基于卷积神经网络的视线跟踪技术研究进展

李钰卿[1]，战荫伟[1]，杨卓[1]，

(广东工业大学计算机学院，广东 广州 510000)

摘 要：专业级眼动仪和消费级眼动仪通常使用传统的基于特征的视线追踪方法，由于这种方法依赖红外光源，具有高成本的特点，不利于视线跟踪技术的广泛应用。基于卷积神经网络的视线跟踪属于基于外观的视线跟踪的一种，是一个新的研究方向；相比于其他基于外观的方法，它具有端到端、不需校准、可在自然光下追踪和适用多种平台等优点。本文对基于外观的视线追踪做了调研和总结，详细介绍了基于卷积神经网络的视线追踪方法，重点讨论了两个关键问题：数据集的采集和 CNN 网络结构的选取。提出理想的数据集应包含头部姿态范围广、光照变化多等特点；视线跟踪 CNN 的网络结构中多输入优于单输入的特性。

关键词：视线跟踪；注视点；卷积神经网络；人机交互

中图分类号：TPxxx〔查询请参考 http://ztflh.xhma.com〕

A survey on CNN based eye tracking methods

Yuqing Li, Yinwei Zhan, Zhuo yang ,

(GDUT, GuangDong, Guangzhou 510000, China)

Abstract: Professional and consumer eye tracking systems usually use feature-based gaze tracking method. This method relies on infrared light source so that it needs high-cost equipment. It is not conducive to wide application of eye tracking technology. Convolution neural network based eye tracking is a new research direction, which belongs to appearance-based eye tracking. It has characteristics such as end-to-end, non-calibration, tracking in natural light and suitable for many platforms. This paper investigates and summarizes the appearance-based eye tracking, which is one of the most important eye tracking methods. We introduce one better method of eye tracking to readers: convolution neural network based eye tracking. In the method of convolutional neural network based eye tracking , this paper discusses two key problems: datasets collection and CNN network structure in detail. We promote that ideal gaze dataset should include various head pose and different lighting condition. We also note that CNN model with multiple input have better precision than model with single input.

Key words: eye tracking; gaze; CNN; HCI

0 引言

视线跟踪能够反映人的注意力，认知过程和情绪状态；在医学诊断、心理学研究和人机交互中，都有着广泛的应用。

早期，视线跟踪通常在侵入式条件[1]下进行，需要被试者佩戴电极或者侵入式镜片，例如：搜寻线圈法和眼电图法。侵入式视线追踪因为受到人体分泌物的影响且佩戴不便，逐渐被非侵入式视线追踪所取代。

非侵入条件下，视线跟踪技术可以分为基于特征的视线追踪和基于外观的视线跟踪。

基于特征的视线跟踪是指使用从图像中提取的眼睛轮廓（contour）、眼角位置（eye corners）和角膜反射位置（glint）等生理特征[42-43]，通过视线变化过程中不变的特征和改变的特征之间的相对关系进行注视点估计的方法。

基于特征的视线追踪方法模块间的关系和流程见图 1。首先，基于特征的视线跟踪方法需要进行人脸追踪，检测眼睛的位置，提取眼睛的生理特征。其次，需要确定眼球转动过程中的不变特征，这一步骤称为校准，校准包含屏幕-相机相对位置校准和个性化校准（生理指标）。最后，通过参数校准得到映射关系，根据参数校准的对象不同，分别为二维多项式映射和三维模型映射。根据映射关系，即可通过图像和视频得到眼睛的注视方向。

基于外观的视线追踪是指将眼睛图像视为高维特征，映射到低维特征（注视点）上的方法。这里的特征指的是图像提取

基金项目：国家自然科学基金资助项目（xxxxxxx，xxxxxxxx）；国家“863”计划资助项目（xxxxxxxx）；XXXX 大学教育创新计划资助项目（xxxxxxxx）；军队科研资助项目〔涉密项目不要填写基金项目编号〕

作者简介：李钰卿（1995-），女，山西运城人，硕士研究生，主要研究方向为深度学习，视线跟踪等；战荫伟(1966-) ，男，吉林长春人，教授，博士，主要研究方向为图像处理、人机交互、虚拟现实、增强现实；

特征，如 HOG、LBP。映射方法通常为基于机器学习的方法，例如 kNN、ALR、SVR、随机森林。基于外观的视线跟踪方法包括眼睛图像定位和模型训练。眼睛图像定位包括人脸追踪和眼睛检测，模型训练包括数据集采集和模型的训练。模块间的关系与流程见图 2。

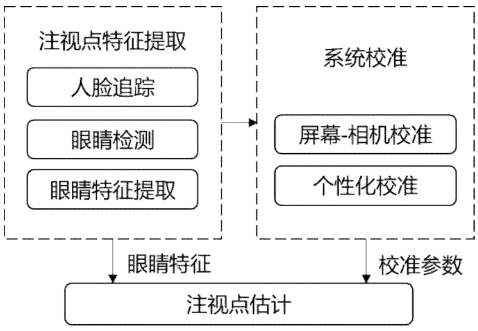

图 1　基于特征的视线跟踪系统

Fig. 1　Feature based eye tracking system

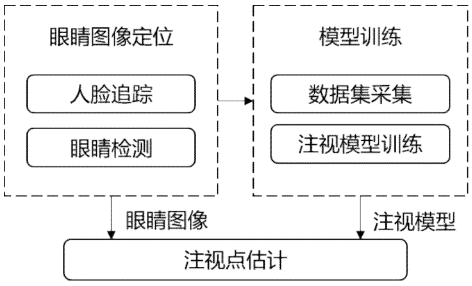

图 2　基于外观的视线跟踪系统

Fig. 2　Appearance based eye tracking system

近年来深度学习也被应用在视线追踪中，且与其他方法在同等条件下相比，精度较高。本文在第一节介绍基于特征的视线跟踪方法；在第一节介绍基于外观的视线跟踪方法；在第三节中，介绍应用于基于外观的方法的凝视数据集；第四节介绍基于卷积神经网络的视线追踪模型。在第五节中，总结基于卷积神经网络的视线追踪的研究趋势。最后，对本文作总结。

