

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 9 月 24 日 (24.09.2020)



(10) 国际公布号
WO 2020/186883 A1

(51) 国际专利分类号:
G06K 9/00 (2006.01)

(21) 国际申请号: PCT/CN2019/129893

(22) 国际申请日: 2019 年 12 月 30 日 (30.12.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910204566.9 2019年3月18日 (18.03.2019) CN

(71) 申请人: 北京市商汤科技开发有限公司(BEIJING SENSETIME TECHNOLOGY DEVELOPMENT CO., LTD.) [CN/CN]; 中国北京市海淀区北四环西路58号11层1101-1117室, Beijing 100080 (CN)。

(72) 发明人: 黄诗尧(HUANG, Shiyao); 中国北京市海淀区北四环西路58号11层1101-1117室, Beijing 100080 (CN)。 王飞(WANG, Fei); 中国北京市海淀区北四环西路58号11层1101-1117室, Beijing 100080 (CN)。 钱晨(QIAN, Chen); 中

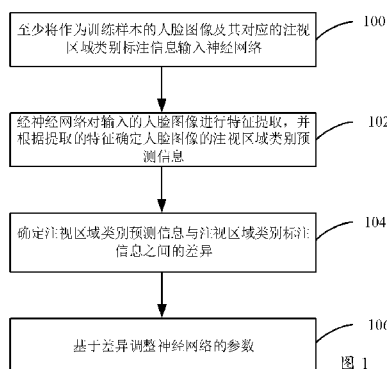
国北京市海淀区北四环西路58号11层1101-1117室, Beijing 100080 (CN)。

(74) 代理人: 北京博思佳知识产权代理有限公司(BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区上地三街9号嘉华大厦B座409, Beijing 100085 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(54) Title: METHODS, DEVICES AND APPARATUSES FOR GAZE AREA DETECTION AND NEURAL NETWORK TRAINING

(54) 发明名称: 注视区域检测及神经网络训练的方法、装置和设备



100 At least input into a neural network a face image used as a training sample and corresponding gaze area category labeling information
102 Extract, by means of the neural network, features of the input face image, and determine gaze area category prediction information of the face image according to the extracted features
104 Determine differences between the gaze area category prediction information and the gaze area category labeling information
106 Adjust parameters of the neural network based on the differences

图 1

(57) Abstract: Embodiments of the present disclosure provide methods, devices and apparatuses for gaze area detection and neural network training. A method for neural network training for use in gaze area detection comprises: at least inputting into a neural network a face image used as a training sample and corresponding gaze area category labeling information, the labeled gaze area category being one of multiple types of defined gaze areas obtained by dividing a specified spatial area in advance; extracting, by means of the neural network, features of the input face image, and determining gaze area category prediction information of the face image according to the extracted features; determining differences between the gaze area category prediction information and the gaze area category labeling information; and adjusting parameters of the neural network based on the differences.

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

(57) 摘要: 本公开的实施例提供一种注视区域检测及神经网络训练的方法、装置和设备。该注视区域检测用神经网络的训练方法包括: 至少将作为训练样本的人脸图像及其对应的注视区域类别标注信息输入所述神经网络, 其中, 标注的注视区域类别属于预先对指定空间区域划分得到的多类定义注视区域之一; 经所述神经网络对输入的所述人脸图像进行特征提取, 并根据提取的特征确定所述人脸图像的注视区域类别预测信息; 确定所述注视区域类别预测信息与所述注视区域类别标注信息之间的差异; 基于所述差异调整所述神经网络的参数。

注视区域检测及神经网络训练的方法、装置和设备

相关申请的交叉引用

[01] 本公开要求于 2019 年 3 月 18 日提交的、申请号为 201910204566.9、发明名称为“注视区域检测及神经网络的训练方法、装置和设备”的中国专利申请的优先权，该中国专利申请公开的全部内容以引用的方式并入本文中。

技术领域

[02] 本公开涉及计算机视觉技术，特别涉及一种注视区域检测及神经网络训练的方法、装置和设备。

背景技术

[03] 随着人工智能和汽车行业的迅速发展，一些人工智能技术已应用到量产车辆上。受关注的一种人工智能产品用于监控驾驶员的驾驶状态，例如，驾驶员在驾驶时是否分心，以在监控到驾驶员分心时及时对驾驶员进行提醒，降低事故风险。

发明内容

[04] 本公开的第一方面提供一种注视区域检测用神经网络的训练方法，所述方法包括：至少将作为训练样本的人脸图像及其对应的注视区域类别标注信息输入所述神经网络，其中，标注的注视区域类别属于预先对指定空间区域划分得到的多类定义注视区域之一；经所述神经网络对输入的所述人脸图像进行特征提取，并根据提取的特征确定所述人脸图像的注视区域类别预测信息；确定所述注视区域类别预测信息与所述注视区域类别标注信息之间的差异；基于所述差异调整所述神经网络的参数。

[05] 本公开的第二方面提供一种注视区域检测方法，所述方法包括：截取在指定空间区域内采集到的图像中的人脸区域，得到人脸图像；将所述人脸图像输入神经网络，其中，所述神经网络预先采用包括多个人脸图像样本及其分别对应的注视区域类别标注信息的训练样本集训练完成，标注的注视区域类别属于预先对所述指定空间区域划分得到的多类定义注视区域之一；经所述神经网络对输入的所述人脸图像进行特征提取，并根据提取的特征确定所述人脸图像对应的注视区域检测类别。

[06] 本公开的第三方面提供一种注视区域检测用神经网络的训练装置，所述装置包括：样本输入模块，用于至少将作为训练样本的人脸图像及其对应的注视区域类别标注信息输入所述神经网络，其中，标注的注视区域类别属于预先对指定空间区域划分得到的多类定义注视区域之一；类别预测模块，用于经所述神经网络对输入的所述人脸图像进行特征提取，并根据提取的特征确定所述人脸图像的注视区域类别预测信息；差异确定模块，用于确定所述注视区域类别预测信息与所述注视区域类别标注信息之间的差异；参数调整模块，用于基于所述差异调整所述神经网络的参数。

[07] 本公开的第四方面提供一种注视区域检测装置，所述装置包括：图像获取模块，用于截取在指定空间区域内采集到的图像中的人脸区域，得到人脸图像；图像输入模块，用于将所述人脸图像输入神经网络，其中，所述神经网络预先采用包括多个人脸图像样本及其分别对应的注视区域类别标注信息的训练样本集训练完成，标注的注视区域类别属于预先对所述指定空间区域划分得到的多类定义注视区域之一；类别检测模块，用于经所述神经网络对输入的所述人脸图像进行特征提取，并根据提取的特征确定所述人脸图像对应的注视区域检测类别。

[08] 本公开的第五方面提供一种注视区域检测用神经网络的训练设备，所述设备包括存储器和处理器，其中，所述存储器中存储可由该处理器执行的计算机指令，所述处理器在执行所述计算机指令时，实现根据本公开的第一方面的注视区域检测用神经网络的训练方法。

[09] 本公开的第六方面提供一种注视区域检测设备，所述设备包括存储器和处理器，其中，所述存储器

中存储有可由该处理器执行的计算机指令，所述处理器在执行所述计算机指令时，实现根据本公开的第二方面的注视区域检测方法。

[10] 本公开的第七方面提供一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时，使该处理器实现根据本公开的第一方面的注视区域检测用神经网络的训练方法，和/或，使该处理器实现根据本公开的第二方面的注视区域检测方法。

[11] 根据本公开的实施例，通过利用作为训练样本的人脸图像及其对应的注视区域类别标注信息训练神经网络，使得能够根据该神经网络直接预测人脸图像对应的注视区域。

附图说明

[12] 图 1 为根据本公开的实施例的一种注视区域检测用神经网络的训练方法的流程图；

[13] 图 2 为根据本公开的实施例的在车辆驾驶员注意力监控场景下预先定义的多个注视区域的示意图；

[14] 图 3 例示了本公开的实施例可以应用的一种神经网络结构的示例；

[15] 图 4 例示了根据本公开的实施例的用于训练神经网络的配置；

[16] 图 5 例示了根据本公开的另一实施例的用于训练神经网络的配置；

[17] 图 6 为与图 5 中的配置对应的神经网络训练方法的流程图；

[18] 图 7 为根据本公开的实施例的获得眼睛图像的示意图；

[19] 图 8 为根据本公开的另一实施例的神经网络训练方法的流程图；

[20] 图 9 例示了与图 8 所示的神经网络训练方法对应的配置；

[21] 图 10 为根据本公开的实施例的一种注视区域检测方法的流程图；

[22] 图 11 为根据本公开的实施例的一种神经网络应用场景的示意图；

[23] 图 12 例示了图 11 所示的应用场景中神经网络输出的注视区域检测类别的示例；

[24] 图 13 为根据本公开的实施例的一种注视区域检测用神经网络的训练装置的框图；

[25] 图 14 为根据本公开的实施例的一种注视区域检测装置的框图；

[26] 图 15 为根据本公开的另一实施例的一种注视区域检测装置的框图；

[27] 图 16 为根据本公开的实施例的一种注视区域检测用神经网络的训练设备的框图；

[28] 图 17 为根据本公开的实施例的一种注视区域检测设备的框图。

具体实施方式

[29] 为了使本技术领域的人员更好地理解本公开，下面将结合附图对本公开的实施例进行详细描述。显然，所描述的实施例仅仅是本公开的一些实施例，而不是全部的实施例。基于本公开，本领域普通技术人员在没有作出创造性劳动的情况下所获得的所有其他实施例，都应当属于本公开保护的范围。

[30] 在本公开使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本公开。在本公开中所使用的单数形式诸如“一种”、“所述”、“该”等也旨在包括复数形式，除非上下文清楚地表示其他含义。还应当理解，本文中使用的术语“和/或”是指并包含一个或多个相关联的列出项目中的任何一个或其所有可能组合。

[31] 应当理解，尽管本公开可能采用术语第一、第二、第三等来描述各种信息，但这些信息不应受这些术语限制。这些术语仅用来将同一类型的信息彼此区分开。例如，在不脱离本公开范围的情况下，第一信息也可以被称为第二信息，类似地，第二信息也可以被称为第一信息。取决于语境，如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于”。

[32] 本公开的实施例提供了一种注视区域检测用神经网络的训练方法。如图 1 所示，该训练方法可以包括步骤 100~106。

[33] 在步骤 100 中，至少将作为训练样本的人脸图像及其对应的注视区域类别标注信息输入神经网络。

所述的神经网络可以包括例如卷积神经网络、深度神经网络等。

[34] 所述的人脸图像可以是在具体的注视区域检测场景中采集的图像。注视区域检测场景可以有很多，例如，通过检测人的注视区域来自动获悉人对智能设备的控制意图，通过检测人的注视区域来得到人的喜好或意愿，通过检测驾驶员的注视区域来判断驾驶员的驾驶专注力，等等。在不同的场景中，可以采集该场景下目标人的人脸图像。

[35] 所标注的注视区域类别属于预先对指定空间区域划分得到的多类定义注视区域之一。例如，在上面列举的任一种注视区域检测场景中，可以预先指定一个空间区域。人脸图像对应的注视区域检测是要检测该人脸图像中的人注视的区域是所述指定空间区域中的哪个位置。不同的注视位置可能对应着不同的含义。比如，不同的注视位置可以表示驾驶员不同的驾驶专注度；又比如，不同的注视位置可以表示目标人的不同意图。为了区分不同的含义，可以将所述的指定空间区域划分为多个不同的子区域，每一个子区域都可以称为一个注视区域。并且，还可以对这些注视区域分别通过不同标识进行区分，例如，注视区域 A、注视区域 B；或者，注视区域 5、注视区域 6 等。上述列举的 A、B、5、6 等都可以称为注视区域类别。该注视区域类别的定义可以方便神经网络的训练，预先标注的类别可以作为用于训练和测试的标签。

[36] 在步骤 102 中，经所述神经网络对输入的所述人脸图像进行特征提取，并根据提取的特征确定所述人脸图像的注视区域类别预测信息。

[37] 神经网络从输入的人脸图像提取的特征包括该人脸图像的多种图像特征。可以根据这些提取的特征输出该人脸图像的注视区域类别预测信息，其可以是预先定义的某个注视区域的类别。例如，该类别可以用字母或数字表示。示例性的，一个人脸图像经神经网络提取特征后，输出的注视区域类别预测信息是“5”，即注视区域 5。

[38] 在步骤 104 中，确定该注视区域类别预测信息与该人脸图像对应的注视区域类别标注信息之间的差异。例如，可以通过损失函数确定注视区域类别预测信息与注视区域类别标注信息之间的差异。

[39] 在步骤 106 中，基于所述差异调整所述神经网络的参数。例如，可以通过梯度反向传播方法调整神经网络的参数。

[40] 根据本实施例，通过利用作为训练样本的人脸图像及其对应的注视区域类别标注信息训练神经网络，使得能够根据该神经网络直接预测人脸图像对应的注视区域。即使驾驶员的视线稍微偏移或改变，也不会影响检测结果，从而可以提高检测的容错性。

[41] 如下的描述中，将对该注视区域检测用神经网络的训练方法进行更详细的描述。下文以车辆驾驶员注意力监控场景为例描述该训练方法，其中，输入神经网络的人脸图像基于针对车的空间区域中的驾驶区域采集到的图像确定。例如，可以对驾驶区域采集一个图像，并裁剪该图像中的人脸区域得到车辆驾驶员的人脸图像。并且，在该车辆驾驶员注意力监控的场景中，预先定义的注视区域是驾驶员在驾驶时可能注视的多个区域。

[42] 但是可以理解的是，其他场景同样可以应用该相同的训练方法，区别在于输入神经网络的人脸图像可以随着应用场景的不同而有所差异，以及在不同的场景中注视区域所在的指定空间区域也可能不同。

