(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 12 月 30 日 (30.12.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/258077 A1

(51) 国际专利分类号:
G06K 9/62 (2006.01) *G06N 3/04* (2006.01)

(21) 国际申请号: PCT/CN2019/093024

(22) 国际申请日: 2019 年 6 月 26 日 (26.06.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人:深圳大学(SHENZHEN UNIVERSITY) [CN/CN]; 中国广东省深圳市南山区南海大道3688号, Guangdong 518000 (CN)。

(72) 发明人:李国法(LI, Guofa); 中国广东省深圳市南山区南海大道3688号, Guangdong 518000 (CN)。 杨一帆(YANG, Yifan); 中国广东省深圳市南山区南海大道3688号, Guangdong 518000 (CN)。 陈耀昱(CHEN, Yaoyu); 中国广东省深圳市南山区南海大道3688号, Guangdong 518000 (CN)。 谢恒(XIE, Heng); 中国广东省深圳市南山区南海大道3688号, Guangdong 518000 (CN)。 李盛龙(LI, Shenglong); 中国广东省深圳市南山区南海大道3688号, Guangdong 518000 (CN)。 赖伟鉴(LAI, Weijian); 中国广东省深圳市南山区南海大道3688号, Guangdong 518000 (CN)。 李晓航(LI, Xiaohang); 中国广东省深圳市南山区南海大道3688号, Guangdong 518000 (CN)。 朱方平(ZHU, Fangping); 中国广东省深圳市南山区南海大道3688号, Guangdong 518000 (CN)。 颜伟荃(YAN, Weiquan); 中国广东省深圳市南山区南海大道3688号, Guangdong 518000 (CN)。

(74) 代理人:深圳中一联合知识产权代理有限公司(SHENZHEN ZHONGYI UNION INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国广东省深

(54) **Title:** PEDESTRIAN DETECTION METHOD AND DEVICE

(54) **发明名称**: 一种行人检测方法及装置

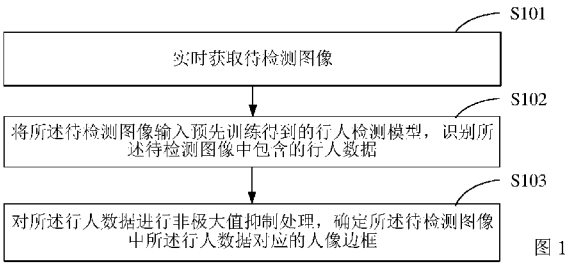


图 1

S101 Obtain an image to be detected in real time
S102 Input said image to a pedestrian detection model obtained by pre-training, and identify pedestrian data comprised in said image
S103 Perform non-maximum suppression processing on the pedestrian data and determine a portrait frame corresponding to the pedestrian data in said image

(57) **Abstract:** The present application is applicable for the technical field of computer applications, and provides a pedestrian detection method and device. The method comprises: obtaining an image to be detected in real time; inputting said image to a pedestrian detection model obtained by pre-training, and identifying pedestrian data comprised in said image; and performing non-maximum suppression processing on the pedestrian data and determining a portrait frame corresponding to the pedestrian data in said image. In this embodiment, the pedestrian detection model is obtained by training on the basis of a depthwise separable convolution mode, the obtained image to be detected is identified according to the pedestrian detection model, and the portrait frame corresponding to the pedestrian comprised in said image is determined, which not only improves the efficiency of portrait detection, so that a user can determine a corresponding processing mode according to the detected portrait frame at the first time, but also improves the accuracy of portrait detection, ensuring that the current pedestrian situation can be clearly detected even in the low-visibility environment such as haze.