1　传统方法

视线追踪传统方法指的是基于特征的视线跟踪方法，使用的特征属于眼睛的生理特征，包括：眼睛轮廓、瞳孔-角膜反射矢量、普尔钦图像等等。

张闯[2]等人使用了瞳孔-角膜反射作为特征的视线追踪方法。红外光照射下眼睛的角膜反射点不变，得到角膜反射点；再利用亮瞳和暗瞳图像的差分图像检测瞳孔，对瞳孔进行椭圆拟合；通过瞳孔-角膜反射矢量进行视线预测。George[56]和 Mohammadi[57]使用了类似特征。这些方法属于二维多项式映射。映射关系 f 可以表示为:

$$f:(X_e,\ Y_e)\ \rightarrow\ (X_s,\ Y_s)$$

(X_e, Y_e)表示实际坐标，(X_s, Y_s)表示屏幕坐标。在[52][53]中描述了(X_e, Y_e) 和(X_s, Y_s)的关系：

$$X_s = a_0 + \sum_{p=1}^{n} * \sum_{i=0}^{p} a_{(i,p)} X_e^{p-i} Y_e^{i}$$

$$Y_s = b_0 + \sum_{p=1}^{n} * \sum_{i=0}^{p} b_{(i,p)} X_e^{p-i} Y_e^{i}$$

其中，n 表示多项式的阶数，a_i 和 b_i 表示系数，a_i 和 b_i 通过校准过程中，注视多个点，标定其视线位置来计算。

Meyer[54]等人使用了三维模型映射的方法估计视线落点。通过图像处理得到人眼参数，包括眼球中心和光轴、角膜半径、视轴和光轴间的偏角、玻璃体折射率、虹膜半径、瞳孔中心到角膜中心的距离、瞳孔半径等[55]，由这些参数建立的模型可估计视线方向。

传统方法的精度高、算法稳定性强。然而与基于外观的方法相比有以下几点不足：

(1) 眼睛特征的提取通常需要红外光源和红外摄像头。在红外光照射下，瞳孔的亮瞳效应和暗瞳效应使眼睛瞳孔的轮廓清晰可辨,见图 3，角膜反射和普尔钦斑（purkinje spot）有利于摄像头捕捉[3]，进而对图像进行处理分析。使用红外光源和红外摄像头的视线追踪方法：提高了眼动仪的成本；红外光容易受到自然环境中其他光线的影响；在户外环境使用可能降低效率；眼睛长时间暴露在红外光下，受到一定损害。

(2) 基于特征的视线跟踪方法通常需要对相机和光线参数、角膜曲率作校准。通过校准获得映射函数或是对眼球进行三维建模，进而通过校准完毕的映射函数与模型计算注视点。基于外观的方法通过训练模型，通过模型预测注视点，通常不需要校准这一步骤[7]。

(3) 基于特征的方法需要摄像机和眼睛的位置相对固定，例如在头戴式设备中的视线追踪[5-6]，头部佩戴支架保证摄像头和眼睛相对位置不变。基于外观的视线跟踪通常使用眼睛图像和学习的方法进行视线跟踪，允许被试者的头部自由运动

2　基于外观的视线追踪

作为红外光的替代。使用自然光的注视估计系统虽然适宜在户外使用，但是自然光下，由于光线反射等原因导致瞳孔捕捉难度大。可见光下的眼睛虹膜图像见图 4。因此，使用基于特征的方法不适用于这种情况。基于外观的视线追踪通常是在自然光条件下进行的，使用普通摄像头，适用于更普遍的视线追踪应用 [44]。

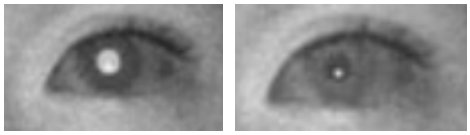

图 3　红外光下的亮瞳和暗瞳图像

Fig. 3　Bright pupil and dark pupil image under IR/NIR light

表 1 基于机器学习的视线跟踪方法

Table 1 Features of learning based eye tracking

分类器	文献	采集设备	平台	单数据集验证精度	跨数据集验证精度	头部运动
kNN	[14]	-	-	9.95°	-	free
	[8]	1 webcam, 1 IR light	head-mounted	-	-	fixed
ALR	[15]	1 camera	desktop	2-5°	-	free
	[16]	1 RGB-D camera and 1 HD camera	desktop	8.1°	-	free
	[9]	-	-	0.62°	-	fixed
	[17]	-	-	0.59°	-	slight
SVR	[18]	1 webcam	desktop	4.06°	-	free
	[19]	-	-	7.1°	8.3°	free
	[20]	cameras	-	5-14°	-	free
	[10]	3 cameras	head-mounted	1-2°	-	fixed
RF	[21]	2 cameras, 2 NIR lights	desktop	2.2cm	-	free
	[11]	8 cameras	desktop	4-6°	-	free
	[22]	front-camera	handheld	1-6cm	-	free
ANN	[23]	4 cameras	head-mounted	1.79°	-	fixed
	[24]	1 camera	desktop	1.5°	-	free
	[25]	1 camera	desktop	1.5°	-	free
	[26]	1 camera, 4 NIR LEDs	head-mounted	0.36°	-	fixed
	[27]	1 camera	desktop	1.5°	-	free