示例性的，指定空间区域可以是车辆的空间区域，也可以是其它空间区域，比如可以是某个智能设备所在的空间；即使是车辆的空间区域，在非驾驶员注意力监控场景中可以是图 2 示例的区域之外的其他车辆空间区域。

[43] 在车辆驾驶员注意力监控应用中，为了降低交通事故，提高行车安全性，一个可行的措施就是通过监控驾驶员的注视区域来判断驾驶员是否分心。驾驶员的注视区域可以指预先对指定空间区域划分得到的多类定义注视区域中驾驶员当前注视的区域。该指定空间区域可以根据车辆结构确定，并且可以被划分成多个注视区域。可以将所述的多个注视区域分别定义为不同的注视区域类别，每种类别以相应的标识来表示。例如，定义某个注视区域的类别是 B。

[44] 图 2 示例了根据本公开的实施例的在车辆驾驶员注意力监控场景下预先定义的多个注视区域。例如，所述的多个注视区域可以包括左前挡风玻璃 21、右前挡风玻璃 22、仪表盘 23、左后视镜 24、右后视镜 25、车内后视镜 26、中控台 27、遮阳板 28、换挡杆 29、方向盘下方 30 以及副驾驶区域、副驾驶前方的杂物箱区域等。需要注意的是，以上仅是示例性的几种，根据实际需求，注视区域的数量可以增加或减少，并且注视区域的范围可以缩放。

[45] 在基于驾驶员注视区域的检测来判断驾驶员是否分心时，可以按照如下方式：例如，正常驾驶时，驾驶员的注视区域通常主要在前挡风玻璃 21，而如果监测到在一段时间内驾驶员的注视区域一直集中在仪表盘 23，则可以确定驾驶员分心。

[46] 基于上述，可以提供一种端到端的用于检测注视区域的神经网络，该神经网络可以用于检测车辆内驾驶员的注视区域。该神经网络的输入可以是摄像头采集的驾驶员人脸图像，神经网络可以直接输出该驾驶员的注视区域的标识。例如，若神经网络检测到驾驶员的注视区域是右前挡风玻璃 22，则神经网络可以直接输出右前挡风玻璃 22 的标识例如“B”。这种端到端的神经网络，可以更快速的检测出驾驶员的注视区域。

[47] 下面将详细描述用于驾驶员注视区域检测的神经网络的训练以及该神经网络的实际应用。

[48] <用于检测驾驶员注视区域的神经网络的训练>

[49] 【准备样本】

[50] 在训练神经网络之前，首先可以准备样本集，该样本集可以包括：用于训练神经网络的训练样本、以及用于测试神经网络的测试样本。

[51] 为了采集样本，可以预先确定待检测的各个注视区域。例如，示例性的，可以预先确定图 2 中所示的十个注视区域。训练该神经网络的目的是使得神经网络能够自动检测输入的驾驶员人脸图像对应这十个注视区域中的哪一个。并且，可以分别为上述十个注视区域指定相应的标识，例如，换挡杆标识“A”、右前挡风玻璃标识“B”等，该标识用于方便后续的神经网络训练和测试。上述的标识在后续描述中也可以称为注视区域的“类别”。

[52] 在确定各个注视区域及对应的类别表示后，可以指示被采集者坐在车辆中的驾驶员位置，并依次注视上述的十个注视区域。每当被采集者注视其中的一个注视区域时，可以通过安装在车辆内的摄像头采集对应此注视区域的驾驶员人脸图像。针对每一个注视区域可以采集该被采集者的多个人脸图像。

[53] 可以建立每个注视区域的类别与针对该注视区域采集的人脸图像之间的对应关系，其中所述“类别”可以作为相应人脸图像的注视区域类别标注信息，即每张人脸图像是当驾驶员注视所述类别标注信息对应的注视区域时采集的图像。最后，可以将采集的大量样本划分为训练集和测试集，训练集中的训练样本用于训练神经网络，测试集中的测试样本用于测试神经网络。每一个训练样本可以包括：驾驶员的一张人脸图像、以及该人脸图像对应的注视区域类别标注信息。

[54] 【确定神经网络结构】

[55] 本公开的实施例中，可以训练用于检测驾驶员注视区域的神经网络。例如，该神经网络可以是卷积神经网络（Convolutional Neural Networks, CNN）或深度神经网络等等。可选地，该神经网络可以包括卷积层（Convolutional Layer）、池化层（Pooling Layer）、修正线性单元（Rectified Linear Unit, ReLU）层、全连接层（Fully Connected Layer）等网络单元，其中上述网络单元按照一定方式堆叠。

[56] 图 3 示例了本公开的实施例可以应用的一种 CNN 300 的网络结构的示例。

[57] 如图 3 所示，CNN 300 可以通过特征提取层 301 由输入图像 302 中提取特征。该特征提取层 301 例如可以包括交替连接在一起的多个卷积层和池化层。每个卷积层可以通过多个卷积核分别提取图像中的不同特征，得到特征图（Feature Map）303。每个池化层位于相应的卷积层之后，可以对该特征图进行局部平均和降采样的操作，以降低特征图的分辨率。随着卷积层和池化层数量的增加，特征图的数目逐渐增多，而且特征图的分辨率逐渐降低。

[58] 将特征提取层 301 最终提取到的特征图中的各个特征进行平铺展开，就可以得到一个特征向量 304，作为全连接层 305 的输入向量。该全连接层 305 可以通过多个隐藏层将特征向量 304 转换成分类器的输入向量 306。由于训练该 CNN 是为了检测出输入图像 302 对应哪个注视区域，所以全连接层 305 最后要通过分类器输出一个分类向量 307，该分类向量 307 中包括输入图像分别对应于各个注视区域的概率。该输入向量 306 中包括的元素数量与分类向量 307 的元素数量相同，都是待检测的注视区域的数目。

[59] 在训练该 CNN 之前，可以设定一些参数。例如，可以设定特征提取层 301 中包括的卷积层和池化层的数量，可以设定每一个卷积层使用的卷积核的数量，还可以设定卷积核的尺寸大小等。而对于卷积核的取值、全连接层的权重等参数，可以通过 CNN 网络的迭代训练进行自学习。具体的 CNN 网络训练方法可以采用常规的训练方式，不再详述。

[60] 在准备了训练样本和初始化 CNN 网络结构的基础上，可以开始进行神经网络训练。下面将描述训练用于检测驾驶员注视区域的神经网络的几种示例方式。

[61] 【训练神经网络的方式一】

[62] 图 4 例示了根据本公开的实施例的用于训练神经网络的配置，其中，CNN 网络的结构可以如图 3 所示，可以将训练样本中的人脸图像输入 CNN 网络。

[63] 示例性的，所述的人脸图像可以是基于由车辆内安装的摄像头采集到的驾驶员的上身图像而得到的。上身图像可以是一个拍摄范围较大的图像，比如，可以涉及人脸、肩膀、脖子等部位。可以通过人脸检测将上身图像裁剪成主要包括驾驶员人脸的人脸图像。

[64] 神经网络可以从输入的人脸图像提取图像特征，并根据该图像特征输出该人脸图像对应的注视区域的类别预测信息，即预测该人脸图像是驾驶员注视哪个类别的注视区域时采集到的。该人脸图像对应的注视区域是根据驾驶员的所乘车辆的结构预先划分的多个注视区域之一，且所述类别作为所述注视区域的标识。

[65] 例如，CNN 网络通过卷积层、池化层、全连接层对输入的人脸图像进行处理后，可以输出分类向量，该分类向量可以包括输入图像分别对应于各个注视区域的概率。如图 4 所示，“A”、“B”、“C”……“J”分别表示十个注视区域的类别，“0.2”表示“输入图像对应于注视区域 A 的概率是 20%”，“0.4”表示“输入图像对应于注视区域 J 的概率是 40%”。假设 J 对应的概率最高，那么“J”将是针对本次输入的人脸图像 CNN 网络得到的注视区域的类别预测信息。如果预先标注的该人脸图像对应的注视区域类别标注信息是 C，那么显然类别预测信息（J）与类别标注信息（C）之间是存在差异的。于是，可以根据类别预测信息与类别标注信息之间的差异得到损失函数的损失（loss）值。

[66] 可以将训练样本分成多个图像子集（batch），以便对神经网络进行迭代训练。每次迭代训练时向神经网络输入一个图像子集。针对所输入的图像子集中的各个训练样本，神经网络输出类别预测结果，并将损失值反馈给神经网络以调整神经网络的参数，例如调整全连接层的权重、卷积核的取值等参数。本次迭代训练完成后，可向神经网络输入下一个图像子集，以进行下一次迭代训练。不同图像子集包括的训练样本至少部分不同。当达到预定的训练结束条件时，就可以获得训练完成的 CNN 网络作为用于检测驾驶员注视区域的神经网络。所述的预定训练结束条件，例如，可以是损失值低于一定阈值，或者达到了预定的迭代训练次数。

[67] 根据本实施例训练得到的神经网络可以将驾驶员的人脸图像作为输入，而输出该人脸图像对应的注视区域检测类别，从而可以快速检测出驾驶员的注视区域，方便后续根据注视区域判断驾驶员是否分心。

[68] 【训练神经网络的方式二】

[69] 为了提高注视区域检测的准确率，本实施例中调整了神经网络的输入配置。

[70] 参见图 5 中的配置，神经网络的输入可以包括：人脸图像和眼睛图像。眼睛图像可以由人脸图像中裁剪得到。例如，可以从人脸图像中检测出人脸的关键点，例如，眼睛关键点、鼻子关键点、眉毛关键点等。然后可以根据检测出的关键点对人脸图像进行裁剪，得到眼睛图像，该眼睛图像主要包括驾驶

员的眼睛。

[71] 需要说明的是，所述的眼睛图像可以包括：左眼图像和右眼图像中的至少一个。例如，神经网络的输入可以包括人脸图像和左眼图像，或者包括人脸图像和右眼图像，或者包括人脸图像、左眼图像以及右眼图像。图 5 中以人脸图像和左右眼图像同时输入为例。

5 [72] 通过将人脸图像和眼睛图像同时输入神经网络进行训练，神经网络能够同时学习脸部和眼睛的特征，增加了特征的多样性和表征能力，从而使得训练后的神经网络能够更准确地检测注视区域类别。

[73] 图 6 是与图 5 中的配置对应的神经网络训练方法的流程图。如图 6 所示，该训练方法可以包括步骤 600~612。

[74] 在步骤 600 中，检测所述人脸图像中的人脸关键点，例如眼睛关键点等。

10 [75] 在步骤 602 中，根据所述人脸关键点裁剪所述人脸图像，得到包括人脸图像中人眼睛的眼睛图像。

[76] 例如，该眼睛图像中包括驾驶员的眼睛。该眼睛图像可以包括驾驶员的左眼图像和右眼图像。图 7 示例了由人脸图像 71 裁剪得到左眼图像 72 和右眼图像 73。

[77] 在步骤 604 中，将所述人脸图像和眼睛图像调整到相同的预定尺寸。

[78] 在步骤 606 中，将调整尺寸后的人脸图像和眼睛图像同时输入同一个神经网络的同一个特征提取层。

15 [79] 在步骤 608 中，神经网络的特征提取层同时提取所述人脸图像中的特征和眼睛图像中的特征，得到提取的特征向量，所述特征向量包括所述人脸图像中的特征和眼睛图像中的特征。

[80] 例如，CNN 的特征提取层可以同时学习人脸的特征和左右眼的特征，提取到包括了人脸图像特征和眼睛图像特征的特征向量。示例性的，CNN 可以通过多个卷积层、池化层等，提取到多个特征图，该多个特征图中包括了人脸图像特征和眼睛图像特征，根据该多个特征图得到所述特征向量。

20 [81] 在步骤 610 中，根据特征向量，确定驾驶员的注视区域类别预测信息。

[82] 例如，可以通过 CNN 中的全连接层，将所述特征向量转化为一个中间向量，该中间向量的维度数量与注视区域的类别数量相同。并且，可以根据该中间向量，通过分类算法计算所述驾驶员的人脸图像分别对应于注视区域的各个类别的概率，并将最大概率对应的类别作为所述类别预测信息。所述的中间向量例如可以是分类器的输入向量 306。

25 [83] 在步骤 612 中，基于该类别预测信息与该人脸图像对应的类别标注信息之间的差异，调整所述神经网络的参数。

[84] 例如，可以基于类别预测信息和类别标注信息之间的差异，计算该训练样本的损失函数的损失值，并可以基于一组训练样本的各个损失函数的损失值，调整 CNN 的参数。

30 [85] 根据本实施例，可以同时将人脸图像和眼睛图像作为神经网络的输入，使得神经网络能够同时学习脸部和眼睛的特征。由于眼睛的特征为注意力检测非常相关的部位，结合人脸图像和眼睛图像可强化所提取出来的特征在注意力方面的表征能力，从而提高神经网络对注视区域类别的检测准确率。

[86] 【训练神经网络的方式三】

[87] 图 8 为根据本公开的另一实施例的神经网络训练方法的流程图，图 9 示例了与该神经网络训练方法对应的配置。如图 8 所示，该训练方法可以包括步骤 800~812。

35 [88] 在步骤 800 中，检测人脸图像中的人脸关键点，例如眼睛关键点等。

[89] 在步骤 802 中，根据所述人脸关键点（如眼睛关键点）裁剪所述人脸图像，得到包括人脸图像中人眼睛的眼睛图像。例如，得到的眼睛图像可以包括左眼图像和/或右眼图像。

