮的交互（例如，按下按钮或单选按钮）
n_{List}	# 与列表的交互（例如，选择建议目的地时）
n_{Map}	# 与地图查看器的交互（例如，缩放或拖动导航地图时）
n_{Slider}	# 与滑块元素的交互（例如，调节音量时）
$n_{\text{主页栏}}$	# 与屏幕底部静态主屏栏的交互
$n_{\text{CoverFlow}}$	# 与 Cover Flow 小组件的交互（例如滚动浏览专辑封面时）
$n_{\text{应用图标}}$	# 与主屏幕上的应用图标进行交互
n_{Tab}	# 与标签栏进行交互
n_{Keyboard}	# 在键盘或数字键盘上的交互（例如输入目的地时）
$n_{\text{浏览器}}$	# 在网页浏览器中的交互
n_{RemoteUI}	# 在 Apple CarPlay 或 Android Auto 中的交互
$n_{\text{ControlBar}}$	# 与控制栏的交互，该控制栏以小型叠加层的形式显示在各种屏幕上
n_{PopUp}	# 与弹出元素的交互
$n_{\text{ClickGuard}}$	# 与不可交互背景元素的交互
n_{Other}	# 与不属于上述任何类别的用户界面元素的交互
$n_{\text{未知}}$	# 与无法提取标识符的用户界面元素的交互
n_{Tap}	#点击手势
n_{Drag}	# 拖动手势
$n_{\text{Multitouch}}$	# 多点触控手势
d_{avg}	连续两次交互之间的平均距离，单位为像素
N	交互次数
驾驶数据	
v_{avg}	平均车速，单位为千米/小时
θ_{avg}	平均方向盘转角，单位为°
a_{acc}	自适应巡航控制状态 $a_{\text{acc}} \in \{0,1\}$
a_{sa}	转向辅助状态 $a_{\text{sa}} \in \{0,1\}$

速度数据，时间范围从交互序列的首次交互前两秒到末次交互后两秒（$t_b = 2\,s$）。数据提取后，将剔除存在缺失值或错误值的序列，以及记录频率出现偏差的序列。

3.3.4. 数据聚合与最终筛选

我们总共提取了 322,425 个触摸屏序列。其中，145,973 个序列具有有效的速度数据，81,150 个序列具有有效的转向数据，111,792 个序列具有有效的注视数据。分别处理后，我们对各数据源取交集，得到 30,158 次完整的次要任务参与。大多数序列被排除，原因包括：序列是在试验台上生成的（没有可用的驾驶数据）、车辆未配备摄像头，或因数据连接中断而不符合采样要求。在数据处理的第二阶段，我们进一步执行筛选步骤，以提高数据质量。为避免数据过于稀疏，我们剔除了交互次数超过 41 次的 342 次次要任务参与（$N > 41$ 对应于分布的 99th 百分位数 N）。这些次要任务参与可视为离群值；对于当前应用场景而言，保留它们并无额外益处。此外，我们剔除了乘客在场的 16,864 次次要任务参与，因为乘客也常与中控台触摸屏交互。这些交互无法对应到驾驶员注视，会使数据偏向于每次次要任务参与中注视次数更少、时长更短的情况，同时记录了许多并非由驾驶员发起的交互。此外，我们还剔除了车辆完全停车期间的 809 次参与，以及因残留速度错误而剔除的一个序列。经过这一步处理后，我们得到最终的 12,142 次次要任务参与。最后，我们根据表1中的说明，计算次要任务参与的汇总统计量，以生成最终的特征集。这些特征将作为下文所述模型的输入。

3.4. 建模

根据问题陈述，我们要解决一个分类任务（长时间注视预测）和一个回归任务（TGD预测）。对于每项任务，我们将基线方法和逻辑回归/线性回归方法与三种机器学习方法进行比较，分别是随机森林、梯度提升树和前馈神经网络（前馈神经网络）。在长时间注视预测任务中，基线方法会随机预测两个类别之一（即在类别均衡的数据集中，正确预测长时间注视的概率约为 50%）。在 TGD 预测任务中，基线方法预测训练数据集的 TGD 中位数。基于机器学习的方法，其参数通过采用随机搜索进行广泛的超参数优化来确定[4]。

4. 评估

本节将介绍最终数据集，结合背景分析实验结果，并详细阐述应用 SHAP 方法生成的可解释预测。

4.1. 数据集

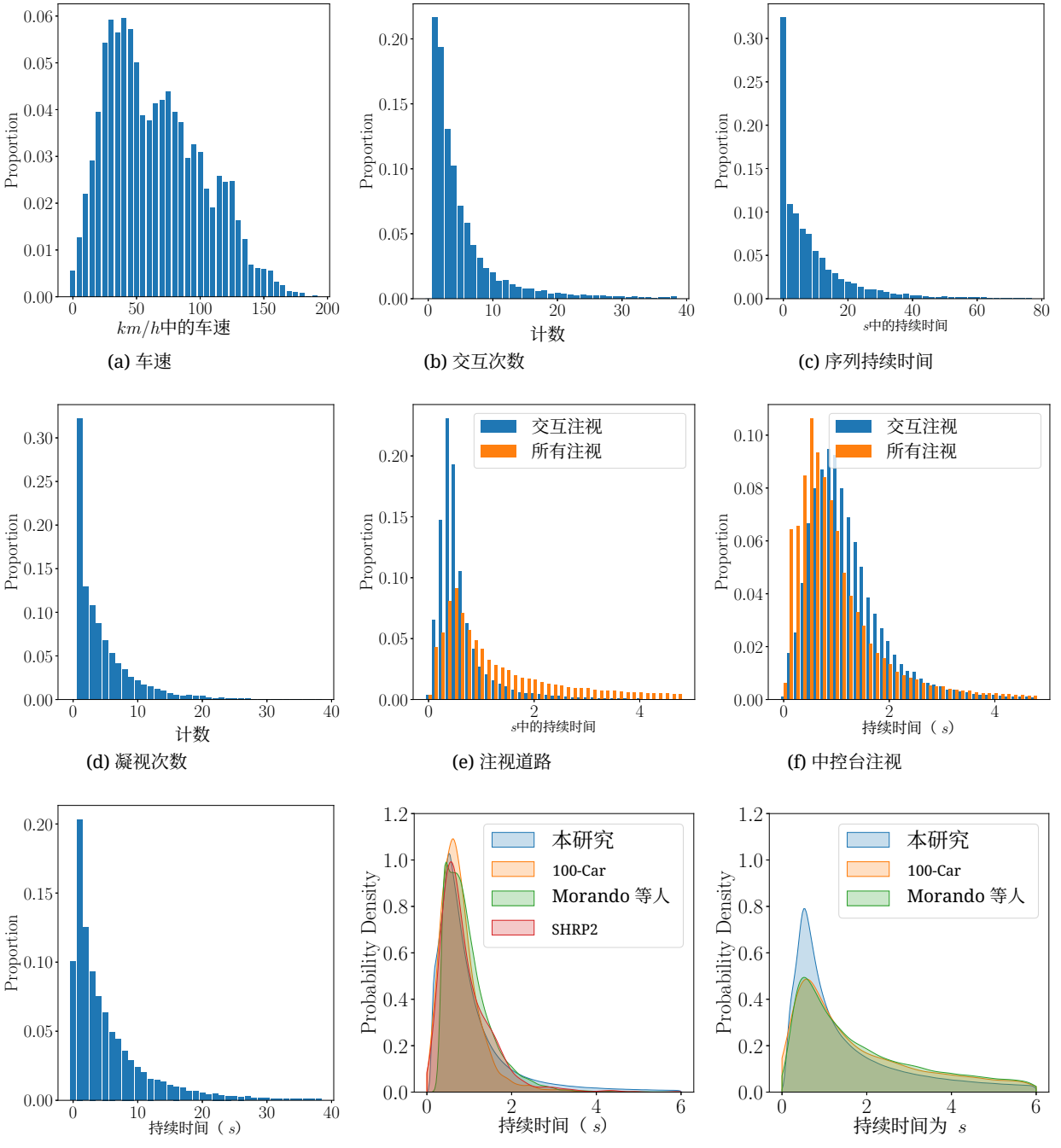
最终数据集包含从 3,046 次独立行程中抽取的 12,142 次次要任务参与记录。大多数次要任务参与记录来自梅赛德斯-奔驰S级（7,342 次）、EQS（3,604 次）和 EQE（824 次）车型。这些车辆配备了 17.7 英寸、12.8 英寸或 11.9 英寸的中控台触摸屏，且像素密度相近。总计采集了 61,943 次触摸交互和 119,770 次独立注视。行程时长的中位数为 34.28 分钟（$Q_1 = 17.49$，$Q_3 = 66.58$）。最终数据集的具体注视和交互统计数据见图2。[5]

在图2e 和图2f中，绘出了次要任务参与期间的注视时长分布（蓝色），并将其与所有会话中的注视时长分布（橙色）进行对比，后者不区分驾驶员是否参与次要任务。这样便可与那些无论驾驶员是否参与次要任务都会使用所收集数据的方法进行比较。