图 4 可见光下的眼睛图像[51]，瞳孔难以识别

Fig. 4 Pupil under natural light is hard to recoginize

在基于外观的方法中，Zhang 等人[8]使用 kNN(k-Nearest Neighbor)和 mRMR 特征进行注视点区域的估计，划分 13 个区域，来预测视线的落点在哪一个区域中。这种方法并非精确的注视点位置预测，且准确率不高。Lu 等人[9]基于 ALR(Adoptive linear regression)对稀疏化的样本进行视线点映射，预测精确的注视点位置，但是精度较低。Matinez 等人[10]提取图像的 HoG 特征，并应用 SVR(Support vector regression)和 RVR(Relevance vector regression)进行视线追踪。Sugano 等人[11]应用 RF(random forests)和基于灰度的特征进行视线追踪。他们的视线追踪方法都达到了不错的精度。

Williams[47]提出一种模型 S^3GP，可以利用稀疏标注的样本，通过高斯过程回归(GPR)进行视线追踪，但要求头部在固定位置上。Lu[48]在高斯过程回归的基础上提出了一种回归模型，可以应用稀疏样本上，但解决了视线追踪中头部需要固定的问题。Williams[47]和 Lu[48]的方法都需要校准过程，增加了系统的复杂度，Sugano[49]提出了不需校准，通过应用特征图和 GPR 的视线跟踪方法，但是精度也随之下降。

Zhang 等人[12]首先使用 CNN 估计注视点，并与其他方法例如 RF, KNN, ALR 和 SVR 进行比较，结果表明 CNN 的性能优于其他方法。Krafka 等人[13]指出深度学习提高视线跟踪性能受到限制的原因在于缺乏大规模数据集。他的研究贡献了大规模数据集，使用 CNN 在移动端估计注视点。在[12]和[13]中，用于注视点估计的图像来源消费级摄像头，且实现了不错的追踪性能，证明了 CNN 可以被用于进行低成本的视线跟踪。

表 1 中展示了不同的学习方法在各个方面的比较。其中，采集设备指得是使用模型预测阶段中采集图像的设备；采集设备由应用 IR/NIR 光源到应用自然光源。平台指的是视线追踪的应用平台，包括头戴式（head-mounted）、桌面式（desktop）和手持式（handheld）,大多数视线追踪实验都在桌面式和头戴式平台进行，头戴式视线追踪设备由于固定在头上，因而精度相对而言更高。验证精度分为两种：第一种，单数据集验证精度是在单一数据集上训练并验证；跨数据集验证精度是在不同的数据集上训练和验证，两种方式都包含角度和距离两种度量，相对而言，跨数据集验证的效果更能证明模型的泛化能力。头部运动表示允许头部的运动范围，包括固定（fixed）、轻微运动（slight）和自由转动（free）；头戴式设备头部与设备间的位置基本都是固定的，而桌面式和手持式设备，由于头部与设备的大多都会存在相对运动，因此大多数为可自由转动。

表 1 中，有些方法通过校准获得了良好的精度，同时也增

表 2 凝视数据集的特性

Table 2 Features of gaze datasets

数据集	分辨率	受试者人数	头部姿态	目标	图像/视频	照明	2D/3D 标注
Columbia [28]	768x480	20	1	16	video	1	2D
Smith et al. [29]	5184x3456	56	5	21	5880	1	3D
EyeDiap [16]	640x480	16	cont	cont	video	2	3D
UT Multiview [30]	1280x1024	50	8+syn	160	64,000	1	3D
MPIIGaze [12]	cont	15	cont	cont	213,659	cont	3D
OMEG [20]	640x480	50	3+cont	10	45,000	1	3D
TabletGaze [22]	1280x1024	51	cont	35	video	cont	2D
GazeCapture [13]	cont	1474	cont	cont	2,445,504	cont	2D
HaopingDeng et al. [31]	cont	200	cont	cont	240,000	cont	3D
ShanghaiTechGaze [32]	cont	137	cont	cont	233,796	cont	2D
RSGD [33]	cont	16	cont	cont	53,180	1	2D

加了系统的复杂度。目前，基于外观的视线跟踪的目标是建立无需校准的，可允许头部自由运动的并且具有良好的精度的系统。

3 凝视数据集

凝视数据集的采集过程为：受试者的头部处于固定的位置；眼睛注视固定目标[28]；布置一个或多个角度不同的相机采集受试者眼睛图像。这样采集到的图像不满足多样化，使模型估计精度收到限制。随后，采集凝视数据集时，受试者允许头部自由运动，视线跟随运动的目标 [16]采集图像。理想的眼动仪应该具有追踪精度高、追踪性能稳定且头部的运动不受限制等特性。这些目标与数据集的特性密不可分。表 2 对一些凝视数据集的基本特性进行了对比。表 2 中，分辨率代表相机分辨率；受试者人数以数字 *n* 表示；头部姿态中，数字 *n* 代表 *n* 个固定位置，cont.代表头部可自由运动；目标中，数字 *n* 代表眼睛注视 *n* 个固定位置的目标，cont.代表连续运动的目标；照明中，数字 *n* 代表 *n* 个固定位置的光源，cont.代表连续变化的光源。