(57) **摘要**: 本申请适用于计算机应用技术领域，提供了行人检测方法及装置，包括：实时获取待检测图像；将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据；对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框。本实施例中通过基于深度可分离卷积方式训练得到行人检测模型，根据行人检测模型对获取到的待检测图像进行识别，确定其中包含的行人对应的人像边框，不仅提高了人像检测的效率，使得用户可以第一时间根据检测到的人像边框确定对应的处理方式，也提高了人像检测的精确度，保证了在雾霾等能见度较低的环境下也能清楚检测出当前的行人情况。

WO 2020/258077 A1

[见续页]

圳市福田区园岭街道深南中路1014号报春大厦9楼（5号信箱）, Guangdong 518028 (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。

(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))。

一种行人检测方法及装置

技术领域

本申请属于计算机应用技术领域，尤其涉及一种行人检测方法及装置。

背景技术

步行是交通出行的基本方式之一，据调查显示，在欧洲区域，每年有超过 7000 名行人死亡，占所有死亡人数的百分之二十七，因此，有效检测各种环境下的行人，将显著提高自动驾驶车辆的行驶安全性。然而，由于行人姿态、定位、服装和天气条件的多样性和复杂性，行人检测问题依然存在。

现有技术中大多数的检测模型只在光照充足的条件下进行了测试，一般而言，它们并没有能力在光照不足的情况下检测行人，比如在雾天，因为恶劣天气的使能见度降低，色彩反射不足，并造成行人轮廓和外观模糊，很难将其与背景区分开来。因此，现有技术中在环境较模糊的情况下，对行人进行检测时很难将其与背景区分开来，而造成行人检测结果不精确的问题。

技术问题

有鉴于此，本申请实施例提供了行人检测方法及装置，以解决现有技术中行人检测结果不精确的问题。

技术解决方案

本申请实施例的第一方面提供了一种行人检测方法，包括：

实时获取待检测图像；

将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据；所述行人检测模型根据预设的深度可分离卷积方式训练得到；

对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框。

本申请实施例的第二方面提供了一种行人检测装置，包括：

获取单元，用于实时获取待检测图像；

识别单元，用于将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据；所述行人检测模型根据预设的深度可分离卷积方式训练得到；

确定单元，用于对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框。

本申请实施例的第三方面提供了一种行人检测装置，包括：处理器、输入设备、输出设备和存储器，所述处理器、输入设备、输出设备和存储器相互连接，其中，所述存储器用于存储支持装置执行上述方法的计算机程序，所述计算机程序包括程序指令，所述处理器被配置用于调用所述程序指令，执行上述第一方面的方法。

本申请实施例的第四方面提供了一种计算机可读存储介质，所述计算机存储介质存储

有计算机程序，所述计算机程序包括程序指令，所述程序指令当被处理器执行时使所述处理器执行上述第一方面的方法。

有益效果

本申请实施例与现有技术相比存在的有益效果是：实时获取待检测图像；将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据；对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框。本实施例中通过基于深度可分离卷积方式训练得到行人检测模型，根据行人检测模型对获取到的待检测图像进行识别，确定其中包含的行人对应的人像边框，不仅提高了人像检测的效率，使得用户可以第一时间根据检测到的人像边框确定对应的处理方式，也提高了人像检测的精确度，保证了在雾霾等能见度较低的环境下也能清楚检测出当前的行人情况。

附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 1 是本申请实施例一提供的行人检测方法的流程图；

图 2 是本申请实施例二提供的行人检测方法的流程图；

图 3 是本申请实施例二提供的图像增强技术的应用示例；

图 4 是本申请实施例二提供的训练模型的训练过程和应用过程示意图；

图 5 是本申请实施例二提供的行人检测方法的结构示意图；

图 6 是本申请实施例二提供的标准卷积与深度可分离卷积的对比示意图；

图 7 是本申请实施例二提供的行人检测方法中瓶颈层结构示意图；

图 8 是本申请实施例二提供的权重连结层的结构示意图；

图 9 是本申请实施例二提供的压缩激励机制的示意图；

图 10 是本申请实施例二提供的历史图像的先验框的分布图；

图 11 是本申请实施例二提供的 MNPB-YOLO 标签示意图；

图 12 是本申请实施例二提供的检测结果示例；

图 13 是本申请实施例三提供的行人检测装置的示意图；

图 14 是本申请实施例四提供的行人检测装置的示意图。

本发明的具体实施方式

以下描述中，为了说明而不是为了限定，提出了诸如特定系统结构、技术之类的具体细节，以便透彻理解本申请实施例。然而，本领域的技术人员应当清楚，在没有这些具体细节的其它实施例中也可以实现本申请。在其它情况中，省略对众所周知的系统、装置、电路以及方法的详细说明，以免不必要的细节妨碍本申请的描述。