我们还将数据与*100*车研究（Dingus等人，2006年）（数据由Custer （2018）提供）、*SHRP2* （Victor 等人，2014年）（数据见Bärgman等人（2015年））以及Morando 等人（2019）研究中报告的数据进行比较（作者应请求提供）。图2i和图2h展示了相应数据集中道路区域注视和非道路区域注视的分布。注视时长分布截断于6秒，因为这与100-Car基线数据集中片段的长度相对应。视觉比较表明，非道路区域注视时长分布与三项相关研究报告的数据吻合良好。不过，道路区域注视在我们的数据与100-Car研究及Morando 等人（2019）研究报告的数据之间存在一些差异。尽管三个数据集的众数相近，我们研究中的道路区域注视时长往往短于另外两项研究。造成这一现象的潜在原因有很多。例如，Morando 等人（2019）只考虑驾驶情境受到严格控制的驾驶片段，排除弯道驾驶、变道以及车速低于60千米/小时的驾驶片段。在这些相对平稳的驾驶情境中，驾驶员无需频繁在道路区域与后视镜或侧窗等非道路区域之间切换视线，因此道路区域连续注视时间更长。与100-Car研究的差异可能是因为该研究的数据距今已近20年，且仅涵盖手动驾驶。当时车辆的技术，尤其是信息娱乐系统和辅助系统的技术，与如今车辆上的技术有根本区别。不过，考虑到数据收集方式的差异，这一比较表明，我们的数据采集与处理流程能够生成具有代表性的数据。

图 2e 显示，在触摸屏] 交互过程中，驾驶员需要在道路和中控台之间分配视觉注意力，因此注视道路的时间较短。另一方面，次要任务参与期间对中控台的注视往往比一般情况下对中控台的注视时间更长（图 2f）。从图2b中可以看出，约 25% 的所有序列仅包含一次交互。因此，许多次要任务参与时间较短，仅包含一次注视中控台（图2d）。这些短暂的任务参与属于真实世界用户行为。然而，在仅评估少数预定义任务的实验室研究中，通常不会体现这类行为。我们认为，分析这些短暂的任务参与仍然很有意义，因此决定将其纳入考量。对于长时注视分类任务，我们通过随机欠采样来平衡数据集。最终的数据集每个类别包含 4,816 个序列。

[4]有关搜索空间的更多详细信息，请参阅：A.2 [5]有关完整数据集统计的信息，请参阅 A.3



(a) 车速 (b) 交互次数 (c) 序列持续时间

(d) 凝视次数 (e) 注视道路 (f) 中控台注视

(g) TGD 中控台 (h) 视线离路凝视 (i) 注视道路

图 2： 用于展示注视和交互数据的直方图。其中显示了：(a) 每次次要任务参与期间的平均速度；(b) 每次次要任务参与期间的交互次数；(c) 每次次要任务参与期间交互序列的持续时间；以及 (d) 每次次要任务参与期间注视触摸屏的次数。图 (e) 比较了次要任务参与期间的道路注视时长分布与不论是否进行触摸屏交互时的注视时长分布。图 (f) 对注视中控台触摸屏的情况进行了同样的比较。图 (g) 显示了次要任务参与期间对中控台触摸屏的总注视时长。图 (h) 和 (i) 将本研究中的视线离路凝视和道路注视时长概率密度函数，与 100车研究、SHRP2研究（仅视线离路凝视）以及Morando 等人 (2019）。

表2不同模型的比较。

	长时间注视预测		总注视时长预测	
模型	准确率	标准差	平均绝对误差	标准差
基线	50.09 %	1.63 %	4378 毫秒	177 毫秒
逻辑回归/线性回归	61.93 %	1.69 %	3778 毫秒	383 毫秒
随机森林	**67.53 %**	**1.38 %**	**2437 毫秒**	**112 毫秒**
XGBoost	67.22 %	1.85 %	2385 毫秒	117 毫秒
FNN	65.90 %	2.02 %	2443 毫秒	109 毫秒

4.2. 实验结果

我们使用重复十折交叉验证（Kohavi，1995）评估回归模型，并使用分层十折交叉验证评估分类模型。结果见表2。模型使用表1中列出的全部输入特征进行拟合。

基于机器学习的方法在两项任务中的表现均优于基线方法、逻辑回归和线性回归方法。预测准确度的差异支持了我们的假设：这里讨论的两个问题都不能视为线性问题，并且不同特征之间存在交互效应。机器学习模型的结果相近。不过，随机森林方法具有两项理想特性，使其尤其适用于当前用例。首先，TreeSHAP（Lundberg 等人，2020年）算法能够高效计算随机森林模型的精确 SHAP值。其次，随机森林可以并行运行，因此适用于未来部署在整个生产车队数据上的用例。因此，我们选择随机森林模型进行后续解释生成。

4.3. 可解释预测

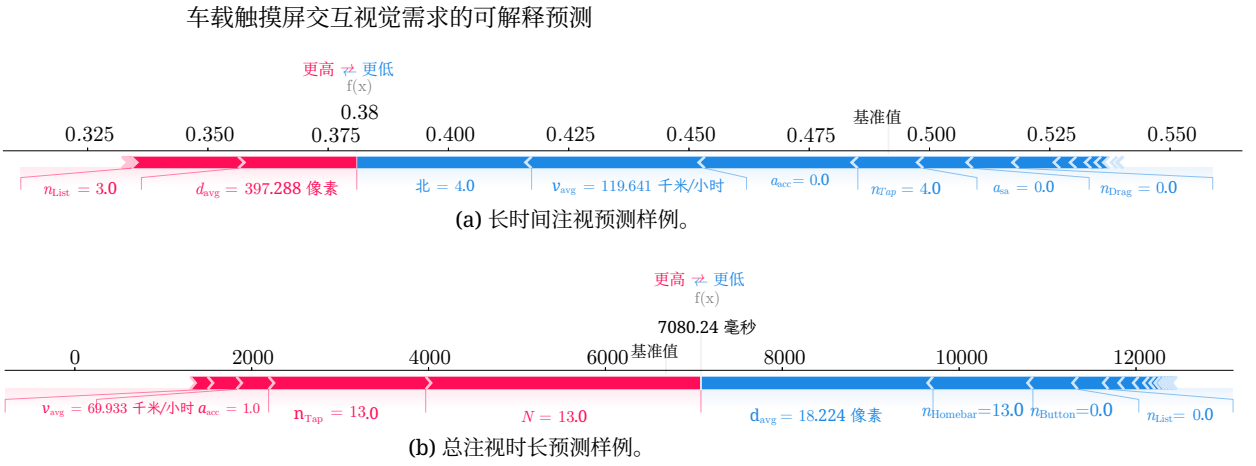
尽管上述结果能够很好地衡量预测准确度，但在理解人类行为方面价值有限。要真正支持研究人员和从业者在设计流程中加深对驾驶员视觉注意力分配的理解，仅仅预测新的用户流程是否可能造成过度分心是不够的（Ebel 等人，2021年b）。因此，我们采用 SHAP。SHAP值代表各特征对模型输出的贡献，并为每个输入样本提供局部解释。通过综合多个局部解释，可以呈现全局结构，从而深入揭示模型行为（Lundberg 等人，2020年）。

4.3.1. 局部解释

图3展示了一项长时间注视预测和一项 TGD预测的解释。这些力图将特定模型输出表示为特征贡献（即 SHAP 值）的累积效应。每个条形的长度表示相应特征值将模型输出从基准值推向更高值（红色，向右）或更低值（蓝色，向左）的程度。基准值是根据训练数据集上的平均模型输出计算得出的。每组中的特征按其影响大小排序，且仅显示影响最大的特征。特征值显示在条形下方。对于长时间注视预测，特征贡献以概率表示；对于 TGD预测，则以毫秒表示。

图3a 展示了某次次要任务参与的解释，模型输出的长时间注视概率为 0.38 = 38 %。由于驾驶员仅进行了4次交互（$N = 4$），且行驶速度为 $v_{\text{avg}} = 119.641\,km/h$，同时自适应巡航控制处于停用状态（$a_{\text{acc}} = 0$），因此长时间注视概率向左移动。另一方面，由于这些交互分布在屏幕上的位置较为分散（$d_{\text{avg}} = 397.288\,px$），且其中三次为列表交互（$n_{\text{List}} = 3$），因此预测向右移动。

图 3b 说明了另一次次要任务参与。在此，约为 $7\,s$ 的 TGD预测接近基准值，因为正向和负向特征贡献相互抵消。在这次次要任务参与期间，驾驶员在以 70 千米/小时的速度驾驶且自适应巡航控制（ACC）已启用的情况下，进行了 13 次触摸交互（$N = n_{\text{Tap}} = 13$）。如果模型只能获取这些信息，它会预测 TGD 约为 13 秒。然而，由于所有交互彼此间隔很近（$d_{\text{avg}} = 18.224\,px$），且都在主页栏上进行（$n_{\text{Homebar}} = 13$，没有列表或按钮交互造成干扰（$n_{\text{List}} = 0$，$n_{\text{Button}} = 0$，最终模型



(a) 长时间注视预测样例。

(b) 总注视时长预测样例。

图 3： 以力图形式呈现的局部解释。

输出仅略高于平均 TGD预测。

这些局部解释表明，并非所有特征始终都相关。次要任务参与的预测可能仅由少数几个主导特征驱动。所呈现的解释使设计师和研究人员能够快速识别单个预测背后的主要驱动因素。它还让他们能够尝试使用人工输入样本，并观察用户流程设计或驾驶情境的某些变化如何影响模型的输出。