[90] 在步骤 804 中，将所述人脸图像、左眼图像和/或右眼图像同时输入神经网络的相应的特征提取分支。

40 [91] 本实施例中，可以不必像图 6 中示出的那样调整人脸图像和眼睛图像的尺寸，而可以将未经尺寸调整的人脸图像和眼睛图像分别输入神经网络的相应特征提取分支，即，输入神经网络的人脸图像和眼睛图像的尺寸可能不同。例如，根据图 9 示例的配置，可以将人脸图像、左眼图像和右眼图像分别输入第一特征提取分支、第二特征提取分支和第三特征提取分支，其中，左眼图像和右眼图像的尺寸可能相同，

人脸图像的尺寸大于左眼图像和右眼图像的尺寸。例如，这三个特征提取分支中的每一个可以包括多个卷积层、池化层等用于提取图像特征，该三个特征提取分支的结构可以相同或不同，例如，可以包括不同的卷积层数量，或者具有不同的卷积核数量。

[92] 在步骤 806 中，神经网络的一个特征提取分支提取所述人脸图像中的特征，得到提取的人脸特征向量；此外，神经网络的其他特征提取分支提取眼睛图像中的特征，得到提取的眼睛特征向量。

[93] 例如，参见图 9，上述的三个特征提取分支可以分别学习各个图像中的特征，其中，第一特征提取分支可以由人脸图像提取到人脸特征向量 91，第二特征提取分支可以由左眼图像提取到左眼特征向量 92，第三特征提取分支可以由右眼图像提取到右眼特征向量 93，左眼特征向量 92 和右眼特征向量 93 均可以称为眼睛特征向量。

[94] 在步骤 808 中，将所述人脸特征向量和眼睛特征向量进行融合，得到融合特征向量即融合特征。例如，参见图 9，可以将人脸特征向量 91、左眼特征向量 92 和右眼特征向量 93 进行融合，得到融合特征向量 94。所述的特征向量融合可以是多个向量按任意顺序拼接组合在一起。

[95] 在步骤 810 中，根据融合特征向量，得到驾驶员的注视区域类别预测信息。

[96] 例如，可以通过 CNN 中的全连接层，将融合特征向量转化为一个中间向量，该中间向量的维度数量与注视区域的类别数量相同。并且，可以根据该中间向量，通过分类算法计算所述驾驶员的人脸图像分别对应于注视区域的各个类别的概率，并将最大概率对应的类别作为所述类别预测信息。

[97] 在步骤 812 中，基于该类别预测信息与该人脸图像对应的类别标注信息之间的差异，调整所述神经网络的参数。

[98] 例如，可以基于类别预测信息和类别标注信息之间的差异，计算该训练样本的损失函数的损失值，并可以基于一组训练样本的各个损失函数的损失值，调整神经网络的参数。

[99] 根据本实施例，可以将未经尺寸调整的人脸图像和眼睛图像输入神经网络，由神经网络中的不同特征提取分支分别提取人脸图像中的特征和眼睛图像中的特征，从而可以减少甚至避免因图像尺寸调整带来的图像质量损失，使得能够更准确的提取到脸部和眼睛的特征。此外，可以将脸部特征和眼睛特征融合来强化特征在注意力方面的表征能力，使得基于融合特征进行注视区域的类别检测更准确。

[100] 在根据本公开的任一实施例的注视区域检测用神经网络的训练方法中，神经网络可以通过分类算法，将对应于不同类别注视区域的特征向量在特征空间内区分开。然而，根据对应于不同注视区域的训练数据提取出的特征向量可能在特征空间中距离很近。在实际使用时，根据训练数据提取出的特征向量在特征空间中有可能离真实注视区域中心的距离大于离相邻注视区域中心的距离，从而可能造成判断错误。

[101] 鉴于此，为了提升神经网络提取出的特征向量的质量，可以将神经网络提取出的图像特征（例如，包括人脸图像特征和眼睛图像特征的特征向量）与多个类别权重分别进行点积运算，得到中间向量。所述多个类别权重分别与所述注视区域的多个类别对应。所述中间向量的维度数量与所述注视区域的类别数量相同。当所述图像特征与所述人脸图像的注视区域类别标注信息对应的类别权重进行点积运算时，调整该图像特征与该类别权重之间的向量夹角余弦值，以增大类间距离且缩小类内距离。

[102] 例如，可以采用大裕量 softmax 算法提升神经网络提取出的特征向量的质量，增强神经网络提取的特征的紧致性，以提升最终注视区域分类的准确性。该算法可以被表示为如下的公式（1），其中， L_i 表示样本 i 的损失函数的损失值， θ_{y_i} 是 W_{y_i} 与 x_i 之间的夹角， W_{y_i} 可以是分别与各个注视区域类别对应的类别权重， x_i 可以是 CNN 提取出的根据特征图得到的图像特征， y_i 可以是各个注视区域的类别， i 可以是第 i 个训练样本， $\|W_{y_i}\| \|x_i\| \varphi(\theta_{y_i})$ 可以称为所述的中间向量，当 $j = y_i$ 时，表示图像特征与人脸图像的注视区域类别标注信息对应的类别权重做点积。

$$L_i = -\log \left(\frac{e^{\|\mathbf{W}_{y_i}\| \|\mathbf{x}_i\| \psi(\theta_{y_i})}}{e^{\|\mathbf{W}_{y_i}\| \|\mathbf{x}_i\| \psi(\theta_{y_i})} + \sum_{j \neq y_i} e^{\|\mathbf{W}_j\| \|\mathbf{x}_i\| \cos(\theta_j)}} \right)$$

$$\psi(\theta) = \begin{cases} \cos(m\theta), & 0 \leq \theta \leq \frac{\pi}{m} \\ \mathcal{D}(\theta), & \frac{\pi}{m} < \theta \leq \pi \end{cases} \quad (1)$$

[103] 上面以驾驶员注意力监控场景为例，详细说明了基于两种可行的神经网络结构的三种训练方法。
5 在驾驶员注意力监控场景之外的其他场景中，可以采用同样的方式来训练所述其他场景中使用的神经网络，只要采用在相应场景中采集的人脸图像以及在相应场景中预定义的注视区域即可。

[104] 下面将说明如何应用训练完成的注视区域检测用神经网络。当然，注视区域检测所使用的神经网络也可以是通过本公开中描述的训练方法之外的其他方式训练得到的。图 10 示例了根据本公开的实施例的一种注视区域检测方法的流程图。如图 10 所示，该方法可以包括步骤 1000~1004。

10 [105] 在步骤 1000 中，截取在指定空间区域内采集到的图像中的人脸区域，得到人脸图像。例如，在指定空间区域采集的图像可以是包括人脸的较大范围的图像，可以由该图像中截取出人脸区域，从而得到人脸图像。

[106] 在步骤 1002 中，将所述人脸图像输入神经网络，其中，所述神经网络预先采用包括多个人脸图像样本及其分别对应的注视区域类别标注信息的训练样本集训练完成，标注的注视区域类别属于预先对
15 所述指定空间区域划分得到的多类定义注视区域之一。

[107] 例如，根据本实施例的神经网络可以是采用图 1 所示的训练方法得到的神经网络，步骤 1000 中获取的人脸图像可以被输入该神经网络。

[108] 在步骤 1004 中，经所述神经网络对输入的所述人脸图像进行特征提取，并根据提取的特征确定所述人脸图像对应的注视区域检测类别。

20 [109] 本步骤中，可以通过所述神经网络预测出人脸图像对应的注视区域，该预测出的注视区域可以称为注视区域检测类别。该注视区域检测类别可以用例如字母、数字、名称等不同的方式表示。

[110] 根据本实施例的端到端的注视区域检测方法，通过预先训练的神经网络，可以直接预测出人脸图像对应的注视区域检测类别。即使驾驶员的视线稍微偏移或改变，也不会影响检测结果，从而可以提高检测的容错性。

25 [111] 下面仍将以驾驶员注意力监控场景为例，说明在该场景下训练的神经网络如何被应用。可以理解的是，其他场景下训练的神经网络也可以被类似地应用。

[112] 参见图 11，可以将上述训练好的任一种神经网络应用于检测驾驶员的注视区域。具体地，驾驶员的所乘车辆 1101 内可以安装有摄像头 1102，该摄像头 1102 可以采集包括驾驶员脸部的图像 1103。图像 1103 可以传输至车辆内的图像处理设备 1104，预先训练完成的神经网络 1108 可以存储在该图像处理
30 设备 1104 中。

[113] 图像处理设备 1104 可以对图像 1103 进行预处理，然后将得到的图像输入神经网络 1108。例如，可以通过例如人脸检测由图像 1103 中截取出人脸区域，得到人脸图像 1105。还可以由人脸图像 1105 裁剪得到左眼图像 1106 和右眼图像 1107。人脸图像 1105、左眼图像 1106 和右眼图像 1107 可以同时输入预先训练好的神经网络 1108，使得神经网络 1108 输出车辆内驾驶员的注视区域检测类别。如前所述，人脸图像 1105、左眼图像 1106 和右眼图像 1107 可以被调整到相同的预定尺寸后输入神经网络 1108，或者
35 可以不经尺寸调整而分别输入神经网络 1108 的相应的特征提取分支。

[114] 图 12 示例了图 11 所示的应用场景中神经网络 1108 输出的注视区域检测类别的示例。图 12 所示的驾驶员图像可以是由驾驶员所乘车辆内部署的摄像头 1102 采集的。该车辆内的图像处理设备 1104 可以由该驾驶员图像中截取出驾驶员的人脸图像 1201。该人脸图像 1201 可以输入图像处理设备 1104 中的神经网络 1108。神经网络 1108 可以输出人脸图像 1201 对应的车辆内驾驶员的注视区域检测类别“[5]: 中控台”, 如图 12 中所示。

[115] 由此可见, 根据本实施例的驾驶员注视区域检测方法实时性较好, 可以快速准确的检测出驾驶员的注视区域。

[116] 此外, 对于不同的注视区域, 同一驾驶员可能有不同的头部姿态。若仅使用单个摄像头采集驾驶员的图像, 则不论摄像头安装在车内什么位置, 都有可能因驾驶员头部转动而导致单只眼睛甚至双眼不可见的情况, 从而影响最终注视区域的判断。另外, 对于佩戴眼镜的驾驶员来说, 摄像头在某个角度刚好拍摄到镜片反光导致眼睛区域被部分或全部遮挡的情况也经常出现。为解决以上问题, 可以在车内不同位置分别安装多个摄像头来采集驾驶员的图像。

[117] 例如, 可以在驾驶员的所乘车辆 1101 内安装多个摄像头 1102, 该多个摄像头 1102 可以分别从不同角度采集在车内驾驶区域的同一驾驶员的图像。可以使多个摄像头的采集时间同步, 或者记录每帧图像的采集时间, 以便后续处理中获取在同一时刻由不同摄像头分别采集的同一驾驶员的多个图像。

[118] 可以理解的是, 在驾驶员注意力监控场景之外的任一其他场景中, 可以在该场景的指定空间区域内部署多个摄像头, 以针对所述指定空间区域的一特定子区域分别采集图像。例如, 在控制智能设备的场景中, 所述的特定子区域可以是控制智能设备的目标人所在的区域。通过针对所述特定子区域采集图像, 可以得到包括人的脸部的图像, 并据此检测人的注视区域。

[119] 仍以驾驶员注意力监控场景为例, 在获取到多个摄像头在同一时刻 T_k 分别采集的同一驾驶员的多个图像之后, 可以根据该多个图像, 按照例如下列方式中任一种确定该驾驶员在该时刻 T_k 的注视区域。

[120] 方式一: 可以根据图像质量评价指标, 确定该多个图像中图像质量评分最高的图像, 并截取图像质量评分最高的图像中的人脸区域而得到驾驶员的人脸图像。所述图像质量评价指标可以包括下列中至少一种: 图像中是否包括有眼睛图像、图像中眼睛区域的清晰度、图像中眼睛区域的遮挡情况、图像中眼睛的睁/闭情况。比如, 若所采集的一个图像中包括清晰的眼睛图像, 眼睛区域没有遮挡, 且眼睛完全睁开, 则可以确定该图像是图像质量评分最高的图像, 并可以从该图像中截取驾驶员的人脸图像, 将该人脸图像输入预先训练好的神经网络, 从而确定该驾驶员在该时刻 T_k 的注视区域检测类别。

[121] 方式二: 可以参照方式一, 根据图像质量评价指标确定该多个图像中图像质量评分最高的图像。可以分别从该多个图像中截取驾驶员的人脸图像, 分别将所截取的多个人脸图像输入预先训练好的神经网络, 得到该多个人脸图像分别对应的多个注视区域检测类别。可以从所述多个注视区域检测类别中, 选择与上述图像质量评分最高的图像关联的人脸图像对应的注视区域检测类别, 作为该驾驶员在该时刻 T_k 的注视区域检测类别。

[122] 方式三: 可以分别从该多个图像中截取驾驶员的人脸图像, 分别将所截取的多个人脸图像输入预先训练好的神经网络, 得到该多个人脸图像分别对应的多个注视区域检测类别。可以选择所述多个注视区域检测类别中的多数结果作为该驾驶员在该时刻 T_k 的注视区域检测类别。比如, 若根据 6 张人脸图像分别得到的 6 个注视区域检测类别中 5 个都是“C”, 则可以选择“C”作为该驾驶员在该时刻 T_k 的注视区域检测类别。

[123] 作为一个例子, 在检测出驾驶员的注视区域后, 可以根据该注视区域执行进一步的操作。例如, 可以根据注视区域类别检测结果, 确定人脸图像对应的人的注意力监控结果。比如, 所述的注视区域类别检测结果可以是预设时间段内的注视区域检测类别。示例性的, 该注视区域类别检测结果可以是“在预设时间段内, 该驾驶员的注视区域一直是区域 B”。如果该区域 B 是前挡风玻璃, 说明该驾驶员的驾驶较为专心。如果该区域 B 是副驾驶前方的杂物箱区域, 说明该驾驶员很有可能分心了, 注意力不集中。