为了说明本申请所述的技术方案，下面通过具体实施例来进行说明。

实施例 1：

参见图 1，图 1 是本申请实施例一提供的一种行人检测方法的流程图。本实施例中行人检测方法的执行主体为具有行人检测功能的装置，包括但不限于计算机、服务器、平板电脑或者终端等装置。如图所示的行人检测方法可以包括以下步骤：

S101：实时获取待检测图像。

步行是交通出行的基本方式之一，据调查显示，在欧洲区域，每年有超过 7000 名行人死亡，占所有死亡人数的百分之二十七，因此，有效检测各种环境下的行人，将显著提高自动驾驶车辆的行驶安全性。然而，由于行人姿态、定位、服装和天气条件的多样性和复杂性，行人检测问题依然存在。随着计算机性能的不断提高，基于深度体系结构的检测方法也得到了广泛的应用。然而，大多数的检测模型只在光照充足的条件下进行了测试，一般而言，它们并没有能力在光照不足的情况下检测行人，比如在雾天。据调查显示，雾霾天气下的交通事故发生率要远高于晴天条件下，因为在雾霾天气下，人眼能感受到的视野范围变小，使得司机难以看清路牌和行人，从而增加了交通事故发生的概率。一般来说，雾霾程度越严重，事故发生率越高。然而，在雾霾中检测行人是一项尤其挑战性的任务，因为恶劣天气的使能见度降低，色彩反射不足，并造成行人轮廓和外观模糊，很难将其与背景区分开来。这些算法主要应用于大气光均匀的白天场景。然而，更为恶劣的雾霾情况通常发生在昏暗的光线下，这使得现有的去雾算法无法适应于更为恶劣的雾霾场景。

本实施例中在检测行人过程中，先实时获取待检测图像。本实施例中的待检测图像可以是单幅图像的形式，也可以是一段视频的形式，当获取到视频之后，可以对视频进行预设周期内的采样处理，得到单帧的图像画面作为待检测图像。同时，本实施例中的待检测图像可以是彩色图像、黑白图像或者红外图像等，此处不做限定。

示例性地，在车辆行进过程中，可以通过安装在车辆前端的摄像装置，例如行车记录仪，来实时拍摄图像或者视频，以将该图像作为待检测图像，对视频进行采样处理，得到单帧的图像画面作为待检测图像。

S102：将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据；所述行人检测模型根据预设的深度可分离卷积方式训练得到。

本实施例中预先训练有行人检测模型，本实施例利用深度可分离卷积和线性瓶颈层技术降低了计算量和参数个数，提高了网络的运行效率。此外，我们还创新性地将多尺度特征融合与压缩激励机制相结合，提出了一种新的特征融合方法——权重连接层。利用上述方法，我们提出了一种高效的在雾霾天气下的行人检测方法 MNPB-YOLO 行人检测模型。

在训练得到行人检测模型之后，将获取到的待检测图像输入该行人检测模型，识别得到待检测图像中包含的行人数据。本实施例中的行人数据可以包括行人在待处理图像中的位置、对应的像素点的位置，以及通过行人检测模型找出了一堆行人对应的方框等，此处不做限定。

S103：对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框。

在通过行人检测模型得到待处理图像中的行人数据之后，对行人数据进行非极大值抑制处理，以确定待检测图像中行人数据对应的人像边框。本实施例中通过非极大值抑制抑

制不是极大值的元素，例如，通过局部最大搜索的方式，确定图像中对应区域的边缘。这个局部用于表示的是一个邻域，邻域有两个参数可变，一是邻域的维数，二是邻域的大小。在本实施例的行人检测中，通过滑动窗口经提取特征，经分类器分类识别后，每个窗口都会得到一个分数。但是滑动窗口会导致很多窗口与其他窗口存在包含或者大部分交叉的情况。这时就需要用到非极大值抑制处理来选取那些邻域里分数最高，即是行人的概率最大，不是行人的概率较小，并且抑制那些分数低的窗口。