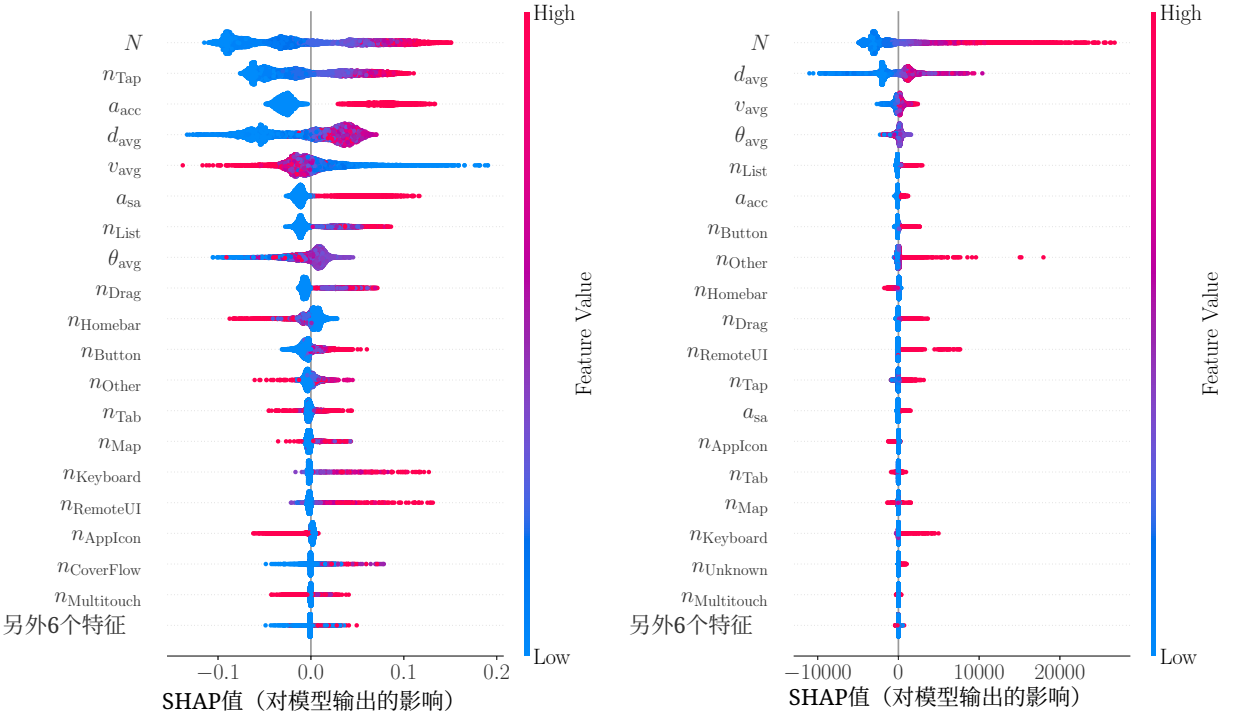
4.3.2. 全局解释

为了解特征如何从全局层面影响模型输出，我们汇总了数据集中的所有局部解释。图4 以蜂群图的形式展示了 SHAP 值的分布（即每个特征对特定预测的影响，见 4.3.1）。每一行中的每个点代表一次次要任务参与。横轴上的位置表示相应特征对模型输出的影响。图 4a中的 SHAP 值处于概率空间，而图4b 中的值表示以毫秒为单位的影响。颜色表示特征值（红色为高，蓝色为低）。特征按全局重要性排序，仅单独显示最重要的 19 个特征。

长时间视线预测模型中最重要的特征（图4a）是交互次数 N、交互之间的平均距离 d_{avg}以及点击手势次数 n_{Tap}。驾驶员进行的触摸屏交互越多、交互之间的距离越大，与之相关的注视中有一次超过 2 秒的预测概率就越高。图 4a还显示，激活自适应巡航控制 a_{acc} 和半自动驾驶 a_{sa}都会提高长时间注视概率。辅助系统未激活（蓝色）时，对所有样本的影响都很小；而辅助系统激活时，影响则有所不同。水平分布表明，辅助驾驶对视觉注意力分配的影响因情境而异，并取决于驾驶情境和交互模式等其他因素。描述车速影响的分布 v_{avg} 呈现明显的长尾特征。对于大多数中等速度下的次要任务参与，影响为负，但相对较小。高车速会降低预测的长时间注视概率，而低车速则会提高该概率。这表明驾驶员具有自我调节行为。

列表交互次数 n_{List} 是与特定用户界面元素相关的最重要特征，其次是与主页栏的交互次数 n_{Homebar}。从图 4a中可以看出，这两者的影响方向相反。随着列表交互次数增加，长时间注视概率会上升；而随着主页栏交互次数增加，长时间注视概率则会下降。这表明，列表交互往往比静态主屏栏上的交互更容易分散注意力。与 Android Auto 或 Apple CarPlay 的交互 n_{RemoteUI} 所产生的影响与列表交互类似。总体而言，与特定用户界面元素类别相关的大多数 SHAP 值分布都以零为中心，并向一侧或两侧延伸出长尾。这是因为大多数元素只出现在少数次要任务参与中。尽管这会导致全局重要性相对较低，但这些特征仍会对特定预测产生重大影响。

对于 TGD预测模型（图 4b）， N 和 d_{avg} 也是最重要的两个特征。它们的分布与长时间注视预测任务中观察到的分布也存在相似之处。不过，车速v_{avg} 的影响与长时间注视预测任务中的情况相反。较高的速度值会使 TGD预测升高，较低的速度值则会使预测值降低。综合这两项发现，驾驶员可能会缩短



（a）长时间注视预测　　（b）总注视时长预测

图4：以一组蜂群图呈现的解释摘要。每张蜂群图展示一个特征的 SHAP值分布。

在更高车速下的单次注视时长，因此需要更多次单独注视才能完成相同任务，进而导致总注视时长更长。

此外，我们可以看到，用户界面交互特征几乎没有负向贡献。这是因为总注视时长任务具有累积性，每次交互都必然会占用一定的视觉注意力。然而，主页栏交互 n_{Homebar}可能会对模型输出产生负面影响。与长时间注视预测中的观察结果一致，列表交互 n_{List}、地图交互n_{Map} 以及与 Android Auto 和 Apple CarPlay 的交互 n_{RemoteUI} 均可能对应更高的视觉需求预测。比较图4a 和图4b 还可以发现，与长时间注视概率相比，总注视时长对驾驶辅助系统状态的依赖程度较低。

为了更详细地了解单个特征对模型输出的影响，我们绘制了 SHAP值（纵轴）与相应特征值（横轴）的关系图。数据集中的每次次要任务参与都以一个点表示（见图5和图6）。同一横轴取值处的纵向离散表明，所示特征与其他特征之间存在非线性依赖关系。为突出特征间的交互，每个点都根据交互作用最强的特征值着色。图底部的直方图展示了数据点的分布。图5a 表明，使用自适应巡航控制会使长时间注视概率预测升高。与车速的交互显示，车速越高，这种影响往往越强。另一方面，图5b 所示数据表明，驾驶员在较低车速（低于 50 千米/小时）时倾向于延长单次注视时长，而在较高车速（高于 125 千米/小时）时则倾向于缩短单次注视时长。然而，在这两个速度范围之间，速度几乎不影响模型输出。这表明驾驶员会根据自己认为合适的速度，自我调节视觉注意力分配。此外，与 ACC状态的交互显示，ACC启用时，速度对模型输出的影响会减弱。与 a_{acc} 的交互效应在一定程度上解释了车速影响中的差异（纵向离散）。不过，所呈现的模型并未考虑道路类型或限速等可能同样影响车速如何作用于驾驶员视觉注意力分配的各种因素。