[124] 在检测出注意力监控结果后，可以输出所述注意力监控结果，例如，可以在车辆内的某个显示区域显示“驾驶很专心”。或者，还可以根据所述注意力监控结果输出分心提示信息，例如在显示屏幕上输出“请注意风险，集中注意力”以提示驾驶员。当然，在具体显示时，可以显示注意力监控结果和分心提示信息中的至少一种。

5 [125] 通过根据注视区域类别检测结果确定人的注意力监控结果或者输出分心提示信息，对于驾驶员注意力监控有着重要的帮助，能够有效检测出驾驶员注意力不集中的情况，及时进行提醒，进而有助于降低事故发生。

[126] 如上的描述中，都是以驾驶员注意力监控场景为例。除此之外，注视区域的检测还可以有其它许多用途。

10 [127] 例如，可以进行基于注视区域检测的车机交互控制。车辆内可以设置有一些电子设备，如多媒体播放器，可以通过检测车辆内人员的注视区域，根据注视区域的检测结果自动控制该多媒体播放器开启播放功能。

[128] 示例性的，通过部署在车辆内的摄像头拍摄得到车内人员（如司机或乘客）的人脸图像，通过预先训练的神经网络检测出注视区域类别检测结果。例如，该检测结果可以是：在一段时间 T 内，该车内人员的注视区域一直是车辆内的某个多媒体播放器上的“注视开启”选项所在的区域。根据上述检测结果可以确定该车内人员要开启该多媒体播放器，从而可以输出相应的控制指令，控制该多媒体播放器开始进行播放。

15 [129] 除了车相关的应用之外，还可以包括游戏控制、智能家居设备控制、广告推送等多种应用场景。以智能家居控制为例，可以采集控制人的人脸图像，通过预先训练的神经网络检测出注视区域类别检测结果。例如，该检测结果可以是：在一段时间 T 内，该控制人的注视区域一直是智能空调上的“注视开启”选项所在的区域。根据上述检测结果可以确定该控制人要启动智能空调，从而可以输出相应的控制指令，控制该空调开启。

[130] 本公开还可以提供与前述方法实施例相对应的装置及设备的实施例。

20 [131] 图 13 为根据本公开的实施例的一种注视区域检测用神经网络的训练装置 1300 的框图。如图 13 所示，该装置 1300 可以包括：样本输入模块 1301、类别预测模块 1302、差异确定模块 1303 和参数调整模块 1304。

[132] 样本输入模块 1301 用于至少将作为训练样本的人脸图像及其对应的注视区域类别标注信息输入神经网络，其中，标注的注视区域类别属于预先对指定空间区域划分得到的多类定义注视区域之一。类别预测模块 1302 用于经所述神经网络对输入的所述人脸图像进行特征提取，并根据提取的特征确定所述人脸图像的注视区域类别预测信息。差异确定模块 1303 用于确定该注视区域类别预测信息与该人脸图像对应的注视区域类别标注信息之间的差异。参数调整模块 1304 用于基于所述差异调整所述神经网络的参数。

30 [133] 根据本公开的实施例，所述样本输入模块 1301 可以在至少将作为训练样本的人脸图像及其对应的注视区域类别标注信息输入神经网络之前，裁剪所述人脸图像中的至少一眼睛区域，得到至少一眼睛图像。

35 [134] 根据本公开的实施例，所述样本输入模块 1301 可以将所述人脸图像和所述至少一眼睛图像调整到相同的预定尺寸后将它们同时输入所述神经网络。所述类别预测模块 1302 可以经所述神经网络同时提取所述人脸图像中的特征和所述至少一眼睛图像中的特征，并根据提取的特征确定所述人脸图像的注视区域类别预测信息。

40 [135] 根据本公开的实施例，所述样本输入模块 1301 可以将所述人脸图像和所述至少一眼睛图像（未经尺寸调整）分别输入所述神经网络的不同的特征提取分支，其中，输入到所述神经网络中的人脸图像和眼睛图像的尺寸可能不同。所述类别预测模块 1302 可以经所述神经网络的相应的特征提取分支分别提

取人脸图像中的特征和眼睛图像中的特征，融合各特征提取分支所提取的特征而得到融合特征，并根据所述融合特征确定所述人脸图像的注视区域类别预测信息。

[136] 根据本公开的实施例，为了根据提取的特征更准确地确定注视区域类别预测信息，所述类别预测模块 1302 可以将所提取的特征与多个类别权重分别进行点积运算，得到中间向量，并根据该中间向量确定该人脸图像的注视区域类别预测信息。所述多个类别权重分别与所述多类定义注视区域对应，所述中间向量的维度数量与所述多类定义注视区域的数量相同。当提取的特征与所述人脸图像的注视区域类别标注信息对应的类别权重进行点积运算时，调整该特征与该类别权重之间的向量夹角余弦值，以增大类间距离且缩小类内距离。

[137] 根据本公开的实施例，所述指定空间区域包括：车的空间区域。

[138] 根据本公开的实施例，所述人脸图像基于针对所述车的空间区域中的驾驶区域采集到的图像确定。所述对指定空间区域划分得到的多类定义注视区域包括下列中至少两类：左前挡风玻璃区域、右前挡风玻璃区域、仪表盘区域、车内后视镜区域、中控台区域、左后视镜区域、右后视镜区域、遮阳板区域、换挡杆区域、方向盘下方区域、副驾驶区域、副驾驶前方的杂物箱区域。

[139] 图 14 为根据本公开的实施例的一种注视区域检测装置 1400 的框图。如图 14 所示，该装置 1400 可以包括：图像获取模块 1401、图像输入模块 1402 和类别检测模块 1403。

[140] 图像获取模块 1401 用于截取在指定空间区域内采集到的图像中的人脸区域，得到人脸图像。图像输入模块 1402 用于将所述人脸图像输入神经网络，其中，所述神经网络预先采用包括多个人脸图像样本及其分别对应的注视区域类别标注信息的训练样本集训练完成，标注的注视区域类别属于预先对所述指定空间区域划分得到的多类定义注视区域之一。类别检测模块 1403 用于经所述神经网络对输入的所述人脸图像进行特征提取，并根据提取的特征确定所述人脸图像对应的注视区域检测类别。

[141] 根据本公开的实施例，用于预先训练所述神经网络的训练样本集中还包括分别从多个人脸图像样本中截取的多个眼睛图像样本。相应地，图像获取模块 1401 可以在得到人脸图像之后，裁剪所述人脸图像中的至少一眼睛区域，得到至少一眼睛图像。

[142] 根据本公开的实施例，所述图像输入模块 1402 可以将所述人脸图像和所述至少一眼睛图像调整到相同的预定尺寸后将它们同时输入所述神经网络。所述类别检测模块 1403 可以经所述神经网络同时提取所述人脸图像中的特征和所述至少一眼睛图像中的特征，并根据提取的特征确定所述人脸图像对应的注视区域检测类别。

[143] 根据本公开的实施例，所述图像输入模块 1402 可以将所述人脸图像和所述至少一眼睛图像（未经尺寸调整）分别输入所述神经网络的不同的特征提取分支，其中，输入到所述神经网络中的人脸图像和眼睛图像的尺寸可能不同。所述类别检测模块 1403 可以经所述神经网络的相应的特征提取分支分别提取人脸图像中的特征和眼睛图像中的特征，融合各特征提取分支分别提取的各特征而得到融合特征，并根据所述融合特征确定所述人脸图像对应的注视区域检测类别。

[144] 可选地，图像获取模块 1401 可以在截取在指定空间区域内采集到的图像中的人脸区域之前，获取通过在指定空间区域部署的多个摄像头，在同一时刻 T_i 针对该指定空间区域的一特定子区域从不同角度分别采集的多个图像。

[145] 可选地，图像获取模块 1401 可以根据图像质量评价指标，确定该多个图像中图像质量评分最高的图像。所述图像质量评价指标可以包括下列中至少一种：图像中是否包括有眼睛图像、图像中眼睛区域的清晰度、图像中眼睛区域的遮挡情况、图像中眼睛的睁/闭情况。

[146] 根据本公开的实施例，图像获取模块 1401 可以截取上述图像质量评分最高的图像中的人脸区域，得到该人脸图像。图像输入模块 1402 可以将该人脸图像输入所述神经网络。类别检测模块 1403 可以经所述神经网络对该人脸图像进行特征提取，并根据提取的特征确定其对应的注视区域检测类别，作为在该时刻 T_i 的注视区域检测类别。

[147] 根据本公开的实施例，图像获取模块 1401 可以分别截取上述多个图像中的人脸区域，得到相应的多个人脸图像。图像输入模块 1402 可以分别将该多个人脸图像输入所述神经网络。针对该多个人脸图像中的每一个，类别检测模块 1403 可以如上所述确定其对应的注视区域检测类别。类别检测模块 1403 可以从所确定的分别与该多个人脸图像对应的多个注视区域检测类别中，选择与上述图像质量评分最高的图像关联的人脸图像对应的注视区域检测类别，作为在该时刻 T_i 的注视区域检测类别。

[148] 根据本公开的实施例，图像获取模块 1401 可以分别截取上述多个图像中的人脸区域，得到相应的多个人脸图像。图像输入模块 1402 可以分别将该多个人脸图像输入所述神经网络。针对该多个人脸图像中的每一个，类别检测模块 1403 可以如上所述确定其对应的注视区域检测类别。类别检测模块 1403 可以选择所确定的分别与该多个人脸图像对应的多个注视区域检测类别中的多数结果，作为在该时刻 T_i 的注视区域检测类别。

[149] 根据本公开的实施例，所述指定空间区域包括：车的空间区域。

[150] 根据本公开的实施例，上述在指定空间区域内采集到的图像包括针对所述车的空间区域中的驾驶区域采集到的图像。所述对指定空间区域划分得到的多类定义注视区域包括下列中至少两类：左前挡风玻璃区域、右前挡风玻璃区域、仪表盘区域、车内后视镜区域、中控台区域、左后视镜区域、右后视镜区域、遮阳板区域、换挡杆区域、方向盘下方区域、副驾驶区域、副驾驶前方的杂物箱区域。

[151] 图 15 为根据本公开的另一实施例的一种注视区域检测装置 1400' 的框图。该装置 1400' 与图 14 中所示的注视区域检测装置 1400 的区别仅在于，该装置 1400' 还可以包括第一类别应用模块 1404 和第二类别应用模块 1405 中的至少一个。

[152] 第一类别应用模块 1404 可以基于类别检测模块 1403 得到的注视区域检测类别而得到注视区域类别检测结果，并根据该注视区域类别检测结果确定该人脸图像对应的人的注意力监控结果。第一类别应用模块 1404 可以输出所述注意力监控结果，和/或，根据所述注意力监控结果输出分心提示信息。

[153] 第二类别应用模块 1405 可以基于类别检测模块 1403 得到的注视区域检测类别而得到注视区域类别检测结果，确定与所述注视区域类别检测结果对应的控制指令，并控制电子设备执行与所述控制指令相应的操作。

[154] 图 16 为根据本公开的实施例的注视区域检测用神经网络的训练设备的框图。如图 16 所示，该设备可以包括存储器 1601 和处理器 1602。所述存储器 1601 中存储有可由处理器 1602 执行的计算机指令。所述处理器 1602 在执行所述计算机指令时，可以实现上述任一种注视区域检测用神经网络的训练方法。

[155] 图 17 为根据本公开的实施例的注视区域检测设备的框图。如图 17 所示，该设备可以包括存储器 1701 和处理器 1702。所述存储器 1701 中存储有可由处理器 1702 执行的计算机指令。所述处理器 1702 在执行所述计算机指令时，可以实现上述任一种注视区域检测方法。

[156] 本公开的实施例还提供了一种计算机可读存储介质，其上存储有计算机程序。该计算机程序被处理器执行时，可以使该处理器实现上述任一种注视区域检测用神经网络的训练方法。

[157] 本公开的实施例还提供了一种计算机可读存储介质，其上存储有计算机程序。该计算机程序被处理器执行时，可以使该处理器实现上述任一种注视区域检测方法。

[158] 本领域技术人员应明白，本公开可提供为方法、装置、系统或计算机程序产品。因此，本公开可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。

[159] 本文中描述的主题及功能操作的实施例可以在以下中实现：数字电子电路、有形体现的计算机软件或固件、包括本文中公开的结构及其结构性等同物的计算机硬件、或者它们中的一个或多个的组合。本文中描述的主题的实施例可以实现为一个或多个计算机程序，即编码在有形非暂时性程序载体上以被数据处理装置执行或控制数据处理装置的操作的计算机程序指令中的一个或多个模块。可替代地或附加地，程序指令可以被编码在生成的传播信号（例如机器生成的电、光或电磁信号）上，该信号被生成以将信息编码并传输到合适的接收机装置以由数据处理装置执行。计算机存储介质可以是机器可读存储设

备、机器可读存储基板、随机或串行存取存储器设备、或它们中的一个或多个的组合。

[160] 本文中描述的处理及逻辑流程可以由执行一个或多个计算机程序的一个或多个可编程计算机执行，以通过根据输入数据进行操作并生成输出来执行相应的功能。所述处理及逻辑流程还可以由专用逻辑电路例如 FPGA（现场可编程门阵列）或 ASIC（专用集成电路）来执行，并且装置也可以实现为专用逻辑电路。