3.1 凝视数据集特性

(1) 头部姿态。理想的视线追踪允许头部自由的运动。在一些凝视数据集的收集程序中，头部被固定在有限数量的位置，使用头部固定在某些位置的方式采集的数据集训练模型会限制卷积神经网络模型的预测性能。Deng 等人[31]指出，凝视数据集头部运动的姿态范围影响着模型预测注视点时头部的运动范围。在采集数据集时，增大头部运动的范围能够提升平均估计精度。头部运动的范围指的是，头部与摄像机的水平距离以及运动占有的摄像机拍摄的视野范围。

(2) 光照变化。理想的视线追踪也应该能够应对自然环境中光线持续不断的变化。照明差异会影响 CNN 的预测性能，Zhang 等人[12]证明了数据集收集连续的光照变化（光照角度、光照强度）可以提升 CNN 预测的性能。在收集数据集时，模拟自然环境光照才能提供给模型应对光照的能力。因此在数据集收集过程中提供不同角度和光强组合的光照是必要的。

(3) 眼睛分辨率。眼睛图像分辨率的大小可以用眼睛图像内眼角和外眼角之间的像素距离（像素数）来表示。收集图像时，眼睛分辨率不仅与摄像机本身相关，同时和人与摄像机间的水平距离相关。Tamura 等人[33]的研究表明眼睛图像的分辨率影响着模型预测注视点的精度和鲁棒性。Lemley 等人 [34]使用相同的卷积神经网络结构，输入不同分辨率的眼睛图像，眼睛图像的分辨率越高，预测误差越小，因而眼睛分辨率在使用卷积神经网络进行视线追踪时具有影响。

(4) 数据集的规模。数据集规模大小影响视线追踪性能。这是因为个体之间存在差异。消除数据集的个体差异通常需要包含多个受试者的大规模数据集。个体差异包括肤色和面部装饰，眼睛等遮挡元素。Krafka[13]指出：（1）保持受试者数量不变，增加每个受试者的样本数量；（2）增加受试者数量，同时保持每个受试者的样本数量不变；前者的预测精度比后者更低。

(5)其他因素。卷积神经网络模型的跨平台适用也是视线追踪系统追求的目标之一。多个研究表明使用相同的数据集进行训练的卷积神经网络模型在不同的平台上预测性能具有显著的差异。为了消除不同设备平台间的预测差异，跨平台收集数据也是必要的。视线追踪数据集标注通常分为两种，三维标注和二维标注。三维标注使用旋转角度标注，二维标注使用平面距离标注。Deng 等人[31]表明 3D 标注的数据集比 2D 标注的数据集更能够减少这种差异，但是相比 3D 标注，使用 2D 标注的数据集在一些特定网络结构上的估计精度更高。

3.2 凝视数据集的收集

大多数情况下，采集凝视数据集时，征集受试者拍摄图像或视频。然而图像标注是一项耗时的工作。为了避免代价昂贵的标注工作，Wood[14]提出了 SynthesEyes，这是使用合成的图像建立数据集的方法，合成的图像均为已标注好的并且具有多样性。Wood 等人[35]提出了 UnityEyes，这是一种基于 SynthesEyes，快速生成标注好的眼睛图像的渲染框架。Shrivastava [36]提出了 SimGAN，这是一种生成对抗网络，输入合成图像，生成高度逼真的图像，用 SimGAN 生成的凝视数据

表 3 基于卷积神经网络模型的视线追踪方法对比

Table 3 CNN based eye tracking methods

方法	模型来源	通道数	数据集	精度	运行效率
Multimodal CNNs [12]	lenet	1	MPIIGaze	1-6deg	-
Spatial weights CNN [37]	alexnet	1	MPIIGaze	4.8deg	-
CNNs with gaze transform layer [31]	alexnet	2	HaopingDeng et al.	5.6deg	1000fps
Multi-Device CNN [39]	alexnet	1	MPIIGaze	5.2deg	-
iTracker [13]	alexnet	4	GazeCapture	2.58deg	50ms
Multiview MTL [32]	resnet	3	ShanghaiTechGaze	4.61deg	21.34ms
$G_{NET}+A_{TN}+A_{TNL}+A_{TNR}$ [40]	-	4	Tablet Gaze	2.08deg	-

集可以进行注视点估计训练。

4 基于卷积神经网络的视线跟踪

使用 CNN 进行注视估计的方法有两类：一类是估计头部旋转矢量和眼球旋转矢量，使用 3D 标注的数据集，计算视线方向与屏幕平面求交得到注视点在屏幕上的坐标，这种方法被称为 3D 注视点估计。另一类是估计屏幕平面上的注视点坐标，使用 2D 标注的数据集，这种方法被称为 2D 注视点估计。本节将根据不同的卷积神经网络的模型架构对注视点估计的方法作介绍分析。

4.1 3D 注视估计

3D 注视估计方法使用卷积神经网络模型预测头部旋转矢量和眼球旋转矢量，并将两者映射在摄像机坐标系中，如图 4 所示。这是对视线方向的估计。视线方向与 3D 空间中物体的交点既是注视点。这种方法适用于 3D 空间中的注视点估计。