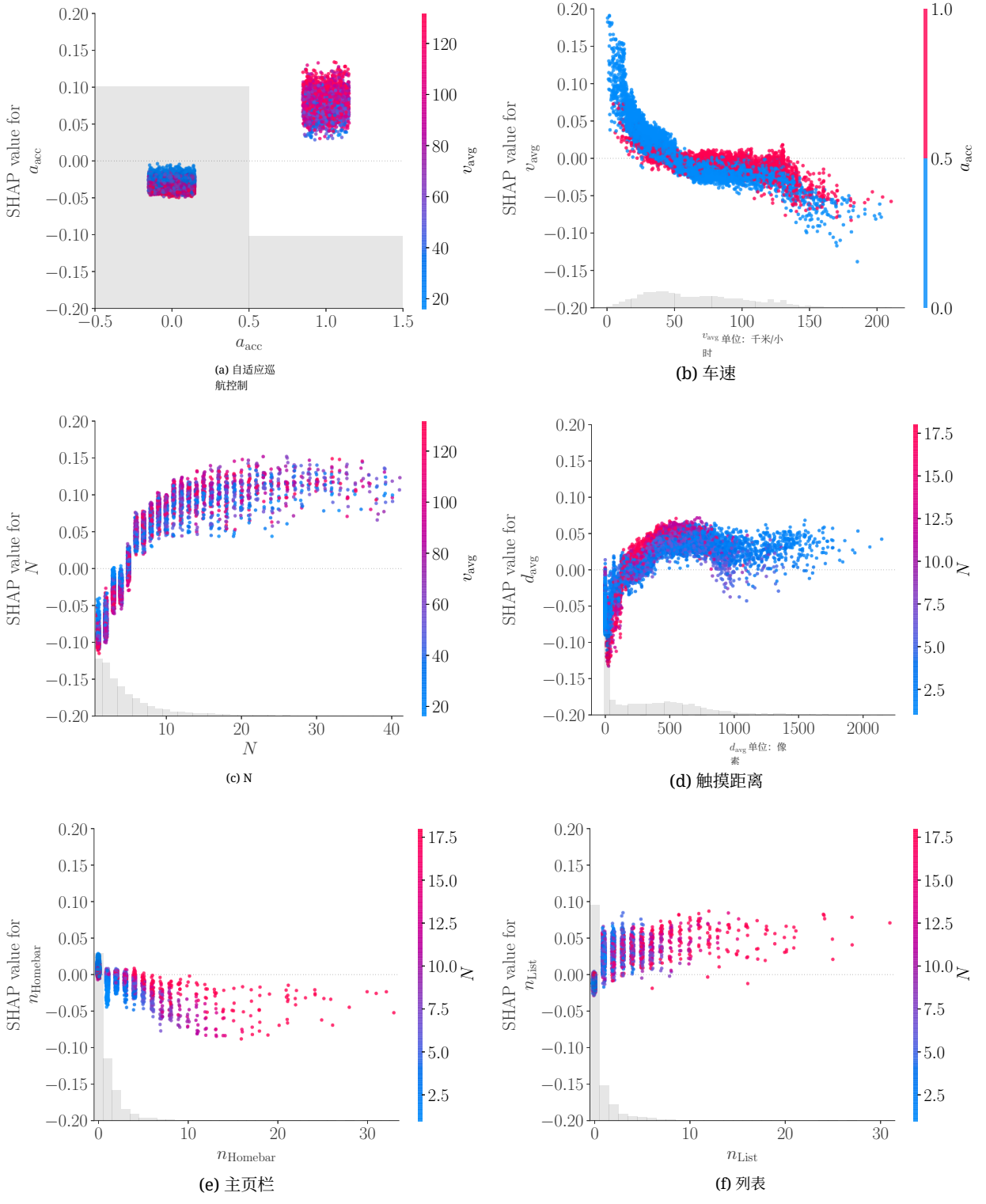
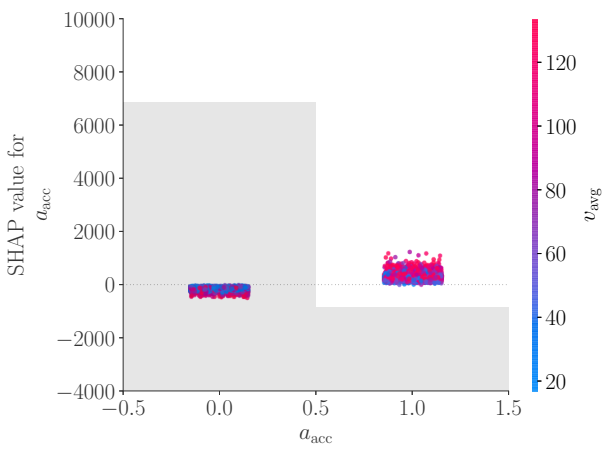
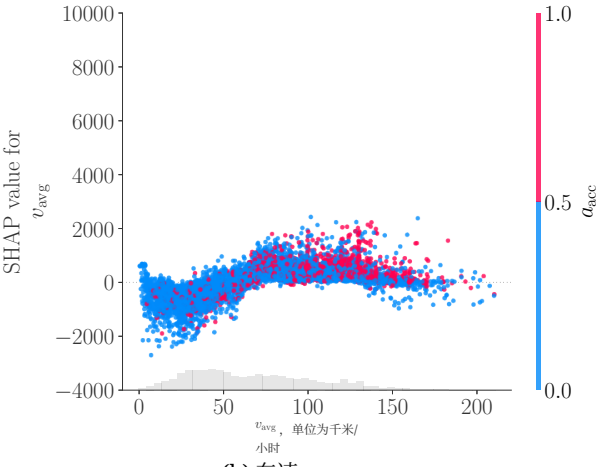


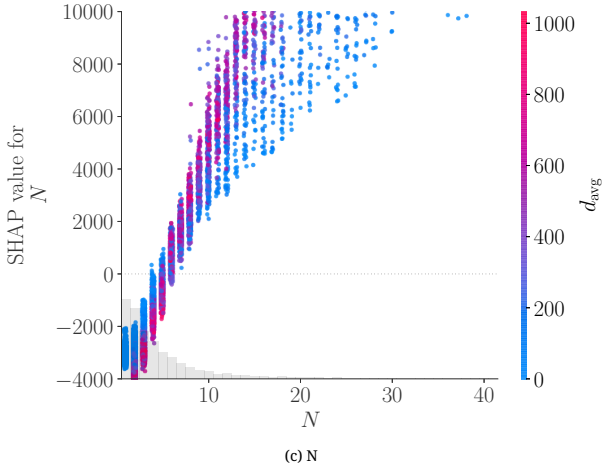
图5： 长时间注视分类模型的特征依赖图。



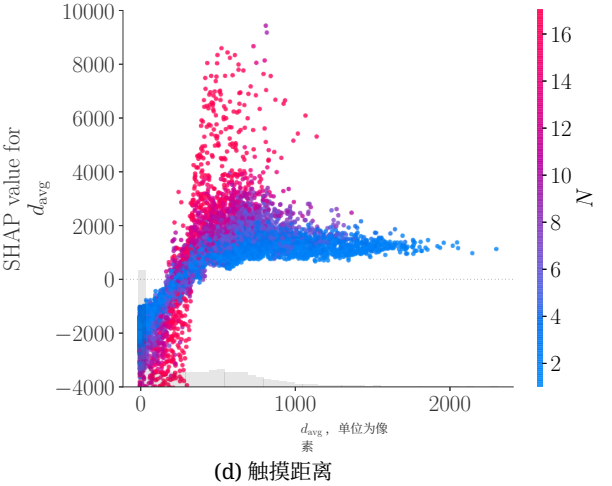
(a) 自适应巡航控制



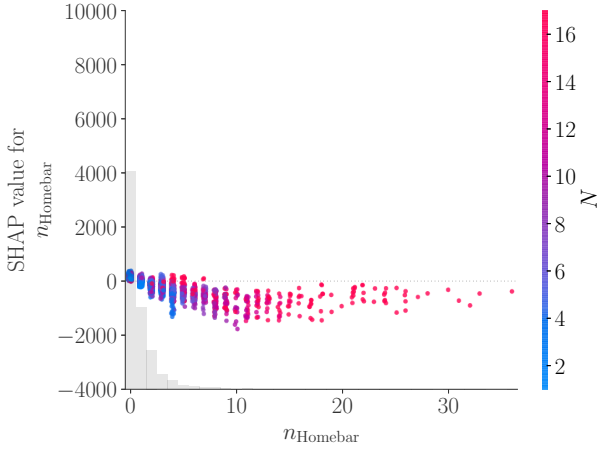
(b) 车速



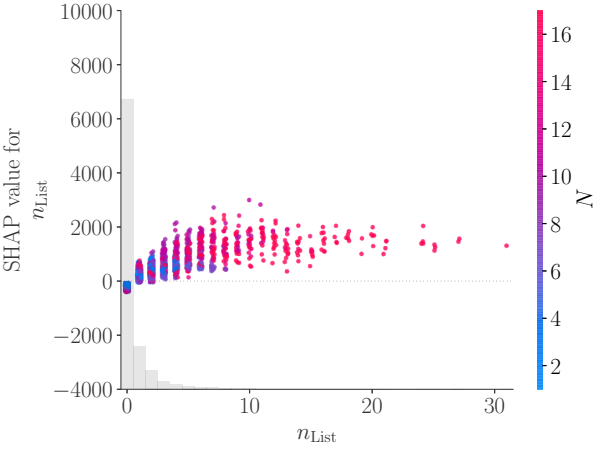
(c) N



(d) 触摸距离



(e) 主页栏



(f) 列表

图6： TGD模型的特征依赖图。

图5c 表明，交互次数与驾驶员发生长时间注视的概率呈正相关。另一方面，图5d 显示，当触摸交互之间的距离超过某个阈值（约为200px）后，其对长时间注视概率的影响保持不变。主页栏交互会降低模型预测发生长时间注视的概率（图5e），而列表交互（图5f）会促使模型预测发生长时间注视。与交互次数的交互效应还表明，两个要素在序列中的占比越高，其影响就越大。

图6 展示了不同特征如何影响TGD预测。交互次数 N （图6c）是主导特征，同时也是对其他所有特征具有最强交互效应的特征。与图5 相比，并且与我们在图4中观察到的情况一致，我们发现，ACC状态 a_{acc} （图6a）和车速v_{avg} （图6b）对TGD预测的影响不如对长时间注视预测的影响大。这一点在交互次数较少的次要任务参与中尤为明显。然而，列表交互n_{List} 和主页栏交互 n_{Homebar} 对TGD的影响与这些交互对长时间注视概率的影响相似（图6e和图6f）。平均触摸距离d_{avg}的影响也是如此。触摸距离增加会导致TGD和长时间注视概率上升，直至达到某个阈值。不过， N 与TGD预测模型的交互效应更强。另一个可能值得进一步探究的方面是x轴截距的位置。该点表示触摸距离达到何值时，特征的影响会从降低视觉需求预测转为提高视觉需求预测（图6d）。

5. 讨论

所提出的方法使用户能够评估早期原型的视觉需求。接下来，我们将结合相关背景审视研究结果，并证明所提出的方法比类似方法更准确。预测结果和解释有助于快速获得见解，无需开展成本高昂且筹划周期长的用户研究。我们通过评估文献中涵盖的三个示例性研究目标来说明这一点。最后，我们讨论该方法的若干局限性。

5.1. 预测车载触摸屏交互的视觉需求

鉴于建模任务的复杂性，所呈现的结果表明，借助大规模自然驾驶数据，机器学习方法能够帮助深入了解驾驶员的多任务行为。与Kujala和Salvucci(2015年)的方法相比，后者报告模型预测与观测结果之间存在显著差异；我们的方法不仅更准确，也涵盖了更多样化的用户界面元素。

在各种触摸屏交互和驾驶场景中，我们的方法预测 TGD 的平均绝对误差约为 2.4 秒。相比之下，Purucker等人(2017年)报告称，对所有评估任务取平均后，平均误差为 4 秒。此外，Purucker等人(2017年)采用恒定速度下的简单跟车任务进行评估。尽管这些比较有助于将研究结果置于相关背景中，但仍需考虑到，正如Janssen 等人(2020年)所述，各方法所处的环境和场景存在很大差异。