[161] 适合用于执行计算机程序的计算机包括例如通用或专用微处理器，或任何其他类型的中央处理单元。通常，中央处理单元将从只读存储器和/或随机存取存储器接收指令和数据。计算机的基本组件包括用于实施或执行指令的中央处理单元以及用于存储指令和数据的一个或多个存储器设备。通常，计算机可以包括用于存储数据的一个或多个大容量存储设备，例如磁盘、磁光盘或光盘等，或者计算机可以可操作地与此大容量存储设备耦接以从其接收数据或向其传送数据。此外，计算机可以嵌入在另一设备（例如移动电话机、个人数字助理（PDA）、移动音频或视频播放器、游戏操纵台、全球定位系统（GPS）接收机、或例如通用串行总线（USB）闪存驱动器的便携式存储设备等）中。

[162] 适合于存储计算机程序指令和数据的计算机可读介质可以包括各种形式的非易失性存储器，例如半导体存储器设备（例如，可擦可编程只读存储器（Erasable Programmable Read Only Memory, EPROM）、电可擦可编程只读存储器（Electrically Erasable Programmable Read Only Memory, EEPROM）和闪存）、磁盘（例如内部硬盘或可移动盘）、磁光盘、光盘只读存储器（Compact Disc Read Only Memory, CD-ROM）、数字多功能光盘（Digital Versatile Disc, DVD）等。处理器和存储器可由专用逻辑电路补充或并入专用逻辑电路中。

[163] 虽然本文包含许多具体实施细节，但是这些不应被解释为限制本公开的范围或所要求保护的范
围，而是主要用于描述本公开的具体实施例的特征。在多个实施例中分别描述的某些特征也可以在单个实施例中组合实施。另一方面，在单个实施例中描述的各种特征也可以在多个实施例中分开实施或以任何合适的子组合来实施。此外，虽然特征可以如上所述在某些组合中起作用并且甚至最初如此要求保护，但是来自所要求保护的组合中的一个或多个特征在一些情况下可以从该组合中去除，并且所要求保护的组合可以指向子组合或子组合的变型。

[164] 类似地，虽然在附图中以特定顺序描绘了操作，但是这不应被理解为要求这些操作以所示的特定顺序执行或顺次执行或者要求所有例示的操作被执行，以实现期望的结果。在某些情况下，多任务和并行处理可能是有利的。此外，上述实施例中的各种系统模块和组件的分离不应被理解为在所有实施例中均需要这样的分离，并且应当理解，所描述的程序组件和系统通常可以一起集成在单个软件产品中，或者封装成多个软件产品。

[165] 以上所述仅为本公开的一些实施例，并不用以限制本公开。凡在本公开的精神和原则之内所做的任何修改、等同替换、改进等，均应包含在本公开的范围之内。

权利要求书

1、一种注视区域检测用神经网络的训练方法，所述方法包括：

至少将作为训练样本的人脸图像及其对应的注视区域类别标注信息输入所述神经网络，其中，标注的注视区域类别属于预先对指定空间区域划分得到的多类定义注视区域之一；

经所述神经网络对输入的所述人脸图像进行特征提取，并根据提取的特征确定所述人脸图像的注视区域类别预测信息；

确定所述注视区域类别预测信息与所述注视区域类别标注信息之间的差异；

基于所述差异调整所述神经网络的参数。

2、根据权利要求1所述的方法，所述方法还包括：在至少将所述人脸图像及其对应的注视区域类别标注信息输入所述神经网络之前，裁剪所述人脸图像中的至少一眼睛区域，得到至少一眼睛图像；

其中，至少将所述人脸图像及其对应的注视区域类别标注信息输入所述神经网络包括：将所述人脸图像和所述至少一眼睛图像同时输入所述神经网络。

3、根据权利要求2所述的方法，其中，

将所述人脸图像和所述至少一眼睛图像同时输入所述神经网络包括：将所述人脸图像和所述至少一眼睛图像调整到相同的预定尺寸后将它们同时输入所述神经网络；

对输入的所述人脸图像进行特征提取包括：经所述神经网络同时提取所述人脸图像中的特征和所述至少一眼睛图像中的特征。

4、根据权利要求2所述的方法，其中，

将所述人脸图像和所述至少一眼睛图像同时输入所述神经网络包括：将所述人脸图像和所述至少一眼睛图像分别输入所述神经网络的不同的特征提取分支，其中，所述人脸图像和所述至少一眼睛图像的尺寸不同；

对输入的所述人脸图像进行特征提取并确定所述注视区域类别预测信息包括：经所述神经网络的相应的特征提取分支分别提取所述人脸图像中的特征和所述至少一眼睛图像中的特征；融合所述神经网络的相应的特征提取分支分别提取的各特征，得到融合特征；根据所述融合特征确定所述人脸图像的注视区域类别预测信息。

5、根据权利要求1~4中任一所述的方法，其中，根据提取的特征确定所述注视区域类别预测信息包括：

将所提取的特征与多个类别权重分别进行点积运算，得到中间向量，其中，所述多个类别权重分别与所述多类定义注视区域对应，所述中间向量的维度数量与所述多类定义注视区域的数量相同，当所提取的特征与所述注视区域类别标注信息对应的类别权重进行点积运算时，调整该特征与该类别权重之间的向量夹角余弦值，以增大大类间距离且缩小类内距离；

根据所述中间向量，确定所述人脸图像的注视区域类别预测信息。

6、根据权利要求1~5中任一所述的方法，其中，

所述指定空间区域包括：车的空间区域。

7、根据权利要求6所述的方法，其中，

所述人脸图像基于针对所述车的空间区域中的驾驶区域采集到的图像确定；

所述多类定义注视区域包括下列中至少两类：左前挡风玻璃区域、右前挡风玻璃区域、仪表盘区域、车内后视镜区域、中控台区域、左后视镜区域、右后视镜区域、遮阳板区域、换挡杆区域、方向盘下方区域、副驾驶区域、副驾驶前方的杂物箱区域。

8、一种注视区域检测方法，所述方法包括：

截取在指定空间区域内采集到的图像中的人脸区域，得到人脸图像；

将所述人脸图像输入神经网络，其中，所述神经网络预先采用包括多个人脸图像样本及其分别对应的注视区域类别标注信息的训练样本集训练完成，标注的注视区域类别属于预先对所述指定空间区域划分得到的多类定义注视区域之一；

5 经所述神经网络对输入的所述人脸图像进行特征提取，并根据提取的特征确定所述人脸图像对应的注视区域检测类别。

9、根据权利要求8所述的方法，其中，

用于预先训练所述神经网络的所述训练样本集中还包括分别从所述多个人脸图像样本中截取的多个眼睛图像样本；

10 所述方法还包括：在得到所述人脸图像之后，裁剪所述人脸图像中的至少一眼睛区域，得到至少一眼睛图像；

将所述人脸图像输入所述神经网络包括：将所述人脸图像和所述至少一眼睛图像同时输入所述神经网络。

10、根据权利要求9所述的方法，其中，

15 将所述人脸图像和所述至少一眼睛图像同时输入所述神经网络包括：将所述人脸图像和所述至少一眼睛图像调整到相同的预定尺寸后将它们同时输入所述神经网络；

对输入的所述人脸图像进行特征提取包括：经所述神经网络同时提取所述人脸图像中的特征和所述至少一眼睛图像中的特征。

11、根据权利要求9所述的方法，其中，

20 将所述人脸图像和所述至少一眼睛图像同时输入所述神经网络包括：将所述人脸图像和所述至少一眼睛图像分别输入所述神经网络的不同的特征提取分支，其中，所述人脸图像和所述至少一眼睛图像的尺寸不同；

25 对输入的所述人脸图像进行特征提取并确定所述注视区域检测类别包括：经所述神经网络的相应的特征提取分支分别提取所述人脸图像中的特征和所述至少一眼睛图像中的特征；融合所述神经网络的相应的特征提取分支分别提取的各特征，得到融合特征；根据所述融合特征确定所述人脸图像对应的注视区域检测类别。

12、根据权利要求8~11中任一所述的方法，

所述方法还包括：在截取在所述指定空间区域内采集到的图像中的人脸区域之前，

30 获取通过在该指定空间区域部署的多个摄像头，在同一时刻针对该指定空间区域的一特定子区域从不同角度分别采集的多个图像；以及

根据图像质量评价指标，确定所述多个图像中图像质量评分最高的图像，

其中，截取在所述指定空间区域内采集到的图像中的人脸区域包括：截取所述图像质量评分最高的图像中的人脸区域。

13、根据权利要求8~11中任一所述的方法，

所述方法还包括：在截取在所述指定空间区域内采集到的图像中的人脸区域之前，

35 获取通过在该指定空间区域部署的多个摄像头，在同一时刻针对该指定空间区域的一特定子区域从不同角度分别采集的多个图像；以及

根据图像质量评价指标，确定所述多个图像中图像质量评分最高的图像，

其中，截取在所述指定空间区域内采集到的图像中的人脸区域以得到人脸图像包括：分别截取所述多个图像中的人脸区域，得到相应的多个人脸图像；

40 将所述人脸图像输入所述神经网络包括：分别将该多个人脸图像输入所述神经网络；

对输入的所述人脸图像进行特征提取并确定所述人脸图像对应的注视区域检测类别包括：针对该多个人脸图像中的每一个人脸图像，经所述神经网络对该人脸图像进行特征提取，并根据提取的特征确定

该人脸图像对应的注视区域检测类别；

所述方法还包括：从所确定的分别与该多个人脸图像对应的多个注视区域检测类别中，选择与所述图像质量评分最高的图像关联的人脸图像对应的注视区域检测类别，作为在所述时刻的注视区域检测类别。

5 14、根据权利要求 12 或 13 所述的方法，其中，所述图像质量评价指标包括下列中至少一种：图像中是否包括有眼睛图像、图像中眼睛区域的清晰度、图像中眼睛区域的遮挡情况、图像中眼睛的睁/闭情况。

15 15、根据权利要求 8~11 中任一所述的方法，

10 所述方法还包括：在截取在所述指定空间区域内采集到的图像中的人脸区域之前，获取通过在该指定空间区域部署的多个摄像头，在同一时刻针对该指定空间区域的一特定子区域从不同角度分别采集的多个图像，

其中，截取在所述指定空间区域内采集到的图像中的人脸区域以得到人脸图像包括：分别截取所述多个图像中的人脸区域，得到相应的多个人脸图像；

将所述人脸图像输入所述神经网络包括：分别将该多个人脸图像输入所述神经网络；

15 对输入的所述人脸图像进行特征提取并确定所述人脸图像对应的注视区域检测类别包括：针对该多个人脸图像中的每一个人脸图像，经所述神经网络对该人脸图像进行特征提取，并根据提取的特征确定该人脸图像对应的注视区域检测类别；

所述方法还包括：选择所确定的分别与该多个人脸图像对应的多个注视区域检测类别中的多数结果，作为在所述时刻的注视区域检测类别。

20 16、根据权利要求 8~15 中任一所述的方法，其中，

所述指定空间区域包括：车的空间区域。

17、根据权利要求 16 所述的方法，其中，

所述在指定空间区域内采集到的图像包括：针对所述车的空间区域中的驾驶区域采集到的图像；

25 所述多类定义注视区域包括下列中至少两类：左前挡风玻璃区域、右前挡风玻璃区域、仪表盘区域、车内后视镜区域、中控台区域、左后视镜区域、右后视镜区域、遮阳板区域、换挡杆区域、方向盘下方区域、副驾驶区域、副驾驶前方的杂物箱区域。

18、根据权利要求 8~17 中任一所述的方法，所述方法还包括：

基于所述注视区域检测类别得到注视区域类别检测结果，并根据该注视区域类别检测结果，确定所述人脸图像对应的人的注意力监控结果；

30 输出所述注意力监控结果，和/或，根据所述注意力监控结果输出分心提示信息。

19、根据权利要求 8~17 中任一所述的方法，所述方法还包括：

基于所述注视区域检测类别得到注视区域类别检测结果，并确定与所述注视区域类别检测结果对应的控制指令；

控制电子设备执行与所述控制指令相应的操作。

35 20、一种注视区域检测用神经网络的训练装置，所述装置包括：

样本输入模块，用于至少将作为训练样本的人脸图像及其对应的注视区域类别标注信息输入所述神经网络，其中，标注的注视区域类别属于预先对指定空间区域划分得到的多类定义注视区域之一；

类别预测模块，用于经所述神经网络对输入的所述人脸图像进行特征提取，并根据提取的特征确定所述人脸图像的注视区域类别预测信息；

40 差异确定模块，用于确定所述注视区域类别预测信息与所述注视区域类别标注信息之间的差异；

参数调整模块，用于基于所述差异调整所述神经网络的参数。

21、根据权利要求 20 所述的装置，其中，

所述样本输入模块用于：在至少将所述人脸图像及其对应的注视区域类别标注信息输入所述神经网络之前，裁剪所述人脸图像中的至少一眼睛区域，得到至少一眼睛图像；将所述人脸图像和所述至少一眼睛图像同时输入所述神经网络。

22、根据权利要求 21 所述的装置，其中，

5 所述样本输入模块用于：将所述人脸图像和所述至少一眼睛图像调整到相同的预定尺寸后将它们同时输入所述神经网络；

所述类别预测模块用于：经所述神经网络同时提取所述人脸图像中的特征和所述至少一眼睛图像中的特征，并根据提取的特征确定所述人脸图像的注视区域类别预测信息。

23、根据权利要求 21 所述的装置，其中，

10 所述样本输入模块用于：将所述人脸图像和所述至少一眼睛图像分别输入所述神经网络的不同的特征提取分支，其中，所述人脸图像和所述至少一眼睛图像的尺寸不同；

所述类别预测模块用于：经所述神经网络的相应的特征提取分支分别提取所述人脸图像中的特征和所述至少一眼睛图像中的特征；融合所述神经网络的相应的特征提取分支分别提取的各特征，得到融合特征；根据所述融合特征确定所述人脸图像的注视区域类别预测信息。