具体的，对于识别出的待检测图像中各个行人数据对应的边界框，我们需要判别哪些矩形框是没用的。本实施例中假设每个行人对应的边界框的集合为 B，选择边界框最大的为检测框 M，将其从 B 集合中移除并加入到最终的检测结果中，计算 B 中每个剩余的边界框与检测框 M 之间的重叠度（Intersection over Union，IOU），将 IOU 大于或者等于预设的重叠度阈值的边界框框从 B 中移除，重复这个过程，直到 B 为空，最后将被保留下来的边界框作为待检测图像中行人数据对应的人像边框。

示例性地，在待检测图像定位一个行人时，通过行人检测模型找出了一堆的方框，我们需要判别哪些矩形框是没用的。本实施例中通过非极大值抑制的方法是：先假设有预设数量的矩形框，即边界框，根据分类器的类别分类概率做排序。假设矩形框的数量为 6 个，从小到大属于行人的概率分别为 A、B、C、D、E、F。从最大概率矩形框 F 开始，分别判断 A~E 与 F 的 IOU 是否大于某个设定的重叠度阈值，假设 B、D 与 F 的重叠度超过重叠度阈值，那么就扔掉 B、D；并标记第一个矩形框 F，是我们保留下来的。从剩下的矩形框 A、C、E 中，选择概率最大的 E，然后判断 E 与 A、C 的重叠度，重叠度大于一定的重叠度阈值，那么就扔掉；并标记 E 是我们保留下来的第二个矩形框，就这样一直重复，找到所有被保留下来的矩形框，作为待检测图像中所述行人数据对应的人像边框。

上述方案，通过实时获取待检测图像；将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据；所述行人检测模型根据预设的深度可分离卷积方式训练得到；对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框。本实施例中通过基于深度可分离卷积方式训练得到行人检测模型，根据行人检测模型对获取到的待检测图像进行识别，确定其中包含的行人对应的人像边框，不仅提高了人像检测的效率，使得用户可以第一时间根据检测到的人像边框确定对应的处理方式，也提高了人像检测的精确度，保证了在雾霾等能见度较低的环境下也能清楚检测出当前的行人情况。

实施例 2：

参见图 2，图 2 是本申请实施例二提供的一种行人检测方法的流程图。本实施例中行人检测方法的执行主体为具有行人检测功能的装置，包括但不限于计算机、服务器、平板电脑或者终端等装置。如图所示的行人检测方法可以包括以下步骤：

S201：实时获取待检测图像。

本实施例中在检测行人过程中，先实时获取待检测图像。本实施例中的待检测图像可以是单幅图像的形式，也可以是一段视频的形式，当获取到视频之后，可以对视频进行预设周期内的采样处理，得到单帧的图像画面作为待检测图像。同时，本实施例中的待检测图像可以是彩色图像、黑白图像或者红外图像等，此处不做限定。示例性地，在车辆行进

过程中，可以通过安装在车辆前端的摄像装置，例如行车记录仪，来实时拍摄图像或者视频，以将该图像作为待检测图像，对视频进行采样处理，得到单帧的图像画面作为待检测图像。

S202：获取包含行人的历史图像。

本实施例中在对待检测图像进行识别时，是根据预先训练好的行人检测模型进行识别。因此，在识别获取到的待检测图像之前，先获取包含行人的历史图像，以根据历史图像训练得到行人检测模型。

本实施例中的历史图像的形式和待检测图像的形式相同，本实施例中的历史图像可以是彩色图像、黑白图像或者红外图像等，此处不做限定。本实施例中的历史图像可以是单幅图像的形式，也可以是一段视频的形式，当获取到视频之后，可以对视频进行预设周期内的采样处理，得到单帧的图像画面作为历史图像。