5.2. 基于真实道路驾驶数据的快速且易于获取的洞见

与传统用户研究相比，我们的方法有两大主要优势。首先，模型能够对尚未出现过的次要任务参与进行预测。传统研究只能用于评估经过明确测试的情境。其次，如果用户界面发生了重大变化（例如采用全新的设计或概念），用户研究的结果便不再有效，需要开展新的研究。同样，我们的计算模型也可能失去泛化能力。然而，自动化数据收集与建模方法的优势在于，一旦新版本部署到测试车辆中，便可收集数据，并根据该新版本拟合新的模型。为证明我们的方法是对传统用户研究方法的有意义补充，我们将研究结果与传统用户研究的结果进行了比较。

车速对驾驶员视觉注意力分配的影响。Risteska 等人（2021年）发现，车速提高会减少驾驶员长时间离路注视的次数。他们认为，驾驶员会根据驾驶需求调整视觉注意力分配。 Tivesten 和 Dozza （2014年）也得出了类似发现：驾驶员执行视觉手动手机任务时，车速与视线偏离道路注视时长之间存在显著相关性。这与我们在图5b 和图4a中展示的结果一致。我们的解释不仅表明，长时间注视的概率会随车速提高而降低，还进一步指出，这种行为可能并非严格成比例，也会受到驾驶辅助系统状态的影响。我们的结果还表明

，预测的总注视时长会随车速提高而增加（见图6b）。综合这两项发现，可以更全面地看出，车速较高时，驾驶员会缩短单次注视时长，因此不得不更频繁地看向中控台触摸屏。反过来，这会导致总注视时长增加，因为人类注视行为的某些方面（例如定位某个项目所需的时间）在每次注视中都是恒定的（Large 等人，2017b）。

驾驶自动化对驾驶员视觉注意力分配的影响。辅助驾驶与次要任务参与期间均值注视时长和总注视时长的增加有关（Large 等人，2017a；Carsten 等人，2012年；Ebel 等人，2022年）。这与图4中呈现的研究结果一致。在一项驾驶模拟器研究中，Carsten 等人（2012年）也发现，横向控制（半自动驾驶）对驾驶员参与的影响大于纵向控制（自适应巡航控制）的影响。根据我们的数据，我们无法证实这一发现。造成这种情况的原因可能有很多，但很可能与真实数据和模拟数据之间的差异有关。我们的结果也不同于Morando 等人（2019）的研究结果；他们报告称，手动驾驶与辅助驾驶的总体离路注视时长分布没有差异。他们仅报告了道路注视分布方面的影响，但指出其眼动追踪器无法提供有关离路兴趣区的详细信息。由于我们能够明确检测到朝向中控台触摸屏的注视，并将其与一般的离路注视区分开来，因此我们认为我们的结果更为可靠。

设计特征对驾驶员视觉注意力分配的影响。目前，尚未有许多方法像我们的方法一样，从元素类型的角度如此详细地研究设计特征对视觉需求的影响。Kujala和Salvucci（2015年）发现，连续两次触摸交互之间的平均距离，是与超过安全上限的长时间注视相关的一个关键因素。这与我们在图5d和6d中展示的结果一致。我们的方法所提供的解释还可以作为首次尝试，用于量化交互元素之间的空间分离对驾驶过程中视觉需求的影响。我们的方法还使我们能够详细说明各个元素的影响。迄今为止，文献中仅以大致相似的详细程度研究了任务交互时间（Green等人，2015年；Schneegaß等人，2011年）。我们发现，与地图、列表的交互，以及在Apple CarPlay和Android Auto中的交互，尤其具有较高的视觉需求。另一方面，在静态主屏栏上的交互、与应用图标的交互以及与一般按钮的交互，视觉需求较低。

5.3. 车载信息娱乐系统设计流程的益处及其对分心驾驶预防的启示

为了开发安全易用的车载信息娱乐系统，驾驶员分心评估从设计早期阶段起就应成为整体设计流程中不可或缺的一部分。然而，驾驶员分心是一个复杂的构念，汽车用户体验专家需要数据驱动的支持，以评估和比较不同设计方案的分心潜势（Ebel 等人，2021年b）。因此，我们的方法旨在自下而上地为车载信息娱乐系统的设计流程提供指导，从而开发出分心程度最低、设计上更安全的解决方案。我们设想将该方法用于动态评估处于早期阶段的车载信息娱乐系统设计。用户可以根据视觉需求，评估假设的车载信息娱乐系统设计所具有的分心潜势。用户还可以尝试使用人工输入样本，了解用户流程或驾驶场景的变化会如何影响驾驶员视觉注意力分配。随后，我们的方法会解释每个参数如何影响整体预测。因此，设计师可以更好地了解各种用户界面元素、驾驶自动化和车速对驾驶员分心的影响。设计师可以利用这些信息设计分心程度更低、能够降低事故风险的车载信息娱乐系统。与可比方法相比，本方法准确率更高，且三个应用示例表明，本方法能够显著帮助人们更好地理解驾驶员分心这一复杂构念，以及驾驶员在执行次要触摸屏任务期间的视觉注意力分配。

5.4. 局限性与未来工作

由于我们利用了研究合作伙伴已商业化的技术，因此收集了大量行为数据。我们观察了驾驶员自然的交互行为，没有明确告知他们要执行哪些触摸屏交互，因此消除了所谓的指令效应（Carsten 等人，2017年）。尽管这种方法有诸多优点，尤其是相较于模拟器研究和试车场研究，但仍存在若干局限性。下文将讨论这些局限性及其可能带来的影响。

数据收集仅涉及公司内部车辆。这些车辆除了用于各种测试程序外，也供员工通勤和休闲出行使用，例如周末出行。我们认为，即使驾驶员遵循旨在评估驾驶相关功能的测试规程，他们使用车载信息娱乐系统的意愿也与真实世界行为相差不大。此外，本研究中的所有驾驶员都应视为专家用户。不过，专家的注视行为与普通用户的注视行为究竟有多大差异，目前尚不完全清楚

。Wikman 等人(1998)报告称，经验丰富的驾驶员能够更恰当地分配视觉注意力(Wikman 等人，1998)，Naujoks 等人(2016年)则表明，与新手用户相比，经验丰富的高级驾驶辅助系统用户倾向于增加次要任务参与。然而，与相关研究方法(Gaspar 和 Carney，2019；Morando 等人，2019；Noble 等人，2021年)的比较显示，在注视总量和平均注视行为方面高度一致。尽管如此，在解读结果时，仍需考虑驾驶员样本有限这一点，以及他们都是独自驾驶这一事实。

需要注意的是，本研究使用的特征并未涵盖影响驾驶员视觉注意力分配的所有因素。本研究仅考虑驾驶自动化等级、车速和方向盘转角来描述驾驶情境。这些特征及其交互效应提供了有价值的信息（比较附录中的图5a、5b、6b、6a和图7），但不足以全面描述驾驶情境。例如，车速的影响不仅会因驾驶自动化等级而异，也会因道路类型和交通状况而异。因此，纳入更多特征不仅可以更全面地描述驾驶情境，还可以通过考虑交互效应，使现有特征更有意义。

此外，重要的是结合具体情境理解研究结果，并进一步阐述其可能产生的实际影响。如研究所示，该方法能够深入揭示设计元素和环境因素如何影响驾驶员视觉注意力分配。预测结果和解释可以指导设计师打造不易造成分心、使用起来更安全的界面。然而，尽管我们的方法优于相关方法，其准确度仍不足以实现像素级精确预测，也无法区分驾驶环境中的细微变化（例如，以 72 千米/小时而非 75 千米/小时的速度行驶）。若要可靠地评估此类细微变化的影响，甚至将其作为制定驾驶员分心指南的依据，就必须提高准确率。此外，我们未考虑光照条件或道路类型（例如乡村道路或高速公路）等环境因素，也未考虑元素颜色和大小等可能影响视觉注意力分配的用户界面元素。纳入这些特征可以更全面地呈现情况，也可能提高预测准确度。此外，驾驶员往往会根据驾驶任务需求，自我调节参与次要任务的意愿（Ebel 等人，2022年；Oviedo-Trespalacios 等人，2018；Hancox 等人，2013）。因此，在某些驾驶情境中，部分交互发生得较少，导致训练数据较少。由此可见，预测准确度可能因驾驶情境而异。

这些解释并不意味着因果关系，因此不能全面评估驾驶员参与次要触摸屏任务时的视觉注意力分配。不过，这些解释有助于设计师识别输入特征与模型输出之间最具信息量的关系，从而帮助他们理解机器学习模型预测的视觉需求。

既然已证明该方法能够取得令人鼓舞的结果，后续迭代的主要目标便是提高预测准确率。首先，通过增加光照条件、周边道路使用者的距离或地图数据等特征，更全面地描述驾驶情境，或许能显著改善效果。其次，考虑年龄或驾驶经验等用户人口统计特征，也可能提高准确率。最后，扩大数据集不仅可能有利于本文提出的算法，也将使得采用循环神经网络等更复杂的方法成为可能，从而捕捉交互序列中包含的序列信息。

6. 结论

本文提出了一种机器学习方法，根据驾驶过程中所执行的交互类型及相关驾驶参数，预测次要任务触摸屏交互的视觉需求。我们的方法生成局部和全局解释，揭示设计元素和驾驶参数如何影响驾驶员的视觉注意力分配。我们在一个包含 12,142 次次要任务参与的真实道路驾驶数据集上评估了该方法。我们的最佳模型能够以 68 % 的准确率识别驾驶员进行长时间注视的次要任务参与，并以 2.4 s 的均值偏差预测 TGD。对生成的解释进行分析后发现，不同触摸屏交互的视觉需求存在明显差异，也表明驾驶员的视觉注意力分配取决于驾驶情境。与相关研究（Risteska 等人，2021年；Tivesten 和 Dozza，2014年）一致，我们表明，驾驶员会根据车速和驾驶自动化等级调整其视觉注意力分配。