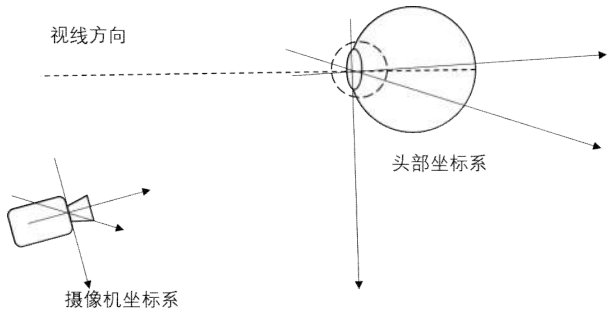

图 4 3D 注视点估计中的摄像机和头部坐标系关系

Fig. 4 Camera coordinates in 3D gaze estimation

4.1.1 Multimodal CNN[50]

多模态卷积神经网络来源于 Lenet 网络架构。该模型采用 SURF 检测人脸标志点。根据标志点得出眼睛图像和头部旋转矢量。将人眼图像和头部姿态归一化到极坐标空间，再将眼睛图像和头部姿态矢量输入 CNN 模型。在[38]中表明了这种对预测数据进行规范化的重要性。将头部姿态矢量被输入到连接层。CNN 网络用于学习从输入的头部旋转矢量特征和裁剪的眼睛图像到注视角度的映射。这种模型利用 SURF 估计头部姿态，使用 CNN 提取眼睛特征，最终在连接层预测视线方向。

4.1.2 Spatial weights CNN[37]

空间权重卷积神经网络模型的特点是在最后一个卷积层前增加了空间权重层用以激活。空间权重指的是三个卷积层，其

中具有 1×1 的滤波器用于线性单元校正。添加空间权重抑制对注视点预测无贡献的图像区域的激活，并增强其他区域的激活。CNN 的任务是学习从单幅面部图像提取特征映射到注视向量。头部旋转矢量在网络中被隐式地检测。优点是无需额外计算头部旋转矢量，缺点是会降低网络的预测精度。

4.1.3 CNNs with gaze transform layer[31]

迁移学习神经网络结构的特点是通过两个 CNN 模型分别估计头部旋转矢量和眼球旋转矢量，再使用连接层将头部旋转矢量和眼球旋转矢量聚合，映射到注视矢量。这种 CNN 模型的优点是能够降低头部与注视点过拟合的风险。这种方法还允许使用其他的头部姿态数据集对网络进行预训练。在数据不足的情况下使用其他数据集增加数据量。

4.1.4 Multi-Device CNN[39]

多设备卷积神经网络来源于 Alexnet 网络架构。网络结构包含编码器、解码器和特征提取层。编码器用于在不同的设备收集的数据集上进行初级特征提取。特征提取层在编码器和解码器之间共享，提取卷积特征。解码器连接来自特征提取层的特征。网络的输入是单幅图像，CNN 模型用于学习从不同设备提取的面部图像到注视向量的映射。这种模型的优点是使用共享的特征提取层，可以使网络架构具有在不同设备上的兼容性，并且便于网络结构扩展到深层网络。

4.2 2D 注视估计

2D 注视估计方法是在以摄像头为原点的屏幕平面坐标系中预测注视点。相比与 3D 注视估计方法，2D 注视点估计的范围被限制在屏幕空间中,见图 5。CNN 模型预测屏幕空间上的注视点坐标。

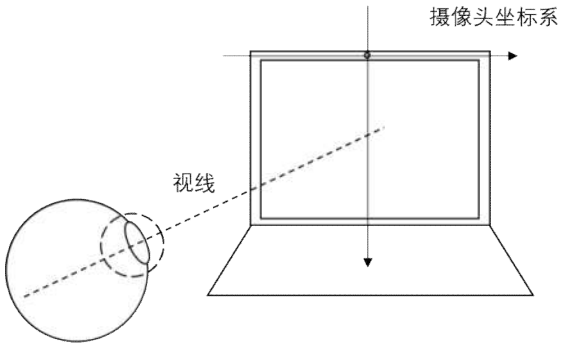

图 5 2D 注视点估计中的摄像机坐标系和头部位置的关系

Fig. 5 Camera coordinates in 2D gaze estimation

4.2.1 iTracker[13]

iTracker 模型来源于 Alexnet，是一种多通道的网络结构。模型的输入为四个部分：裁剪后的面部图像、人脸在整个图像中的位置（面部网格，以二进制掩模表示）和裁剪后的左右眼睛图像。模型用于学习从输入的一组图像到屏幕平面二维注视点的映射。研究[12]中表明，删除模型输入的四个部分中的一个，预测性能都明显下降。多个图像输入使 CNN 模型能够识别眼睛的细微变化，从而提高注视预测的精度。iTracker 是一种端到端的视线跟踪网络结构，不需要对头部姿态的额外估计。

4.2.2 Multiview MTL[32]

Multiview MTL 是多个模块组合的框架，包括注视方向预测模块和注视点坐标预测模块。注视方向预测模块的任务是使用 CNN 模型预测眼睛的三维注视方向。注视点坐标预测的任务是对视线凝视点坐标进行预测。注视点坐标预测模块的输入是由三台摄像机拍摄的眼睛图像，网络使用了 ResNet，然后由单视图特征融合网络（SVFFN）和交叉视图特征融合网络（CVFFN）学习从 ResNet 的输出到注视点坐标的映射。从不同视角的相机拍摄的图像组合中定位眼睛位置并进行视线估计。这种框架的优点在于能够减少子网络中的参数数量，而且还减少了眼镜反射对视线跟踪的影响。

4.2.3 G_{NET} +A_{TN} +A_{TNL} +A_{TNR}[40]

这种网络结构包括四个通道：一个使用 CNN 模型提取面部几何信息的通道（G_{NET}）和三个分别提取左眼（A_{TNL}）、右眼（A_{TNR}）和双眼图像（A_{TN}）的外观信息的通道。几何信息是指检测面部特征点得到的头部位姿。外观信息是指通过 CNN 提取的特征。通过连接层连接四个通道输出，协同地估计二维注视点。在文[40]中，眼睛的外观信息与面部的几何特征相结合，使模型预测达到了最佳的预测性能。而单独使用四个通道中的其中一个来估计注视点都会导致性能下降，表明了几何信息和外观信息的组合在注视点预测中更有效，多输入的网络结构优于单输入的网络结构。