S203：根据预设的权重连接层构建训练模型，根据所述历史图像对所述训练模型进行训练，得到所述行人检测模型。

在获取到历史图像之后，根据预设的权重连接层构建训练模型，并根据历史图像对训练模型进行训练，得到行人检测模型来检测时待检测图像中的行人。需要说明的是，为了区别不同处理方式对应的训练模型，本实施例中的训练模型包括第一训练模型和第二训练模型，其中，第一训练模型用于表示对历史图像进行扩充之后得到的训练模型，第二训练模型用于表示通过步骤 S2031~S2034 训练得到的 S2031~S2034，两个训练模型可以单独实现训练和图像识别，也可以将两个训练模型结合起来进行训练和图像识别。

进一步的，步骤 S203 可以具体包括：

对所述历史图像进行图像增强处理，得到所述历史图像对应的至少两个扩充图像；

根据预设的权重连接层构建第一训练模型，根据所述历史图像及其对应的所述扩充图像对所述第一训练模型进行训练，得到所述行人检测模型。

具体的，请一并参阅图 3 所示，图 3 展示了一些图像增强技术的应用示例，其中，(a)~(f)分别为历史图像的原图、随机翻转之后的图像、随机对比度改变之后的图像、随机裁剪之后的图像、随机颜色改变之后的图像以及随机仿射变换之后的图像。为了让模型具有更好的泛化性能，我们在训练时增加了图像增强技术，通过随机的裁剪、翻转，颜色改变、仿射变换和高斯噪声等操作，来扩充原有的数据集，以保证在具有强大的数据集基础的条件下，增加模型训练的次数，提高模型训练的精度。根据预设的权重连接层构建第一训练模型，根据历史图像及其对应的扩充图像对第一训练模型进行训练，得到行人检测模型。

请一并参与图 4，图 4 为本实施例提供的训练模型的训练过程和应用过程示意图，其中，在进行训练时，先获取红绿蓝（Red Green Blue，RGB）形式的历史图像，即 RGB 图像，再对 RGB 图像进行图像扩充，例如，进行图像裁剪、图像增强等处理，再通过预先涉及的算法架构，即 MNPB-YOLO 行人检测方法对历史图像及其扩充之后的图像进行识别，并根据预设的真值标签确定方式计算识别结果与原图之间的损失函数，最后通过损失函数来对 MNPB-YOLO 行人检测方法中的模型参数进行更新，最后得到训练模型中固定的参数和权重，得到高效、精确的训练模型。

进一步的，步骤 S203 可以具体包括步骤 S2031~S2034：

S2031：根据预设的权重连接层，基于预设的深度可分离卷积方式以及预设的线性瓶颈层技术，构建第二训练模型。

请一并参与图 5，图 5 为本实施例提供的行人检测方法的结构示意图，其中，图 5 中的各个数字之间的相乘计算公式用于表示在当前参与计算的数据量。为了有效的实现雾霾天气下的行人检测，本实施例提出了一新的基于 YOLO 的深度学习方法，其中包括了基础卷积模型部分、权重连接层、检测模块和分类模块。利用深度可分离卷积和线性瓶颈层技术构成基础卷积模型，降低了计算量和参数个数，提高了网络的运行效率。此外，我们还将空间到深度变换的多尺度特征融合与压缩激励机制相结合，提出了一种新的特征融合方法——权重连接层，最后通过检测模块和分类模块完成实现待处理图像的中的行人检测。

利用上述方法，我们提出了一种高效的在雾霾天气下的行人检测方法——MNPB-YOLO，请一并参与图 6，图 6 为本实施例提供的标准卷积与深度可分离卷积的对比示意图，其中，H，W 用于表示卷积核的高和宽，M 用于表示输入特征图的通道数或卷积核的通道数，N 用于表示卷积核个数。深度可分离卷积与普通卷积的效果差别主要是其可以有效的减少网络参数量和计算量，为让 MNPB-YOLO 可以更高速地运行在一般的处理器上，我们采用深度可分离卷积用来搭建整体的 MNPB-YOLO 模型。其卷积方式和普通卷积的对比如图 6 所示，通过图 6 可以看到，深度可分离卷积由两部分组成，纵向 depthwise 卷积和逐点 pointwise 卷积，depthwise 卷积在不同通道上分别对特征图进行卷积，然后 pointwise 卷积在对特征图的所有通道进行卷积。