我们的主要贡献回应了以往研究（Risteska 等人，2021年；Large 等人，2017b；Janssen 等人，2015年；Victor 等人，2014年）提出的诸多期望：（1）该方法利用持续采集的大规模真实世界数据，对驾驶员执行次要任务期间的视觉注意力分配进行逼真的预测

。（2）该方法易于调整，可纳入更多特征，并预测除 TGD 和长时间注视概率之外的多种指标（例如凝视次数、总离路注视时间、平均凝视时长）。（3）局部和全局解释详细揭示了设计元素和场景参数对驾驶员分心预测的影响。（4）该方法可帮助设计师了解其设计可能带来的影响，并指导他们设计安全易用的车载触摸屏界面。

参考文献

Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y., 2004 111, 1036 1060 10 1037 0033 295 111 4 1036。《整合心智理论》。《心理学评论》–。doi:。/ - x. . . . Angell, L., Perez, M., Hankey, J., 2008。辅助信息系统的驾驶员使用模式，载于：《首届自然驾驶方法与分析人因研讨会》特邀论文。Bärgman, J., Lisovskaja, V., Victor, T., Flannagan, C., Dozza, M., 2015。注视行为如何影响碰撞与伤害风险？基于SHRP2碰撞事故和险情数据的“假设情境”反事实模拟 35, 152 169 10 1016 2015 10 011。《交通研究F辑：交通心理学与行为》–。doi:。/j.trf. . . . Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F., 2020。可解释人工智能（XAI）：迈向负责任人工智能的概念、分类体系、机遇与挑战。《信息融合》58, 82–115。doi:10.1016/j.inffus.2019.12.012。Burns, P., Harbluk, J., Foley, J.P., Angell, L., 2010。评估车载任务的分心潜势时任务持续时间及相关测量指标的重要性，载于：第2届国际汽车用户界面与交互式车载应用会议论文集 - AutomotiveUI 10，ACM出版社。doi:10.1145/1969773.1969776。Card, S., 1983。《人机交互心理学》。L. Erlbaum Associates，希尔斯代尔，新泽西州。Card, S.K., Moran, T.P., Newell, A., 1980。交互系统用户操作时间的击键层级模型。《ACM通讯》23, 396–410。doi:10.1145/358886.358895。Carsten, O., Hibberd, D., Bärgman, J., Kovaceva, J., Pereira Cocron, M.S., Dotzauer, M., Utesch, F., Zhang, M., Stemmler, E., Guyonvarch, L., Sagberg, F., Forcolin, F., 2017。UDRIVE交付成果43.1：《驾驶员分心与注意力不集中》，欧盟第七框架计划UDRIVE项目。第一版，UDRIVE联盟，比利时。Carsten, O., Lai, F.C.H., Barnard, Y., Jamson, A.H., Merat, N., 2012。半自动驾驶中的控制任务替代。《人因：人因与工效学学会期刊》54, 747–761。doi:10.1177/0018720812460246。Custer, K., 2018。百车数据。doi:10.15787/VTT1/CEU6RB。de Winter, J.C., Happee, R., Martens, M.H., Stanton, N.A., 2014。自适应巡航控制和高度自动化驾驶对工作负荷和情境意识的影响：实证证据综述。《交通研究F辑：交通心理学与行为》27, 196–217。doi:10.1016/j.trf.2014.06.016。Dingus, T.A., Guo, F., Lee, S., Antin, J.F., Perez, M., Buchanan-King, M., Hankey, J., 2016。利用自然驾驶数据评估驾驶员碰撞风险因素及其普遍程度。《美国国家科学院院刊》113, 2636–2641。doi:10.1073/pnas.1513271113。Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sud weeks, J., Perez, M.A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z.R., Jermeland, J., Knipling, R.R., 2006。《百车自然驾驶研究：第二阶段——百车实地实验结果》。doi:10. 1037/e624282011-001。Donmez, B., Boyle, L.N., Lee, J.D., 2010。视线离开道路的注视差异：对年轻驾驶员表现的影响。《交通工程学报》136, 403–409。Doshi-Velez, F., Kim, B., 2017。迈向可解释机器学习的严谨科学。arXiv预印本 arXiv:1702.08608 arXiv:1702.08608。Ebel, P., Berger, M., Lingenfelder, C., Vogelsang, A., 2022。驾驶员如何自我调节次要任务参与？驾驶自动化对触摸屏交互和注视行为的影响，载于：第14届国际汽车用户界面与交互式车载应用会议论文集，ACM，韩国首尔。第263–273页。doi:10.1145/3543174.3545173。Ebel, P., Brokhausen, F., Vogelsang, A., 2020。实地用户交互数据在汽车用户体验开发生命周期中的作用与潜力：行业视角，载于：第12届国际汽车用户界面与交互式车载应用会议，ACM，线上会议，美国哥伦比亚特区。第141–150页。doi:10.1145/3409120.3410638。Ebel, P., Lingenfelder, C., Vogelsang, A., 2021a。用于评估车载信息系统用户行为的事件序列数据可视化，载于：第13届国际汽车用户界面与交互式车载应用会议，计算机协会，纽约，美国。第1–10页。doi:10.1145/3409118.3475140。Ebel, P., Orlovska, J., Hünemeyer, S., Wickman, C., Vogelsang, A., Söderberg, R., 2021b。汽车用户体验设计与数据驱动开发：缩小差距以支持从业者。《交通研究：跨学科视角》11, 100455。doi:10.1016/j.trip.2021. 100455。Engström, J., Johansson, E., Östlund, J., 2005。真实与模拟高速公路驾驶中的视觉与认知负荷效应。《交通研究F辑：交通心理学与行为》8, 97–120。doi:10.1016/j.trf.2005.04.012。Fisher, D.L.（编），2011。《工程、医学与心理学驾驶模拟手册》。CRC出版社，博卡拉顿。Fitts, P.M., 1954。人体运动系统在控制运动幅度时的信息容量。《实验心理学杂志》47, 381–391。doi:10.1037/h0055392。Gaspar, J., Carney, C., 2019。部分自动化对驾驶员注意力的影响：一项自然驾驶研究。《人因：人因与工效学学会期刊》61, 1261–1276。doi:10.1177/0018720819836310。Green, P., 1999。《驾驶员信息系统的视觉与任务需求》。技术报告。密歇根大学交通研究