4.3 其他基于 CNN 的注视估计

基于 CNN 的视线追踪系统应用平台不仅包含桌面式和移动手持式，还包含在智能汽车上的应用。与桌面式和移动手持式不同的是，智能汽车的视线追踪通常用于检测疲劳驾驶；这一需求不要求视线追踪系统精确跟踪人的注视点而是注视的区域，并且对注视区域预测的精确率极高。Vora[45]和 Rizwan[46]等人为汽车驾驶舱内的视野划分区域，并通过 CNN 输入司机面部图像来预测注视区域，精确率均在 90%以上。此外，预测注视区域还可用于辅助司机与驾驶系统的交互，在智能汽车未来的发展路上是重要的组成部分。

5 基于卷积神经网络的视线跟踪研究趋势

结合对凝视数据集和基于卷积神经网络的注视跟踪方法的总结分析，我们对未来的视线跟踪研究趋势做出一些预测。

5.1 凝视数据集

随着基于外观的视线追踪方法的发展，这一领域会出现更多数据集。首先，可能出现更大规模的数据集，规模增大包含两个方面：其一是受试者的数量。大部分数据集受试者的数量在 100 名以下。增大研究对象的数量，在数据集中包含更多的外观、头部姿态和照明的组合，能够促进卷积神经网络在视线跟踪领域更广泛地使用。其二是不同的数据采集设备平台，在不同的平台上采集数据，训练模型，能够使基于卷积神经网络的视线追踪应用在不同的平台上预测注视点。

5.2 高性能视线追踪

提高预测性能有三个方面：首先，采用外观信息与几何信息结合作为 CNN 模型的输入。这种方式需要额外的计算几何信息，但是可以提高预测的精度。其次，多输入的卷积神经网络模型比单输入的卷积神经网络模型更能提高视线追踪的精度。但是其中也要注意头部姿态和眼睛方向估计的过拟合问题。最后，使用更深层的网络的模型可以展示出更好的性能，未来可能使用更深层网络进行注视预测，但同时，深层网络会带来计算耗时的增加。

5.3 更广泛的应用：

基于特征的视线跟踪技术已应用于台式机、头戴式设备、手持设备[58-59]、汽车[60]和智能电视平台。然而由于高成本的红外光源和头部固定器，使得视线跟踪技术无法成为普及技术。栗战恒[41]研究中提出移动设备与固定设备在硬件构成、使用环境、操作方式和用户心理间存在差异。基于外观的视线跟踪方法中，基于卷积神经网络的方法对硬件成本要求低、可以在户外使用、对移动设备的适用性好，推动了视线追踪的广泛应用。虽然卷积是一种高耗的计算，但是随着计算硬件成本的降低，基于卷积神经网络的视线追踪能够带来预测精度的提高。

6 结语

本文对基于卷积神经网络的凝视数据集和网络体系结构的特点进行了总结。展示了收集凝视数据集应该满足的需求，以及基于卷积神经网络的视线跟踪方法模型架构。凝视数据集的收集需要考虑头部姿态、光照、眼睛分辨率、标注、数据多样性等因素。基于卷积神经网络的视线跟踪采用外观信息与多输入结构可以提高追踪精度。此外，卷积神经网络模型有潜力

扩展到更深层的网络架构，以提高眼动仪系统的预测性能。在人机交互方面，视线跟踪有着巨大的潜力，基于卷积神经网络的眼球跟踪对于人机交互方式的改变、人机交互的体验提升都有巨大的作用。

7 参考文献

[1] Eye tracking methodology. New York, NY: Springer Berlin Heidelberg, 2017.

[2] ZHANG C, CHI J-N, ZHANG Z-H 等. A Novel Eye Gaze Tracking Technique Based on Pupil Center Cornea Reflection Technique: A Novel Eye Gaze Tracking Technique Based on Pupil Center Cornea Reflection Technique[J]. Chinese Journal of Computers, 2010, 33(7): 1272–1285.

[3] ZHANG W, LI B, DENG H 等. Distorted Pupil Localization in Eye Tracking: Distorted Pupil Localization in Eye Tracking[J]. JOURNAL OF ELECTRONICS INFORMATION & TECHNOLOGY, 2010, 32(2): 416–421.

[4] ZHANG C, CHI J-N, ZHANG Z-H 等. The Research on Eye Tracking for Gaze Tracking System: The Research on Eye Tracking for Gaze Tracking System[J]. Acta Automatica Sinica, 2010, 36(8): 1051–1061.

[5] 潘世豪,赵新灿,王雅萍,高明磊.人体自由运动状态下的视线追踪算法研究[J].计算机应用研究,2017,34(03):911-914.

[6] 王林,李斌.头部可自由运动的头戴式视线跟踪系统设计[J].计算机应用与软件,2015,32(07):163-166.

[7] HANSEN D W, QIANG JI. In the Eye of the Beholder: A Survey of Models for Eyes and Gaze[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(3): 478–500.

[8] Zhang Y., Bulling A., and Gellersen. H., “Discrimination of gaze directions using low-level eye image features,” in Proceedings of the 1st international workshop on Pervasive eye tracking & mobile eye-based interaction - PETMEI ’11, Beijing, China, 2011, p. 9.

[9] Lu F., Sugano Y., Okabe T., and Sato Y., “Inferring human gaze from appearance via adaptive linear regression,” in 2011 International Conference on Computer Vision, Barcelona, Spain, 2011, pp. 153–160.