请一并参与图 7，图 7 为本实施例提供的行人检测方法中瓶颈层结构示意图。深度可分离卷积搭配 Relu 激活函数会造成一定的信息损失，为了减少这种信息损失，需要用到瓶颈层的技术。具体的，瓶颈层技术在本方案的实施过程如表 1 所示，图像先经过一个 3×3 卷积输出一个特征图，该特征图就作为瓶颈层结构中的 Input，即图 6 中所示的根据卷积步幅为 1 或 2 采用不同的方法再进行卷积，当卷积步幅为 1 时，先通过 1×1 卷积，激活函数为 Relu6 进行升维，即图 7 中的 Conv 1×1，Relu6；升维的倍数体现在表 1 中的“扩展系数”一栏，通过此步骤可以将信息更广阔的散布在特征图中，以防止 depthwise 卷积（激活函数为 Relu6 时）造成的信息损失）；将升维后的特征图用 3×3 Depthwise 卷积进行降维处理，即图 7 中的 Dwise3×3，Relu6；然后通过 1×1 普通卷积、激活函数为线性函数来融合不同通道的信息，即图 7 中的 Conv 1×1，Linear；之后将输入与此部分输出做元素相加 element-wise，即图 7 中的 ADD；最后将此特征图输出到下一层，作为下一层的输入。当卷积步幅为 2 时，即图 7 中的 Stride=2，与卷积步幅 1 的区别只在于没有 element-wise 相加的操作，这是因为卷积步幅为 2 时，主要目的是对特征图高宽进行下采样，高宽缩减一倍，缩减后的特征图无法与原特征图做 element-wise 相加操作。

行人检测方法中详细的网络结构参数如表 1 所示，需要说明的是，本实施例中文字和附图中的 Conv 用于表示代表卷积；“数字×数字×数字”格式的字符均用于表示特征图的高×宽×通道数，之后不再一一赘述。

表 1：MNPB-YOLO 网络配置参数

输入大小	操作	扩展系数	卷积核数量	重复次数	步幅	多尺度融合
224×224×3	3x3 卷积	-	32	1	2	-
112×112×32	瓶颈层	1	16	1	1	-
112×112×16	瓶颈层	6	24	2	2	-
56×56×24	瓶颈层	6	32	3	2	-
28×28×32	瓶颈层	6	64	4	2	√
14×14×64	瓶颈层	6	96	3	1	-
14×14×96	瓶颈层	6	160	3	2	√
7×7×160	瓶颈层	6	320	3	1	-
7×7×320	1×1 卷积	-	1280	1	1	-
7×7×1280	3×3 卷积	-	-	1	1	√
权重连结层						
7×7×768	3×3+1×1 卷积	-	512	1	1	-
7×7×512	3×3+1×1 卷积	-	256	1	1	-
7×7×256	3×3+1×1 卷积	-	128	1	1	-
7×7×128	3×3+1×1 卷积	-	64	1	1	-
7×7×64	3×3+1×1 卷积	-	32	1	1	-
7×7×32	3×3+1×1 卷积	-	12	1	1	-
输出:7×7×12						

请一并参与图 8，图 8 为本实施例提供的权重连结层的结构示意图。权重连结层能自动筛选来自于不同特征尺度的特征图的重要性，然后过滤掉那些不重要的信息，以此来提升网络的性能。首先，收集来自于不同层数、尺度不同的多个特征图的信息，这些特征图大小不一致，即特征图的高×宽×通道数不同，例如图中的 28x28x16、14x14x48 或者 7x7x320 等，不能直接拼接，所以采用分割和拼接的方法，先把不同大小的特征图调整至大小一致，即空间到深度变换，然后将所有调整后的特征图拼接在一起，一同构成多尺度的特征信息，例如图中的 7x7x256+7x7x192+7x7x320，最后由压缩激励机制模型筛选出重要的特征。