15 24、根据权利要求 20~23 中任一所述的装置，其中，所述类别预测模块用于：

将所提取的特征与多个类别权重分别进行点积运算，得到中间向量，其中，所述多个类别权重分别与所述多类定义注视区域对应，所述中间向量的维度数量与所述多类定义注视区域的数量相同，当所提取的特征与所述注视区域类别标注信息对应的类别权重进行点积运算时，调整该特征与该类别权重之间的向量夹角余弦值，以增大类间距离且缩小类内距离；

20 根据所述中间向量，确定所述人脸图像的注视区域类别预测信息。

25、根据权利要求 20~24 中任一所述的装置，其中，所述指定空间区域包括：车的空间区域。

26、根据权利要求 25 所述的装置，其中，

所述人脸图像基于针对所述车的空间区域中的驾驶区域采集到的图像确定；

25 所述多类定义注视区域包括下列中至少两类：左前挡风玻璃区域、右前挡风玻璃区域、仪表盘区域、车内后视镜区域、中控台区域、左后视镜区域、右后视镜区域、遮阳板区域、换挡杆区域、方向盘下方区域、副驾驶区域、副驾驶前方的杂物箱区域。

27、一种注视区域检测装置，所述装置包括：

图像获取模块，用于截取在指定空间区域内采集到的图像中的人脸区域，得到人脸图像；

30 图像输入模块，用于将所述人脸图像输入神经网络，其中，所述神经网络预先采用包括多个人脸图像样本及其分别对应的注视区域类别标注信息的训练样本集训练完成，标注的注视区域类别属于预先对所述指定空间区域划分得到的多类定义注视区域之一；

类别检测模块，用于经所述神经网络对输入的所述人脸图像进行特征提取，并根据提取的特征确定所述人脸图像对应的注视区域检测类别。

28、根据权利要求 27 所述的装置，其中，

35 用于预先训练所述神经网络的所述训练样本集中还包括分别从所述多个人脸图像样本中截取的多个眼睛图像样本；

所述图像获取模块还用于：在得到所述人脸图像之后，裁剪所述人脸图像中的至少一眼睛区域，得到至少一眼睛图像；

所述图像输入模块用于：将所述人脸图像和所述至少一眼睛图像同时输入所述神经网络。

40 29、根据权利要求 28 所述的装置，其中，

所述图像输入模块用于：将所述人脸图像和所述至少一眼睛图像调整到相同的预定尺寸后将它们同时输入所述神经网络；

所述类别检测模块用于：经所述神经网络同时提取所述人脸图像中的特征和所述至少一眼睛图像中的特征，并根据提取的特征确定所述人脸图像对应的注视区域检测类别。

30、根据权利要求 28 所述的装置，其中，

5 所述图像输入模块用于：将所述人脸图像和所述至少一眼睛图像分别输入所述神经网络的不同的特征提取分支，其中，所述人脸图像和所述至少一眼睛图像的尺寸不同；

所述类别检测模块用于：经所述神经网络的相应的特征提取分支分别提取所述人脸图像中的特征和所述至少一眼睛图像中的特征；融合所述神经网络的相应的特征提取分支分别提取的各特征，得到融合特征；根据所述融合特征确定所述人脸图像对应的注视区域检测类别。

31、根据权利要求 27~30 中任一所述的装置，其中，

10 所述图像获取模块用于：获取通过在所述指定空间区域部署的多个摄像头，在同一时刻针对该指定空间区域的一特定子区域从不同角度分别采集的多个图像；根据图像质量评价指标，确定所述多个图像中图像质量评分最高的图像；截取所述图像质量评分最高的图像中的人脸区域，得到所述人脸图像。

32、根据权利要求 27~30 中任一所述的装置，其中，

15 所述图像获取模块用于：获取通过在所述指定空间区域部署的多个摄像头，在同一时刻针对该指定空间区域的一特定子区域从不同角度分别采集的多个图像；根据图像质量评价指标，确定所述多个图像中图像质量评分最高的图像；分别截取所述多个图像中的人脸区域，得到相应的多个人脸图像；

所述图像输入模块用于：分别将该多个人脸图像输入所述神经网络；

20 所述类别检测模块用于：针对该多个人脸图像中的每一个人脸图像，经所述神经网络对该人脸图像进行特征提取，并根据提取的特征确定该人脸图像对应的注视区域检测类别；从所确定的分别与该多个人脸图像对应的多个注视区域检测类别中，选择与所述图像质量评分最高的图像关联的人脸图像对应的注视区域检测类别，作为在所述时刻的注视区域检测类别。

33、根据权利要求 31 或 32 所述的装置，其中，所述图像质量评价指标包括下列中至少一种：图像中是否包括有眼睛图像、图像中眼睛区域的清晰度、图像中眼睛区域的遮挡情况、图像中眼睛的睁/闭情况。

25 34、根据权利要求 27~30 中任一所述的装置，其中，

所述图像获取模块用于：获取通过在所述指定空间区域部署的多个摄像头，在同一时刻针对该指定空间区域的一特定子区域从不同角度分别采集的多个图像；分别截取所述多个图像中的人脸区域，得到相应的多个人脸图像；

所述图像输入模块用于：分别将该多个人脸图像输入所述神经网络；

30 所述类别检测模块用于：针对该多个人脸图像中的每一个人脸图像，经所述神经网络对该人脸图像进行特征提取，并根据提取的特征确定该人脸图像对应的注视区域检测类别；选择所确定的分别与该多个人脸图像对应的多个注视区域检测类别中的多数结果，作为在所述时刻的注视区域检测类别。

35、根据权利要求 27~34 中任一所述的装置，其中，所述指定空间区域包括：车的空间区域。

36、根据权利要求 35 所述的装置，其中，

35 所述在指定空间区域内采集到的图像包括：针对所述车的空间区域中的驾驶区域采集到的图像；

所述多类定义注视区域包括下列中至少两类：左前挡风玻璃区域、右前挡风玻璃区域、仪表盘区域、车内后视镜区域、中控台区域、左后视镜区域、右后视镜区域、遮阳板区域、换挡杆区域、方向盘下方区域、副驾驶区域、副驾驶前方的杂物箱区域。

37、根据权利要求 27~36 中任一所述的装置，所述装置还包括：

40 第一类别应用模块，用于：基于所述类别检测模块得到的所述注视区域检测类别而得到注视区域类别检测结果，并根据该注视区域类别检测结果，确定所述人脸图像对应的人的注意力监控结果；输出所述注意力监控结果，和/或，根据所述注意力监控结果输出分心提示信息。

38、根据权利要求 27~36 中任一所述的装置，所述装置还包括：

第二类应用模块，用于：基于所述类别检测模块得到的所述注视区域检测类别而得到注视区域类别检测结果，并确定与该注视区域类别检测结果对应的控制指令；控制电子设备执行与所述控制指令相应的操作。

5 39、一种注视区域检测用神经网络的训练设备，包括存储器和处理器，其中，所述存储器中存储有可由所述处理器执行的计算机指令，所述处理器在执行所述计算机指令时，实现根据权利要求 1 至 7 中任一所述的方法。

40、一种注视区域检测设备，包括存储器和处理器，其中，所述存储器中存储有可由所述处理器执行的计算机指令，所述处理器在执行所述计算机指令时，实现根据权利要求 8 至 19 中任一所述的方法。

10 41、一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时，使该处理器实现根据权利要求 1 至 7 中任一所述的方法。

42、一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时，使该处理器实现根据权利要求 8 至 19 中任一所述的方法。