[10] Martinez F., Carbone A., and Pissaloux, E. “Gaze estimation using local features and non-linear regression,” in 2012 19th IEEE International Conference on Image Processing, Orlando, FL, USA, 2012, pp. 1961–1964.

[11] SUGANO Y, MATSUSHITA Y, SATO Y. Learning-by-Synthesis for Appearance-Based 3D Gaze Estimation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 1821–1828.

[12] ZHANG X, SUGANO Y, FRITZ M 等. Appearance-based gaze estimation in the wild[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015: 4511–4520.

[13] KRAFKA K, KHOSLA A, KELLNHOFER P 等. Eye Tracking for Everyone[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 2176–2184.

[14] WOOD E, BALTRUŠAITIS T, MORENCY L-P 等. Learning an appearance-based gaze estimator from one million synthesised images[C]//Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications - ETRA ’16. Charleston, South Carolina: ACM Press, 2016: 131–138.

[15] Lai C.-C., Chen Y.-T., Chen K.-W., Chen S.-C., 等, “Appearance-Based Gaze Tracking with Free Head Movement,” in 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 2014, pp. 1869–1873.

[16] MORA K A F, MONAY F, ODOBEZ J-M. EYEDIAP: a database for the development and evaluation of gaze estimation algorithms from RGB and RGB-D cameras[C]//Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA ’14. Safety Harbor, Florida: ACM Press, 2014: 255–258.

[17] Lu F., Sugano Y., Okabe T., 等 “Adaptive Linear Regression for Appearance-Based Gaze Estimation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 10, pp. 2033–2046, Oct. 2014.

[18] Huang M. X., Kwok,T. C. K. Ngai G. 等“Building a Self-Learning Eye Gaze Model from User Interaction Data,” in Proceedings of the ACM International Conference on Multimedia - MM ’14, Orlando, Florida, USA, 2014, pp. 1017–1020.

[19] Park S., Zhang X., Bulling A.等 “Learning to find eye region landmarks for remote gaze estimation in unconstrained settings,” in Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications - ETRA ’18, Warsaw, Poland, 2018, pp. 1–10.

[20] HE Q, HONG X, CHAI X 等. OMEG: Oulu Multi-Pose Eye Gaze Dataset[G]//PAULSEN R R, PEDERSEN K S. Image Analysis. Cham: Springer International Publishing, 2015, 9127: 418–427.

[21] Luong D., Kang J., Nguyen 等 “Focus Assessment Method of Gaze Tracking Camera Based on ε-Support Vector Regression,” Symmetry, vol. 9, no. 6, p. 86, Jun. 2017.

[22] HUANG Q, VEERARAGHAVAN A, SABHARWAL A. TabletGaze: Unconstrained Appearance-based Gaze Estimation in Mobile Tablets[J]. arXiv:1508.01244 [cs], 2015.

[23] Tonsen M., Steil J., Sugano Y., 等 “Invisible Eye: Mobile Eye Tracking Using Multiple Low-Resolution Cameras and Learning Based Gaze Estimation,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous T echnologies,vol.1,no.3,pp.1–21,Sep.2017.

[24] Sugano Y., Matsushita Y., Sato Y. 等.“An Incremental Learning Method for Unconstrained Gaze Estimation,” in Computer Vision – ECCV2008, vol.5304, D.Forsyth ,P.Torr, and A.Zisserman,Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 656–667.

[25] Xu L.-Q., Machin D., and Sheppard P., “A Novel Approach to Real-time Non-intrusive Gaze Finding,”in Procedings of the British Machine Vision Conference