请一并参与图 9，图 9 为本实施例提供的压缩激励机制的示意图，压缩激励机制实际上是一种通道注意力机制，通过给每个通道赋以不同的权重ω来筛选出重要的特征，权重ω由学习得到，其更新的方向为损失下降的方向。如图所示，先通过 1×1 卷积进行降维处理，然后通过全局池化获取通道的特征值，之后通过全连接层计算出权重ω，再用权重ω乘以压

缩后的特征图，最后得到重新标定后的特征。如图 9 所示，H，W，C 分别用于表示某一特征图的高、宽、通道数，且带上标'也是同样的意思；F_{tr}用于表示一个 1×1 卷积的降维操作，目的是通过降维的方式来减少通道数，以减少后面步骤所需要的计算量，提高本实施例中的方法检测行人的效率；$F_{sq}(\cdot)$用于表示一个在通道上的压缩操作，在实际中采用全局平均池化代替；$F_{ex}(\cdot, W)$代表将形状为$1 \times 1 \times C$的特征信息映射成另一个形状为$1 \times 1 \times C$的特征信息，即映射后的特征信息用于表示通道的重要性系数，映射方案采用多层感知机；W 用于表示多层感知机的权重，此权重的更新方向为损失梯度下降的方向；$F_{scala}(\cdot,\cdot)$用于表示将特征图 U 与映射后的特征信息，即通道重要性系数，做通道上的乘法运算；最后得到重新标定后的特征$\tilde{X}$；X 用于表示输入的特征图。在本实施例中，X 用于表示多尺度特征信息，$\tilde{X}$为筛选后的信息（请一并参阅图 8）。

S2032：将所述历史图像输入所述第二训练模型中，检测所述历史图像中的行人图像，并确定每个行人图像对应的真值标签；所述真值标签用于表示识别出的历史图像的先验框基于真值框的变换系数。

MNPB-YOLO 检测思想受 YOLO 启发，将图像分成 N×N 的网格，每个网格预测 B 个检测框，每个检测框由描述检测框位置的参数及物体种类的编码构成。在 YOLOv1(第一版本)和 YOLOv2(第二版本中)，都采用物体中心位置位于哪个网格内，就由哪个网格内的检测框进行预测，但是在 MNPB-YOLO 中，我们先计算网格内先验框与真值框的 IOU 值，然后按 IOU 值降序对先验框排序，选取前 k 个先验框来复杂预测物体的大小。

请一并参与图 10，图 10 为本实施例提供的历史图像的先验框的分布图。我们采用聚类方法先预先获取数据集中行人大小的聚类中心，我们将行人大小分为两类，将聚类中心的参数作为先验框的高宽，然后均匀分布在图像的 N×N 网格中，黑色的粗框就是两个位于网格中心的先验框，黑色网格线用于表示 N×N 网格，在这里 N 取为 7。

请一并参与图 11，图 11 为本实施例提供的 MNPB-YOLO 标签示意图，在 MNPB-YOLO 中，每个目标人物对应一个真值标签，由 6 个参数进行描述，分别用于表示先验框向真值框 x 方向的偏置和 y 方向的偏置，先验框向真值框在高、宽上的变换系数，以及用于区分人物及背景的特征向量，以 one-hot 编码表示。在图 11 的左图中，细线表示的框为先验框，粗线表示的框为真值框，假设细线表示的框就是负责预测真值框的，那么其标签就是$[\frac{X_g - X_a}{W_a} \frac{y_g - y_a}{h_a} \log(\frac{h_g}{h_a}) \cdots \log(\frac{W_g}{W_a}) \cdots 0 \cdots 1] \in \Re^6$，其中，$w_a$，$h_a$用于表示先验框的宽和高，$x_a$，$y_a$用于表示先验框距离图像左边缘和上边缘的距离。$w_g$，$h_g$用于表示真值框的宽和高，$x_g$，$y_g$用于表示真值框距离图像左边缘和上边缘的距离。否则，则其标签为$[a\ b\ c\ d\ 1\ 0] \in \Re^6$，其中 a、b、c、d 用于表示此处的值随意设置。