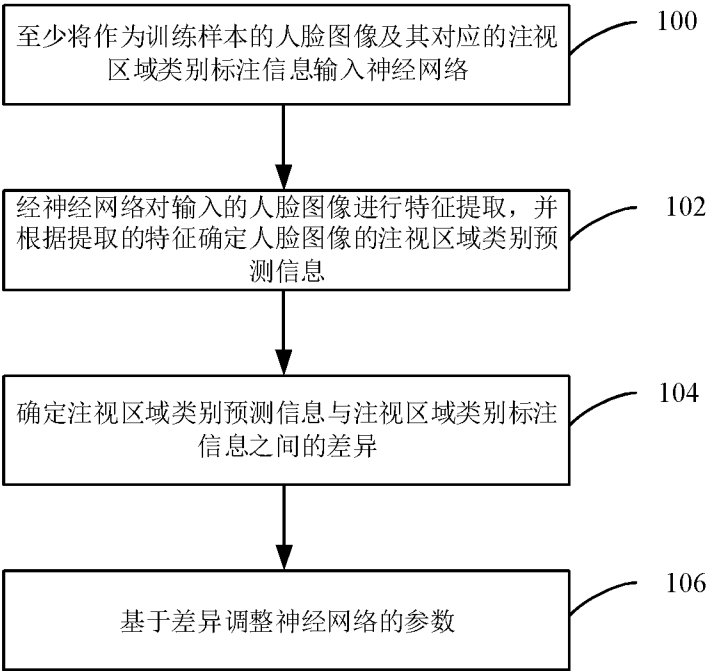


图 1

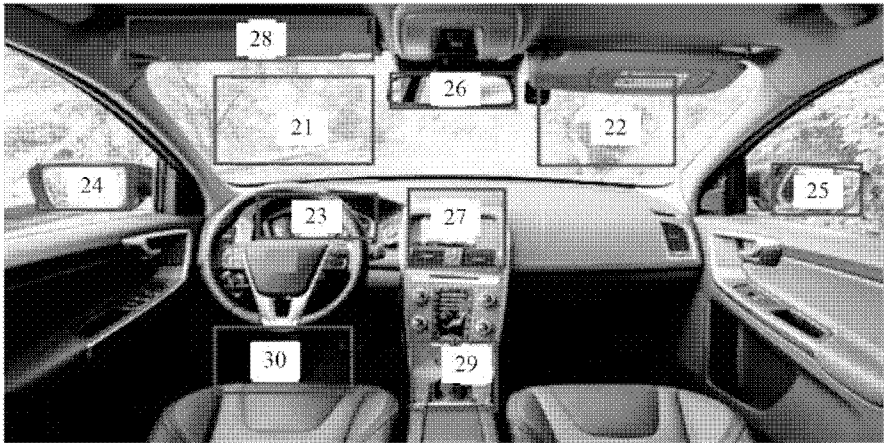


图 2

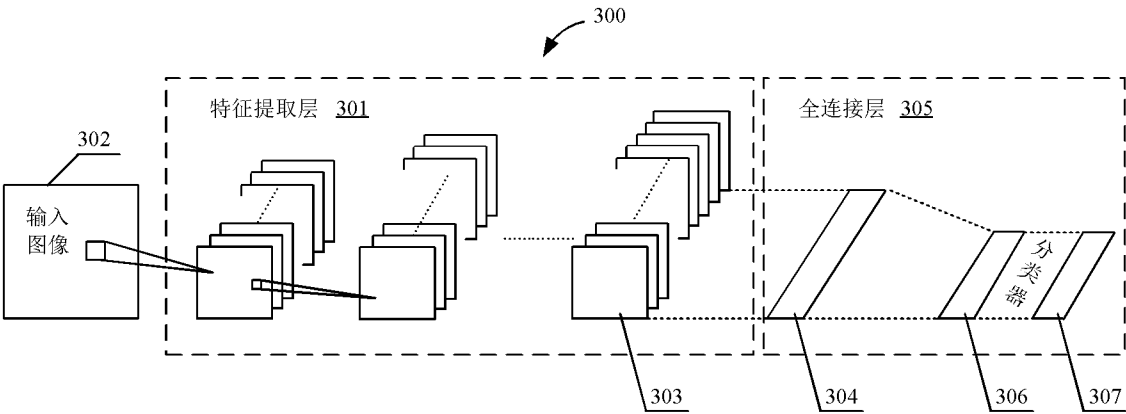


图 3

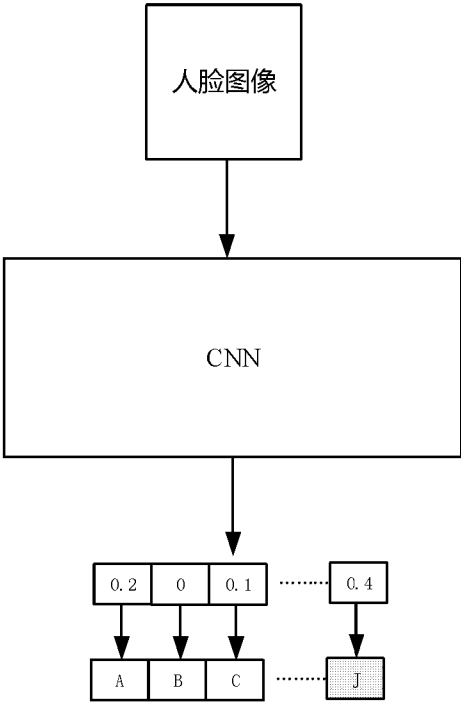


图 4

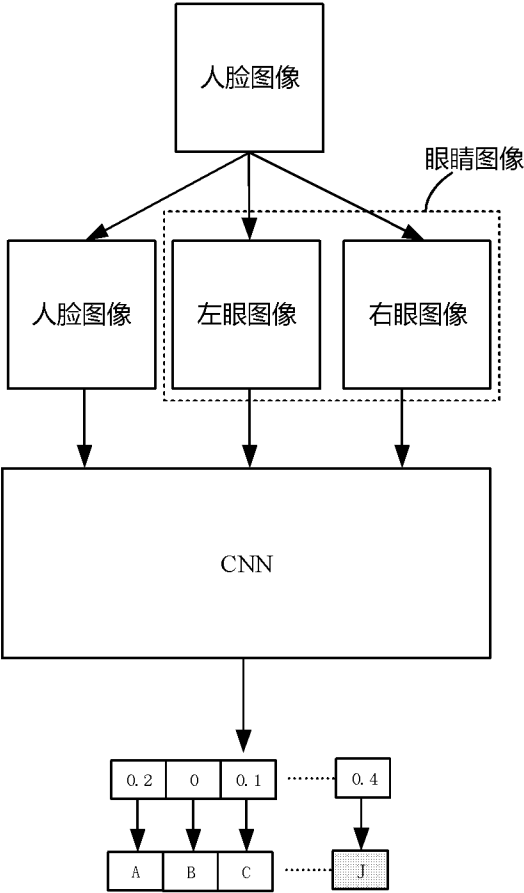


图 5

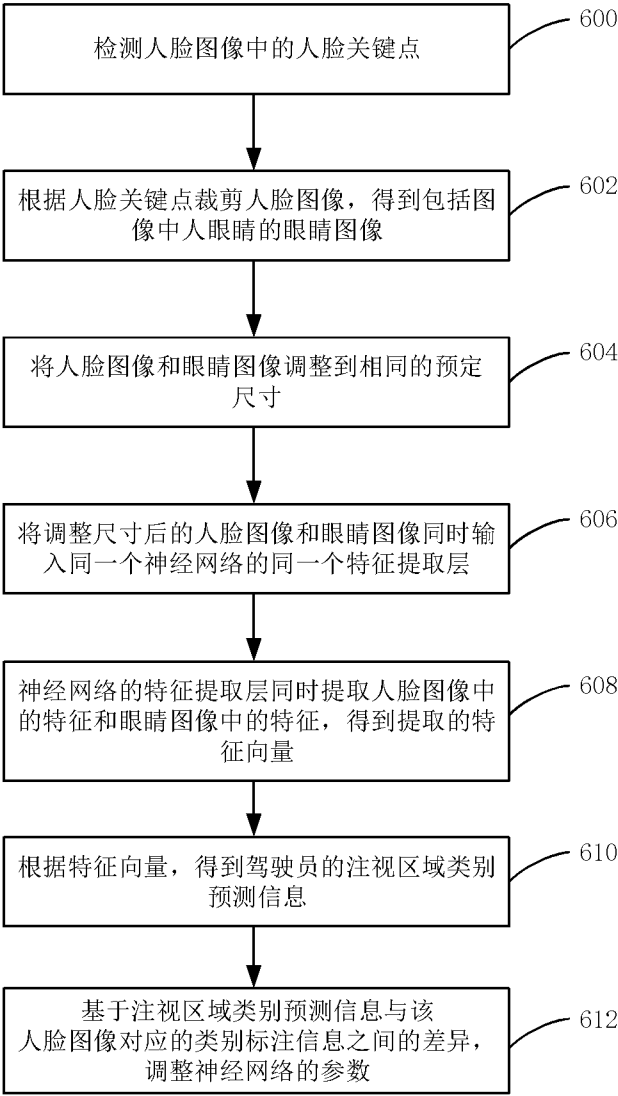


图 6

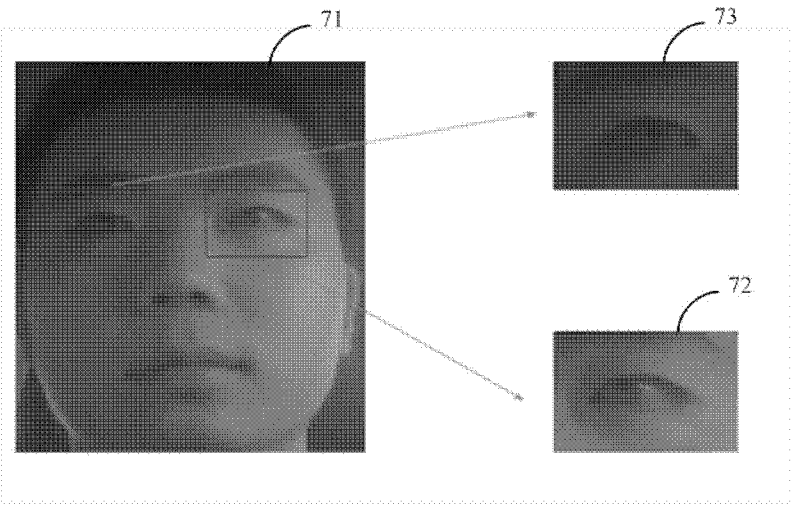


图 7

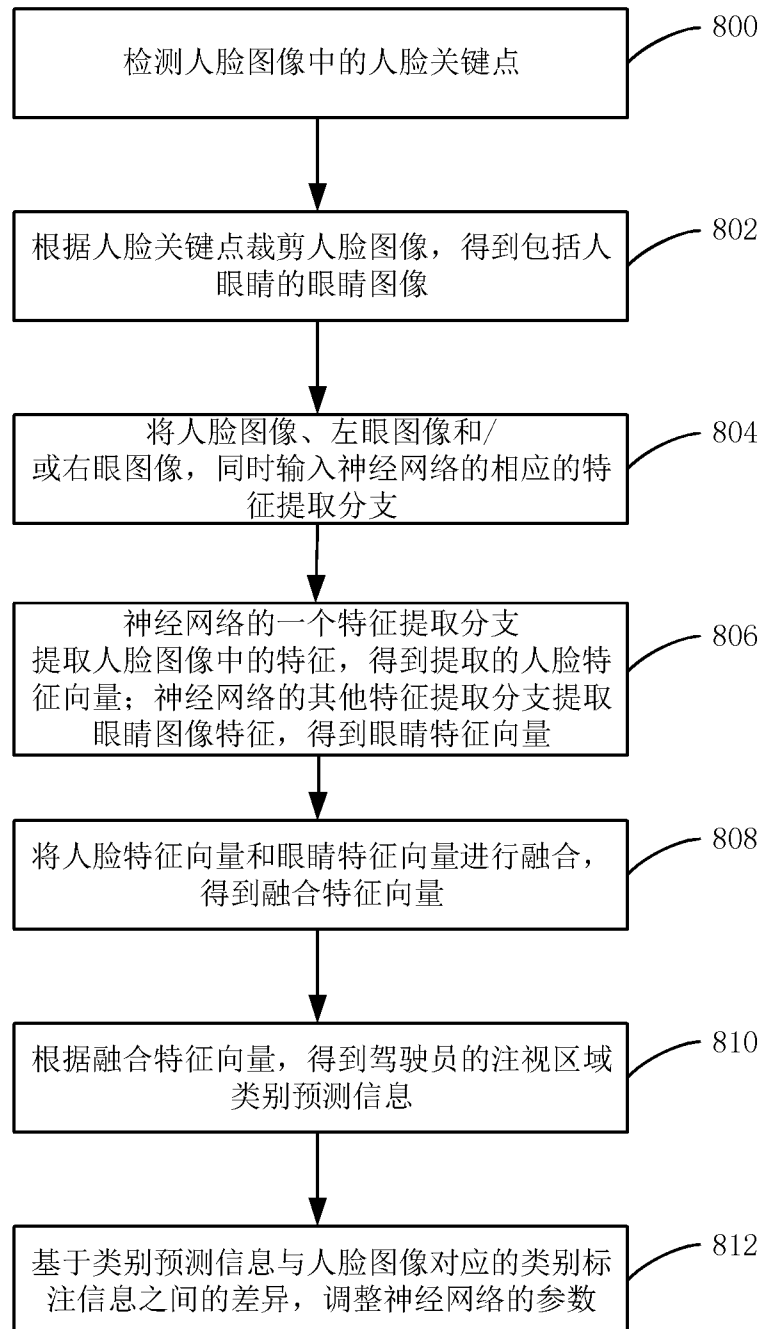


图 8

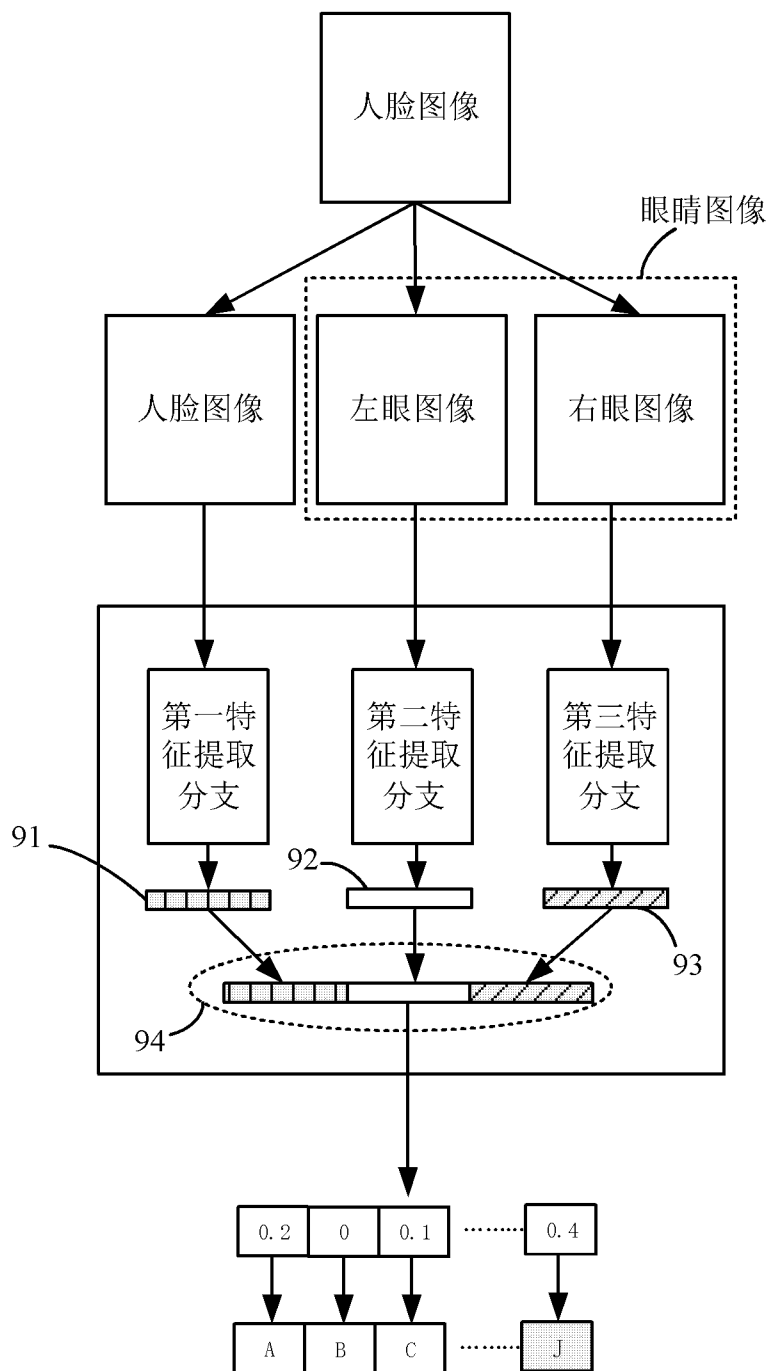


图 9

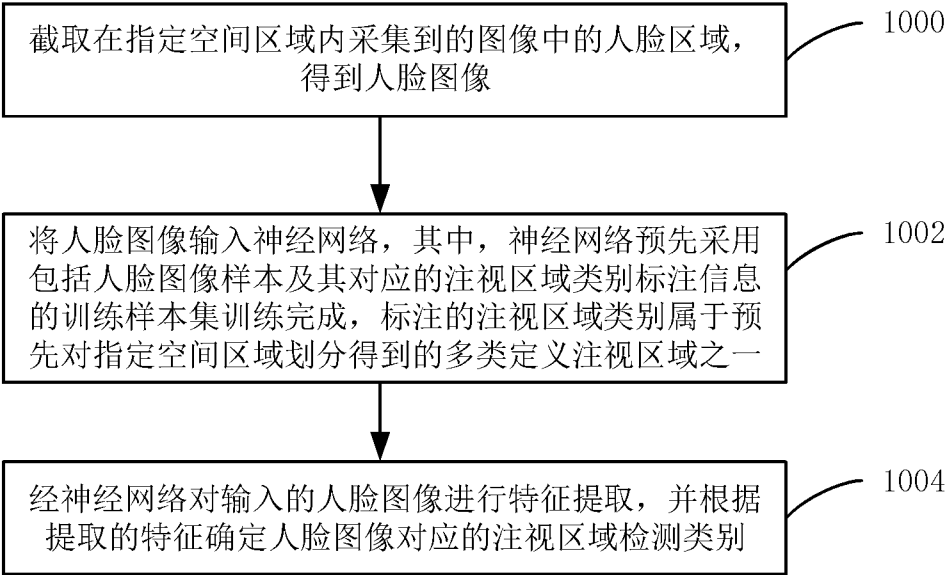


图 10

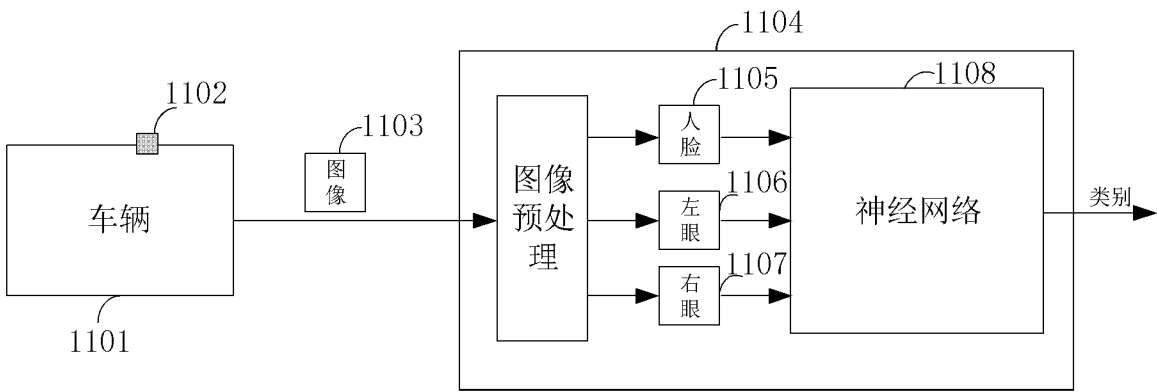


图 11

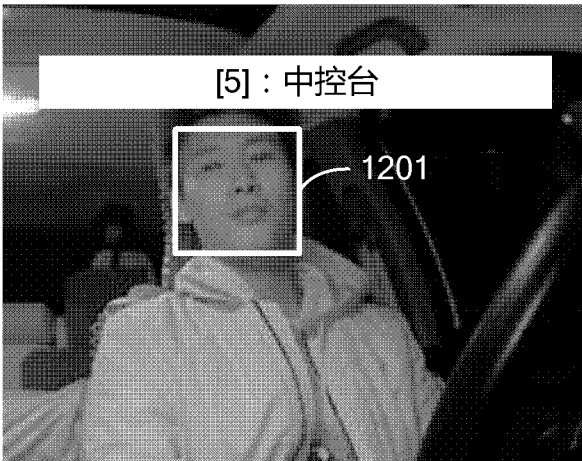


图 12

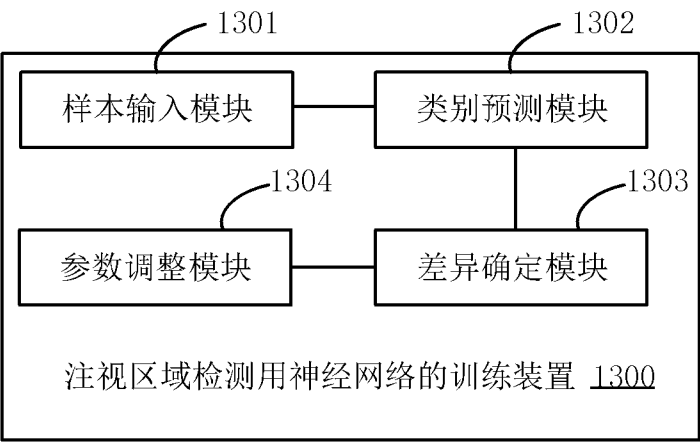


图 13

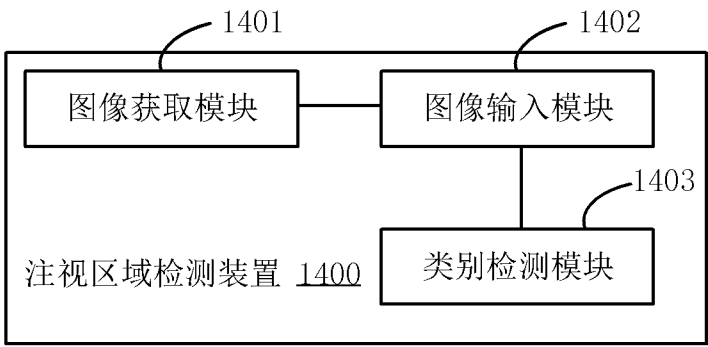


图 14

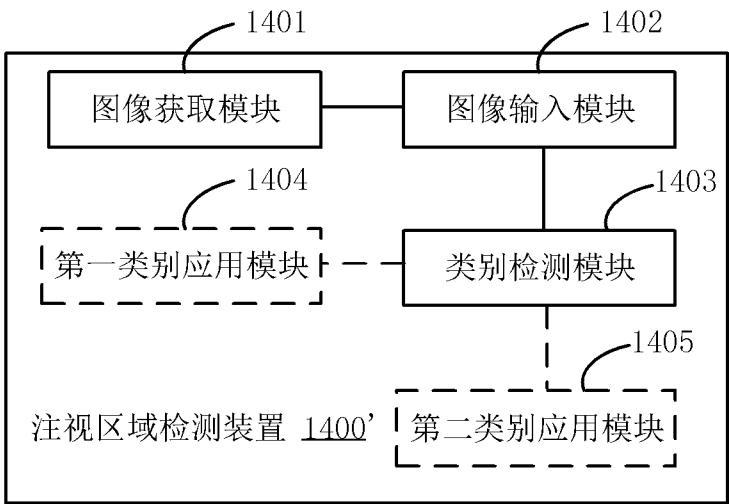


图 15

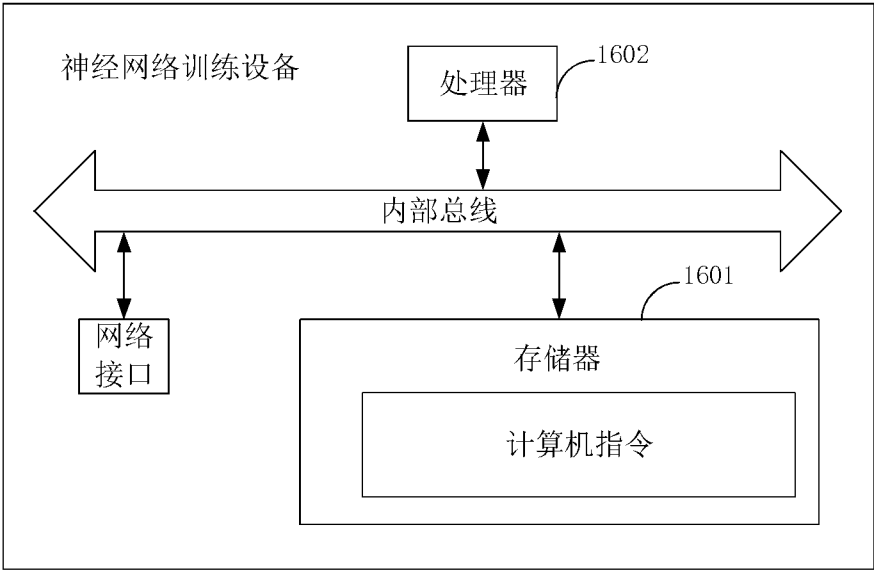


图 16

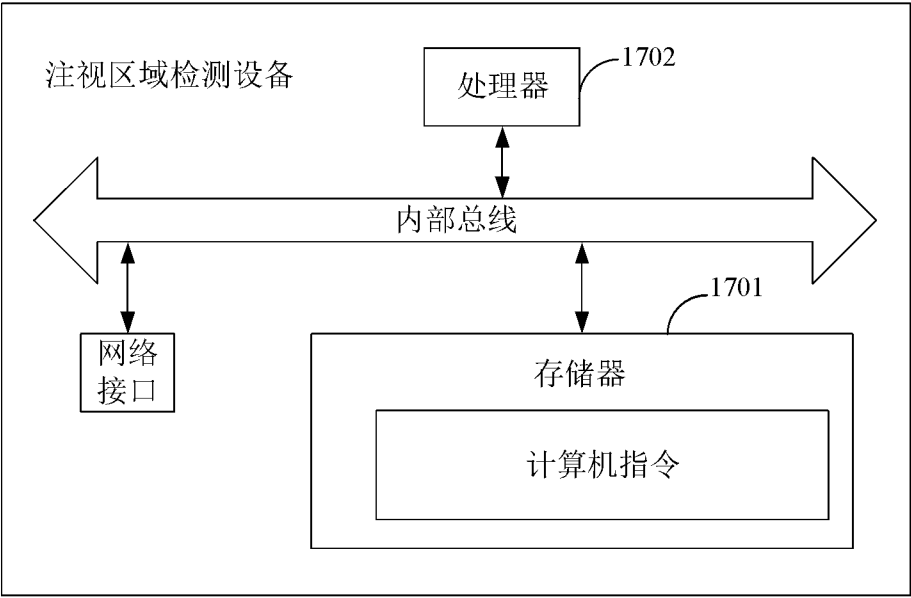


图 17

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/129893

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/00(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

USTXT; DWPI; SIPOABS; CNABS; CNKI; CNTXT: 神经网络, 人脸图像, 眼睛图像, 注视区域, neural network, sight region, face image, attention region, eye image

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 108229284 A (BEIJING SENSETIME TECHNOLOGY DEVELOPMENT CO., LTD.) 29 June 2018 (2018-06-29) description, paragraphs 176-350	1-42
A	WO 2018167991 A1 (OMRON CORPORATION) 20 September 2018 (2018-09-20) entire document	1-42
A	CN 109446892 A (UNIVERSAL UBIQUITOUS TECHNOLOGY CO., LTD.) 08 March 2019 (2019-03-08) entire document	1-42
A	CN 107697069 A (SAIC MOTOR CORPORATION LIMITED) 16 February 2018 (2018-02-16) entire document	1-42
A	CN 109460780 A (DEEPBLUE TECHNOLOGY (SHANGHAI) CO., LTD.) 12 March 2019 (2019-03-12) entire document	1-42



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

26 February 2020

Date of mailing of the international search report

27 March 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/129893

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)		Publication date (day/month/year)
CN	108229284	A	29 June 2018	None		
WO	2018167991	A1	20 September 2018	JP	6264492 B1	24 January 2018
				JP	2018152038 A	27 September 2018
				JP	6264494 B1	24 January 2018
				US	2019370580 A1	05 December 2019
				JP	6264495 B1	24 January 2018
				JP	2018152034 A	27 September 2018
				WO	2018168040 A1	20 September 2018
				WO	2018168039 A1	20 September 2018
				JP	2018152037 A	27 September 2018
				CN	110268456 A	20 September 2019
				DE	112017007252 T5	19 December 2019
CN	109446892	A	08 March 2019	None		
CN	107697069	A	16 February 2018	None		
CN	109460780	A	12 March 2019	None		

国际检索报告

国际申请号

PCT/CN2019/129893

A. 主题的分类 G06K 9/00 (2006.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类																				