1998, Southampton, 1998, pp. 43.1-43.10.
[26] Wang J., Zhang G., and Shi J., “2D Gaze Estimation Based on Pupil-Glint Vector Using an Artificial Neural Network,” Applied Sciences, vol. 6, no. 6, p. 174, Jun. 2016.
[27] Baluja, Shumeet, and D.Pomerleau." Non-Intrusive Gaze Tracking Using Artificial Neural Networks." Advances in Neural Information Processing Systems 6, [7th NIPS Conference, Denver, Colorado, USA, 1993] 1993.
[28] MCMURROUGH C D, METSIS V, RICH J 等. An eye tracking dataset for point of gaze detection[C]//Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA ′12. Santa Barbara, California: ACM Press, 2012: 305.
[29] SMITH B A, YIN Q, FEINER S K 等. Gaze locking: passive eye contact detection for human-object interaction[C]//Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST ′ 13. St. Andrews, Scotland, United Kingdom: ACM Press, 2013: 271–280.
[30] SESMA L, VILLANUEVA A, CABEZA R. Evaluation of pupil center-eye corner vector for gaze estimation using a web cam[C]//Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA ’12. Santa Barbara, California: ACM Press, 2012: 217.
[31] DENG H, ZHU W. Monocular Free-Head 3D Gaze Tracking with Deep Learning and Geometry Constraints[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 3162–3171.
[32] LIAN D, HU L, LUO W 等. Multiview Multitask Gaze Estimation With Deep Convolutional Neural Networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018: 1−14.
[33] TAMURA K, CHOI R, AOKI Y. Unconstrained and Calibration-Free Gaze Estimation in a Room-Scale Area Using a Monocular Camera[J]. IEEE Access, 2018, 6: 10896–10908.
[34] LEMLEY J, KAR A, DRIMBAREAN A 等. Efficient CNN Implementation for Eye-Gaze Estimation on Low-Power/Low-Quality Consumer Imaging Systems[J]. arXiv:1806.10890 [cs], 2018.
[35] WOOD E, BALTRUAITIS T, ZHANG X 等. Rendering of Eyes for Eye-Shape Registration and Gaze Estimation[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015: 3756−3764.
[36] SHRIVASTAVA A, PFISTER T, TUZEL O 等. Learning from Simulated and Unsupervised Images through Adversarial Training[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, 2017: 2242−2251.
[37] ZHANG X, SUGANO Y, FRITZ M 等. It′s Written All Over Your Face: Full-Face Appearance-Based Gaze Estimation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, HI, USA: IEEE, 2017: 2299−2308.
[38] ZHANG X, SUGANO Y, BULLING A. Revisiting data normalization for appearance-based gaze estimation[C]//Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications - ETRA ’18. Warsaw, Poland: ACM Press, 2018: 1–9.
[39] ZHANG X, HUANG M X, SUGANO Y 等. Training Person-Specific Gaze Estimators from User Interactions with Multiple Devices[C]//Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ′ 18. Montreal QC, Canada: ACM Press, 2018: 1 −12.
[40] JYOTI S, DHALL A. Automatic Eye Gaze Estimation using Geometric & Texture-based Networks[C]//2018 24th International Conference on Pattern Recognition (ICPR). Beijing, China: IEEE, 2018: 2474–2479.
[41] 栗战恒,郑秀娟,刘凯.移动设备视线跟踪技术研究进展[J].计算机工程与应用,2018,54(24):6-11+148.
[42] WANG H, PAN C, CHAILLOU C. Tracking Eye Gaze under Coordinated Head Rotations with an Ordinary Camera[G]//ZHA H, TANIGUCHI R, MAYBANK S. Computer Vision – ACCV 2009. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, 5995: 120–129.
[43] CHEN B-C, WU P-C, CHIEN S-Y. Real-time eye localization, blink detection, and gaze estimation system without infrared illumination[C]//2015 IEEE International Conference on Image Processing (ICIP). Quebec City, QC, Canada: IEEE, 2015: 715–719.
[44] KAR A, CORCORAN P. A Review and Analysis of Eye-Gaze Estimation Systems, Algorithms and Performance Evaluation Methods in Consumer Platforms[J]. IEEE Access, 2017, 5: 16495–16519.
[45] Vora S., Rangesh A., and Trivedi M. M., “Driver Gaze Zone Estimation Using Convolutional Neural Networks: A General Framework and Ablative Analysis,” IEEE Transactions on Intelligent Vehicles,vol.3,no.3,pp. 254–265, Sep. 2018.
[46] Rizwan Naqvi, Muhammad Arsalan, Ganbayar Batchuluun 等，“Deep Learning-Based Gaze Detection System for Automobile Drivers Using a NIR Camera Sensor,” Sensors, vol. 18, no. 2, p. 456, Feb. 2018.
[47] Williams O., Blake A., and Cipolla R., “Sparse and Semi-supervised Visual Mapping with the S3GP,” in 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR’06), New York, NY, USA, 2006, vol. 1, pp. 230–237.
[48] Lu F., Okabe T., Sugano Y., 等 “A Head Pose-free Approach for Appearance-based Gaze Estimation,” in Procedings of the British Machine Vision Conference 2011, Dundee, 2011, pp. 126.1-126.11.
[49] Y. Sugano, Y. Matsushita, and Y. Sato, “Calibration-free gaze sensing using saliency maps,” in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 2010, pp. 2667–2674.
[50] Zhang X., Sugano Y., Fritz M., 等 “MPIIGaze: Real-World Dataset and Deep Appearance-Based Gaze Estimation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 1, pp. 162–175, Jan. 2019.
[51] Chen B.-C., Wu P.-C., and Chien S.-Y., “Real-time eye localization, blink detection, and gaze estimation system without infrared illumination,” in 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 2015, pp. 715–719.
[52] P. Blignaut, ``Mapping the pupil-glint vector to gaze

coordinates in a simple video-based eye tracker," J. Eye Movement Res., vol. 7, no. 1, pp. 1-11, 2013.

[53] Cherif Z. R., Nait-Ali,A. Motsch J. F.``An adaptive calibration of an infrared light device used for gaze tracking," in Proc. 19th IEEE Instrum. Meas. Technol. Conf., vol. 2. May 2002, pp. 1029-1033.

[54] Meyer A., Böhme M., Martinetz T., ``A single-camera remote eye tracker," Perception and Interactive Technologies (Lecture Notes in Computer Science), vol. 4021. New York, NY, USA: Springer, 2006, pp. 208-211.

[55] 周小龙, 汤帆扬, 管秋, 等. 基于 3D 人眼模型的视线跟踪技术综述[J]. 计算机辅助设计与图形学学报, 2017(9).

[56] George A. and Routray A., “Fast and accurate algorithm for eye localization for gaze tracking in low-resolution images,” IET Computer Vision, vol. 10, no. 7, pp. 660–669, Oct. 2016.

[57] Mohammadi M. R. and Raie,A. “Selection of unique gaze direction based on pupil position,” IET Computer Vision, vol. 7, no. 4, pp. 238–245, Aug. 2013.

[58] Biedert R., Dengel A., Buscher G.等``Reading and estimating gaze on smart phones," in Proc. Symp. Eye Tracking Res. Appl., New York, NY, USA, 2012, pp. 385_388.

[59] Li Z., Sun G., Zhang F,等 ``Smartphone based fatigue detection system using progressive locating method," IET Intell. Transp. Syst., vol. 10, no. 3, pp. 148_156, 2016.

[60] D. Kern, A. Mahr, S. Castronovo, A. Schmidt, and C. Müller, ``Making use of drivers' glances onto the screen for explicit gaze-based interaction," in Proc. 2nd Int. Conf. Autom. User Interfaces Interact. Veh. Appl., 2010, p. 110.