研究所（UMTRI）。Green, P., Kang, T., Lin, B., 2015。用于改进SAE建议实践J2365的触摸屏任务元素时间：首份提案。Hancox, G., Richardson, J., Morris, A., 2013。驾驶员使用手机的意愿：手机功能和道路需求的影响。《IET智能交通系统》7, 215–222。doi:10.1049/iet-its.2012.0133。Horrey, W.J., Wickens, C.D., 2007。车内注视时长。《交通研究委员会记录》2018, 22–28。doi:10.3141/2018-04。Hutchinson, T., White, K., Martin, W., Reichert, K., Frey, L., 1989年11–12月。使用眼动输入的人机交互。《IEEE系统、人和控制论汇刊》19, 1527–1534。doi:10.1109/21.44068。ISO 15007, 2020。《道路车辆——与交通信息和控制系统相关的驾驶员视觉行为测量与分析》。标准。国际标准化组织。瑞士日内瓦。Janssen, C.P., Boyle, L.N., Ju, W., Riener, A., Alvarez, I., 2020。智能体、环境、情景：考察人车交互模型与模拟的框架。《交通研究：跨学科视角》8, 100214。doi:10.1016/j.trip.2020.100214。Janssen, C.P., Gould, S.J., Li, S.Y., Brumby, D.P., Cox, A.L., 2015。整合不同视角和研究方法下的多任务与中断知识。《国际人机研究期刊》79, 1–5。doi:10.1016/j.ijhcs.2015.03.002。Kanaan, D., Ayas, S., Donmez, B., Risteska, M., Chakraborty, J., 2019。利用基于车辆的自然驾驶数据预测分心和环境需求。《国际移动人机交互期刊》11, 59–70。doi:10.4018/ijmhci.2019070104。Kaptein, N.A., Theeuwes, J., van der Horst, R., 1996 1550, 30 36 10 1177 0361198196155000105, 10 1177 0361198196155000105。驾驶模拟器效度：若干考量。《交通研究委员会记录》–。doi:。/arXiv:https://doi.org/。/。Klauer, S., Dingus, T., Neale, T., Sudweeks, J., Ramsey, D., 2006。《驾驶员注意力不集中对险情/碰撞风险的影响：基于百车自然驾驶研究数据的分析》。技术报告。美国交通部、美国国家公路交通安全管理局/弗吉尼亚理工运输研究所。运输研究广场3500号（0536），弗吉尼亚州布莱克斯堡24061。Kohavi, R., 1995。交叉验证和自助法在准确率估计与模型选择中的研究，载于：第14届国际人工智能联合会议论文集——第2卷，Morgan Kaufmann出版社，美国加利福尼亚州旧金山。第1137–1143页。Kujala, T., Salvucci, D.D., 2015。车载显示屏视觉采样建模：预测安全关键控制失误的挑战。《国际人机研究期刊》79, 66–78。doi:10.1016/j.ijhcs.2015.02.009。Kutila, M., Jokela, M., Markkula, G., Rue, M.R., 2007。利用摄像头视觉系统检测驾驶员分心，载于：2007年IEEE国际图像处理会议，IEEE。doi:10.1109/icip.2007.4379556。Large, D., Banks, V., Burnett, G., Baverstock, S., Skrypchuk, L., 2017a。探索不同自动化等级下分心驾驶员的行为。Large, D.R., Burnett, G., Crundall, E., van Loon, E., Eren, A.L., Skrypchuk, L., 2017b。开发预测方程以模拟车载触摸屏人机界面的视觉需求。《国际人机交互期刊》34, 1–14。doi:10.1080/10447318.2017.1306940。Lee, S.C., Yoon, S.H., Ji, Y.G., 2019。使用击键层级建模，预测驾驶时车载信息系统的任务完成时间。《国际工业工效学期刊》72, 252–260。doi:10.1016/j.ergon.2019.06.001。Li, L., Zhong, B., Hutmacher, C., Liang, Y., Horrey, W.J., Xu, X., 2020。通过基于图像的手部和耳部识别检测驾驶员手动分心。《事故分析与预防》137, 105432。doi:10.1016/j.aap.2020.105432。Li, Z., Bao, S., Kolmanovsky, I.V., Yin, X., 2018。利用自然驾驶数据和驾驶表现指标检测视觉-手动分心。《IEEE智能交通系统汇刊》19, 2528–2535。doi:10.1109/tits.2017.2754467。Liang, Y., Lee, J.D., 2010。结合认知分心和视觉分心：效果并非简单相加。《事故分析与预防》42, 881–890。doi:10.1016/j.aap.2009.05.001。Liao, Q.V., Gruen, D., Miller, S., 2020。质疑人工智能：为可解释人工智能用户体验提供设计实践指导，载于：2020年CHI计算系统中的人因会议论文集，ACM，美国夏威夷州檀香山。第1–15页。doi:10.1145/3313831.3376590。Lipton, Z.C., 2018。模型可解释性的迷思。《Queue》16, 31–57。doi:10.1145/3236386.3241340。Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020。利用面向树模型的可解释人工智能，从局部解释走向全局理解。《自然机器智能》2, 56–67。doi:10.1038/s42256-019-0138-9。Lundberg, S.M., Erion, G.G., Lee, S.I., 2018。树集成模型的一致性个体化特征归因。arXiv预印本 arXiv:1802.03888 arXiv:1802.03888。Lundberg, S.M., Lee, S.I., 2017。解释模型预测的统一方法，载于：第31届神经信息处理系统国际会议论文集，Curran Associates Inc.，美国纽约州雷德胡克。第4768–4777页。Manes, D.I., 1997。利用击键层级模型预测目的地输入和检索时间。Merchant, J., 1967。《眼动仪》。技术报告NASA-CR-805。美国国家航空航天局。Merlhiot, G., Bueno, M., 2021。困倦和分心如何影响接管表现：系统性综述与元分析。《事故分析与预防》，106536, doi:10.1016/j.aap.2021.106536。Molnar, C., 2020。《可解释机器学习》。Lulu.com。Morando, A., Gershon, P., Mehler, B., Reimer, B., 2021。视觉注意力与方向盘控制：从参与到脱离特斯拉Autopilot。《人因与工效学学会年会论文集》65, 1390–1394。doi:10.1177/1071181321651118。Morando, A., Victor, T., Dozza, M., 2019。自动化情境下驾驶员注意力的参考模型：横向与纵向辅助期间注视行为的变化。《IEEE智能交通系统汇刊》20, 2999–3009。doi:10.1109/tits.2018.2870909。Naujoks, F., Purucker, C., Neukum, A., 2016。次要任务参与与车辆自动化——比较不同自动化等级在道路实地实验中的影响。《交通研究F辑：交通心理学与行为》38, 67–82。doi:10.1016/j.trf.2016. 01.011。美国国家公路交通安全管理局，2012。《NHTSA车载电子设备驾驶员视觉-手动分心指南》。标准。美国国家公路交通安全管理局（NHTSA）。Noble, A.M., Miles, M., Perez, M.A., Guo, F., Klauer, S.G., 2021。评估驾驶员眼动注视行为和次要任务参与情况，同时

使用驾驶自动化系统。《事故分析与预防》151，105959。doi：10.1016/j.aap.2020.105959。Oviedo-Trespalacios, O., Haque, M.M., King, M., Washington, S., 2018. 我应该在这里发短信还是打电话？基于情境分析驾驶员认为自己参与手机多任务处理的可能性：手机多任务处理参与度。《风险分析》38，2144–2160。doi：10.1111/risa.13119。Pampel, S.M., Burnett, G., Hare, C., Singh, H., Shabani, A., Skrypchuk, L., Mouzakitis, A., 2019. Fitts模型驶上高速公路，载于：第11届国际汽车用户界面与交互式车载应用会议论文集，ACM。doi：10.1145/3342197.3344538。Pettitt, M., Burnett, G., Stevens, A., 2007年。扩展的击键层级模型（KLM）：预测车载信息系统的视觉需求，载于：SIGCHI计算机系统中的人为因素会议论文集，ACM。doi：10.1145/1240624.1240852。Purucker, C., Naujoks, F., Prill, A., Neukum, A., 2017年。通过击键层级建模预测视线离开道路的总时长，评估驾驶过程中车载信息系统造成的分心。《应用人体工程学》58，543–554。doi：10.1016/j.apergo.2016.04.012。Ribeiro, M.T., Singh, S., Guestrin, C., 2016年。“我为什么要相信你？”，载于：第22届ACM SIGKDD知识发现与数据挖掘国际会议论文集，ACM。doi：10.1145/2939672.2939778。Riener, A., 2011年。评估模拟器逼真度和有效性：模拟器研究与道路实测研究。《系统与测量进展国际期刊》3，110–124 (15)。Risteska, M., Kanaan, D., Donmez, B., Chen, H.Y.W., 2021年。驾驶需求对分心参与和注视行为的影响：自然驾驶数据的研究结果。《安全科学》136，105123。doi：10.1016/j.ssci.2020.105123。SAEJ3016，2021年。SAE J3016：道路机动车辆驾驶自动化系统相关术语的分类与定义。标准。汽车工程师学会（SAE）。沃伦代尔。Schneegaß, S., Pfleging, B., Kern, D., Schmidt, A., 2011年。支持汽车用户界面交互建模，载于：第3届国际汽车用户界面与交互式车载应用会议论文集，计算机协会，纽约，美国。第71–78页。doi：10.1145/2381416.2381428。Shapley, L.S., 1953年。n人博弈的一个值，载于：《博弈论研究》（AM-28），第二卷。普林斯顿大学出版社，第307–318页。doi：10.1515/9781400881970-018。Shin, D., 2021年。可解释性与可解释能力对感知、信任和接受度的影响：对可解释人工智能的启示。《国际人机研究期刊》146，102551。doi：10.1016/j.ijhcs.2020.102551。Tivesten, E., Dozza, M., 2014年。驾驶情境和手机视觉-手动任务会影响自然驾驶中的注视行为。《交通研究F辑：交通心理学与行为》26，258–272。doi：10.1016/j.trf.2014.08.004。Tsimhoni, O., Green, P., 2001年。驾驶的视觉需求与显示密集型车载任务的执行。《人因与工效学学会年会论文集》45，1586–1590。doi：10.1177/154193120104502305。Verma, T., Lingenfelder, C., Klakow, D., 2020年。在人工智能语境中定义解释，载于：第三届BlackboxNLP自然语言处理神经网络分析与解释研讨会论文集，计算语言学协会。doi：10.18653/v1/2020.blackboxnlp-1. 29。Victor, T., Dozza, M., Bärgman, J., Boda, C.N., Engström, J., Markkula, G., 2014年。自然驾驶研究数据分析：更安全的注视、驾驶员注意力不集中与碰撞风险。Wang, C., An, P., 2021年。一款通过与日常环境交互，帮助非专业人士理解预训练卷积神经网络的移动工具，载于：第23届移动人机交互国际会议附属出版物，ACM，图卢兹及线上，法国。第1–5页。doi：10.1145/3447527.3474873。Wiegand, G., Eiband, M., Haubelt, M., Hussmann, H., 2020年。“我希望能对此得到解释！”探索人们对意外自动驾驶的反应，载于：第22届移动设备与服务人机交互国际会议，ACM，奥尔登堡，德国。第1–11页。doi：10.1145/3379503.3403554。Wikman, A.S., Nieminen, T., Summala, H., 1998年。驾驶经验与在不同宽度道路上执行车内任务时的时间分配。《人体工效学》41，358–372。doi：10.1080/001401398187080。Wintersberger, P., Schartmüller, C., Shadeghian-Borojeni, S., Frison, A.K., Riener, A., 2021年。在测试跑道上评估配备真实自动化系统的即将到来的接管请求。《人因：人因与工效学学会期刊》，001872082110514doi：10.1177/ 00187208211051435。Wollmer, M., Blaschke, C., Schindl, T., Schuller, B., Farber, B., Mayer, S., Trefflich, B., 2011年。使用长短期记忆网络进行在线驾驶员分心检测。《IEEE智能交通系统汇刊》12，574–582。doi：10.1109/tits.2011.2119483。Zhang, Y., Liao, Q.V., Bellamy, R.K.E., 2020年。置信度和解释对人工智能辅助决策中准确率与信任校准的影响，载于：2020年公平性、问责制与透明度会议论文集，ACM，巴塞罗那，西班牙。第295–305页。doi：10.1145/ 3351095.3372852。