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06K 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) USTXT; DWPI; SIPOABS; CNABS; CNKI; CNTXT: 神经网络, 人脸图像, 眼睛图像, 注视区域, neural network, sight region, face image, attention region, eye image																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 108229284 A (北京市商汤科技开发有限公司) 2018年 6月 29日 (2018 - 06 - 29) 说明书第176-350段</td> <td>1-42</td> </tr> <tr> <td>A</td> <td>W0 2018167991 A1 (欧姆龙株式会社) 2018年 9月 20日 (2018 - 09 - 20) 全文</td> <td>1-42</td> </tr> <tr> <td>A</td> <td>CN 109446892 A (杭州宇泛智能科技有限公司) 2019年 3月 8日 (2019 - 03 - 08) 全文</td> <td>1-42</td> </tr> <tr> <td>A</td> <td>CN 107697069 A (上海汽车集团股份有限公司) 2018年 2月 16日 (2018 - 02 - 16) 全文</td> <td>1-42</td> </tr> <tr> <td>A</td> <td>CN 109460780 A (深兰科技上海有限公司) 2019年 3月 12日 (2019 - 03 - 12) 全文</td> <td>1-42</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 108229284 A (北京市商汤科技开发有限公司) 2018年 6月 29日 (2018 - 06 - 29) 说明书第176-350段	1-42	A	W0 2018167991 A1 (欧姆龙株式会社) 2018年 9月 20日 (2018 - 09 - 20) 全文	1-42	A	CN 109446892 A (杭州宇泛智能科技有限公司) 2019年 3月 8日 (2019 - 03 - 08) 全文	1-42	A	CN 107697069 A (上海汽车集团股份有限公司) 2018年 2月 16日 (2018 - 02 - 16) 全文	1-42	A	CN 109460780 A (深兰科技上海有限公司) 2019年 3月 12日 (2019 - 03 - 12) 全文	1-42
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
X	CN 108229284 A (北京市商汤科技开发有限公司) 2018年 6月 29日 (2018 - 06 - 29) 说明书第176-350段	1-42																		
A	W0 2018167991 A1 (欧姆龙株式会社) 2018年 9月 20日 (2018 - 09 - 20) 全文	1-42																		
A	CN 109446892 A (杭州宇泛智能科技有限公司) 2019年 3月 8日 (2019 - 03 - 08) 全文	1-42																		
A	CN 107697069 A (上海汽车集团股份有限公司) 2018年 2月 16日 (2018 - 02 - 16) 全文	1-42																		
A	CN 109460780 A (深兰科技上海有限公司) 2019年 3月 12日 (2019 - 03 - 12) 全文	1-42																		
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																				
<table border="0"> <tr> <td> * 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td> “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																			
国际检索实际完成的日期 2020年 2月 26日		国际检索报告邮寄日期 2020年 3月 27日																		
ISA/CN的名称和邮寄地址 中国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 王艳坤 电话号码 86-(10)-62412330																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/129893

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	108229284	A	2018年 6月 29日	无	
WO	2018167991	A1	2018年 9月 20日	JP	6264492 B1 2018年 1月 24日
				JP	2018152038 A 2018年 9月 27日
				JP	6264494 B1 2018年 1月 24日
				US	2019370580 A1 2019年 12月 5日
				JP	6264495 B1 2018年 1月 24日
				JP	2018152034 A 2018年 9月 27日
				WO	2018168040 A1 2018年 9月 20日
				WO	2018168039 A1 2018年 9月 20日
				JP	2018152037 A 2018年 9月 27日
				CN	110268456 A 2019年 9月 20日
				DE	112017007252 T5 2019年 12月 19日
CN	109446892	A	2019年 3月 8日	无	
CN	107697069	A	2018年 2月 16日	无	
CN	109460780	A	2019年 3月 12日	无	