A. 附录

A.1. 定义

定义 1： 一次交互序列 $I = (i_n)_{n=1}^{N}$ 是指在一次行程中记录的一系列触摸屏交互，其中，i_n 是用户执行的一次触摸屏交互，N 表示 I的交互次数。一次触摸屏交互 $i = (t, e, p, c)$ 由其时间戳 t、元素类型 e、手势类型 p 和坐标对 $c = (x, y)$组成。在 I中，连续两次交互 $t(i_{n+1}) - t(i_n)$ 之间的持续时间必须小于 Δt_{max}，使得 $t(i_{n+1}) - t(i_n) \leq \Delta t_{max}$。

定义 2： A注视序列 $G = (g_n)_{n=1}^{N}$ 是由驾驶员注视构成的一系列互不重叠的时间区间，其中g_n 是用户进行的一次注视，N 表示 G的注视次数。每次注视 $g_n = (t^s, t^e, r)_n$ 由

其开始时间 t^s、结束时间 t^e和兴趣区域（AOI） r构成，用于描述在 t^s 与 t^e之间注视的位置。对于一次行程中除第一次注视外的所有注视，其开始时间等于前一次注视 $t_s(g_n) = t_s(g_{n-1})$的结束时间。

定义 3： A 驾驶序列 $D = (d_n)_{n=1}^N$ 是由驾驶数据观测组成的序列，其中 d_n 是单个观测， N 表示 D的观测数量。每个观测定义为 $d_n = (t, v, \theta, a_{ACC}, a_{SA})_n$，其中 t 表示时间戳， v 表示车速， θ 表示方向盘转角，a_{ACC} 和 a_{SA} 分别表示 ACC 和 SA 的状态。

定义 4： A 次要任务参与 S 定义为一个交互序列及其对应的注视序列和驾驶序列 $S = (I, G, D)$。我们考虑从首次交互之前开始、直到末次交互之后的所有驾驶观测，使得 $t(i_1) - t_b < t(d_n) < t(i_N) + t_b$。其中， t_b 表示缓冲时长。对于注视序列 G，我们考虑其开始时间或结束时间落在 I 的首次交互与末次交互之间的所有注视，使得 $t(i_1) < t^s(g_n) < t(i_N) \vee t(i_1) < t^e(g_n) < t(i_N)$。

问题 1： 我们将长时间注视预测任务 定义为：识别所有发生长时间注视的次要任务参与S。对于任意 $g_n \in G$，给定相应的交互序列 s^i 和驾驶序列 s^d

问题 2： 我们将总注视时长预测任务 定义为：预测次要任务参与 S期间对中控台触摸屏的 TGD。

A.2. 超参数优化

在下文中　　我们报告各个模型的超参数优化结果　　.

A.2.1. 随机森林模型

实现及相关说明基于 scikit-learn Python 软件包。**n_估计器** = [100, 200, 400, 800, 1200, 1600, 2000] – 森林中的树的数量。**最大值_特征** = [“auto”、“sqrt”] 寻找最佳划分时考虑的特征数量**最大值_深度** = [10, 20, 30, 40, 60, 80, 100] – 树的最大深度。**最小值_样本_划分** = [2, 5, 10] – 划分内部节点所需的最少样本数。**最小值_样本_叶节点** = [1, 2, 4] – 叶节点所需的最少样本数。**自助采样** = [True、False] – 构建树时是否使用自助采样样本。

表3随机森林模型中表现最佳的参数组合。

特征	长时间注视预测	TGD预测
n_估计器	200	1600
max_特征	auto	auto
max_深度	10	60
min_样本 _划分	5	2
min_样本 _leaf	2	4
自举法	True	True

A.2.2. XGBoost 模型

实现和描述基于 XGBoost Python 软件包。**n_估计器** = [20, 100, 500, 1000, 5000, 10000, 20000] – 提升轮数。**子采样** = [0.2,0.4,0.6,0.8,1] – 训练实例的子采样比例。**最大值_深度** = [5,10,50,100] – 基学习器的最大树深度。**学习_率** = [0.0005, 0.001, 0.01, 0.1, 1] – 提升学习率（xgb 的“eta”）**列采样_按树** = [0.2,0.4,0.6,0.8,1] – 构建每棵树时的列子采样比例。**列采样_按层级** = [0.2,0.4,0.6,0.8,1] – 每层的列子采样比例。

A.2.3. 前馈神经网络

该实现和超参数优化使用了 Keras API。需要注意的是，不同的超参数组合在预测性能上没有显著差异。

表4XGBoost 模型的最佳参数组合。

特征	长时间注视预测	TGD预测
n_估计器	5000	5000
子采样	0.6	0.8
max_深度	10	10
min_子节点_权重	4	10
学习_rate	0.01	0.0005
列采样_按树采样	0.2	0.6
列采样_按层采样	0.2	1

n_隐藏_层 = [1,2,3,4,5] – 神经网络输入层与输出层之间的层数**n_神经元** = [32, 64, 128, 256, 512]) – 每层的神经元数量。
激活函数 = ["ReLU"、"Sigmoid"] – 相应层中神经元的激活函数。**丢弃_率** = [0,0.1,0.2,0.3] – 训练期间随机将单元设为零的概率。**学习_率** = [0.01,0.001,0.0001] – ADAM 优化器的初始学习率。

表5前馈神经网络模型中表现最佳的参数组合。

特征	长时间注视预测	TGD预测
n_隐藏_层	3	1
学习_rate	0.0001	0.001
n_第1层神经元	512	512
第1层激活函数	Sigmoid	relu
drop_第1层输出	0.3	0.1
n_第2层神经元	64	-
第2层激活层	relu	-
drop_第2层输出	0.1	-
n_第3层神经元	256	-
第3层激活层	Sigmoid	-
drop_第3层输出层	0.1	-

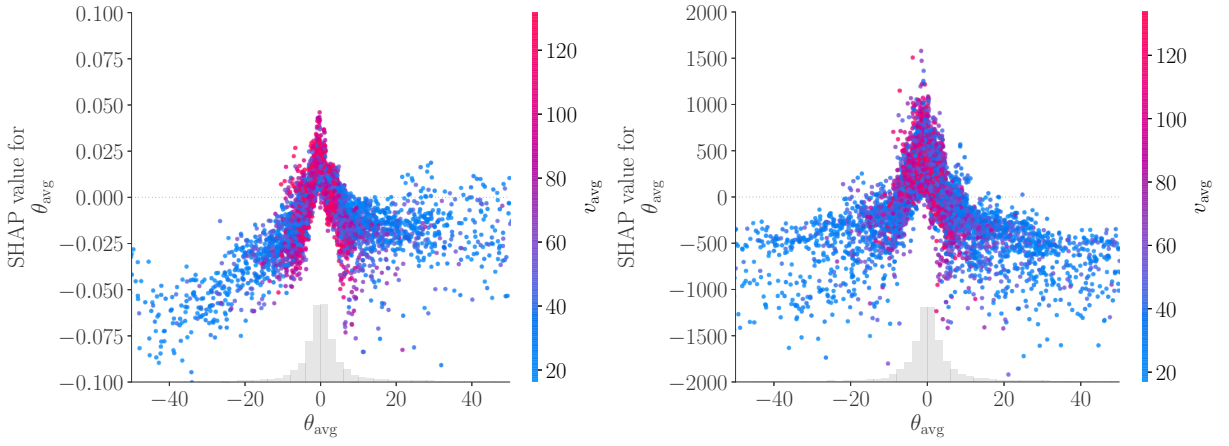
A.3. 数据集汇总统计

表6概述了所有特征。

表6数据集汇总统计

统计量	Mean	标准差	Min	Q_1(25)	中位数	Q_3(75)	Max
交互次数	4.431	4.993	1	1	3	5	41
点击手势次数	3.814	4.495	0	1	2	5	40
拖动手势次数	0.363	1.324	0	0	0	0	31
多点触控手势次数	0.240	1.108	0	0	0	0	26
平均注视时长（毫秒）	1,441.491	929.736	120.000	960.000	1,241.000	1,659.000	26,801.000
凝视次数	4.367	4.998	1	1	3	6	50
长时间注视次数	0.569	1.102	0	0	0	1	13
总注视时长（毫秒）	5,742.751	7,049.487	120.000	1,590.500	3,499.000	7,354.000	262,416.000
平均速度（千米/小时）	70.516	36.935	0.633	40.881	66.230	96.567	209.883
自适应巡航控制已启用	0.206	0.404	0	0	0	0	1
半自动驾驶已启用	0.099	0.299	0	0	0	0	1
应用图标交互	0.196	0.586	0	0	0	0	13
CoverFlow 交互	0.038	0.434	0	0	0	0	16
未知交互	0.049	0.450	0	0	0	0	23
其他交互	0.731	1.568	0	0	0	1	37
列表交互	0.518	1.652	0	0	0	0	31
选项卡交互	0.385	1.335	0	0	0	0	35
控制栏交互	0.012	0.135	0	0	0	0	4
按钮交互	0.640	1.515	0	0	0	1	36
主页栏交互	0.892	2.032	0	0	0	1	36
滑块交互	0.015	0.228	0	0	0	0	9
点击保护交互	0.058	0.313	0	0	0	0	8
弹出窗口交互	0.030	0.232	0	0	0	0	9
键盘交互	0.184	1.435	0	0	0	0	31
地图交互	0.508	2.181	0	0	0	0	39
远程用户界面交互	0.173	1.179	0	0	0	0	31
浏览器交互	0.002	0.093	0	0	0	0	8

A.4. 方向盘特征依赖图



(a) 长时间注视预测样例。

(b) 总注视时长预测样例。

图 7： 方向盘转角及其与车速交互项在长时间视线转移分类和 TGD模型中的特征依赖图。