S2033：根据所述真值标签，确定所述第二训练模型对应的损失函数。

对于检测任务(检测框的回归)，我们采用 smooth L1 对网络输出及真值标签进行损失评价，对于分类任务(区分行人与背景)，我们采用交叉熵作为损失，而 YOLOV1 和 V2 均采用 L2 作为检测与分类的损失。

YOLO 损失函数如下所示：

$$L_{center} = \lambda_{corrd} \sum_{i=0}^{S} \sum_{j=0}^{S} \sum_{k=0}^{B} 1_{i,j,k}^{obj} [(x_{i,j,k} - \hat{x}_{i,j,k})^2 + (y_{i,j,k} - \hat{y}_{i,j,k})^2]$$

$$L_{size} = \lambda_{corrd} \sum_{i=0}^{S} \sum_{j=0}^{S} \sum_{k=0}^{B} 1_{i,j,k}^{obj} \left[\left(\sqrt{w_{i,j,k}} - \sqrt{\hat{w}_{i,j,k}} \right)^2 + \sqrt{h_{i,j,k}} - \sqrt{\hat{h}_{i,j,k}})^2 \right]$$

$$L_{socre} = \sum_{i=0}^{S} \sum_{j=0}^{S} \sum_{k=0}^{B} [1_{i,j,k}^{obj} (c_{i,j,k} - \hat{c}_{i,j,k})^2 + \lambda_{noobj} [1_{i,j,k}^{noobj} [(c_{i,j,k} - \hat{c}_{i,j,k})^2]]$$

$$L_{class} = \sum_{i=0}^{S} \sum_{j=0}^{S} \sum_{k=0}^{B} 1_{i,j,k}^{obj} [(p_{i,j,k} - \hat{p}_{i,j,k})^2]$$

$$L_{total} = L_{center} + L_{size} + L_{socre} + L_{class}$$

其中，L_{center}，L_{size}和L_{score}用于表示检测框中心、大小及框内有无物体的置信度损失，L_{class}用于表示分类损失，$c_{i,j,k}$，$\hat{c}_{i,j,k}$分别用于表示网络预测的网格 I,j 处检测框 k 内有无物体的置信度及其真值，$x_{i,j,k}$,$y_{i,j,k}$，$w_{i,j,k}$,$h_{i,j,k}$用于表示网络预测的网格 I,j 处的检测框 k 的位置及大小，$\hat{x}_{i,j,k}$,$\hat{y}_{i,j,k}$,$\hat{w}_{i,j,k}$,$\hat{h}_{i,j,k}$用于表示与之对应的真值，$p_{i,j,k}$ 和 $\hat{p}_{i,j,k}$分别用于表示网络预测的网格 I,j 处的检测框 k 内物体所属类别的概率及其真值，$1_{i,j,k}^{obj}$用来指示网格 I,j 处的检测框 k 是否用来负责检测物体，$1_{i,j,k}^{noobj}$用来指示网格 I,j 处的检测框 k 是否用来负责检测背景。

而本实施例中的，MNPB-YOLO 损失如下所示：

$$L_{class} = \sum_{i=0}^{S} \sum_{j=0}^{S} \sum_{k=0}^{B} CrossEntropy(class_{i,j,k}, \widehat{class}_{i,j,k})$$

$$L_{box} = \sum_{i=0}^{S} \sum_{j=0}^{S} \sum_{k=0}^{B} 1_{i,j,k}^{pos} \left[smooth_{L1} \left(t_g^{i,j,k} - t_p^{i,j,k} \right) \right]$$

$$L_{total} = \lambda_{class} L_{class} + \lambda_{box} L_{box}$$

$$smooth_{L1}(x) = \begin{cases} 0.5x^2, if|x| < 1 \\ |x| - 0.5, otherwise \end{cases}$$

其中，$class_{i,j,k}$用于表示网络预测的物体种类的 one-hot 编码，$\widehat{class}_{i,j,k}$用于表示物体种类的真值，$t_p^{i,j,k}$用于表示网络预测的先验框向真值框变换的值，$t_g^{i,j,k}$用于表示先验框向真值框变换的真值。λ_{class}和λ_{box}是两个常数，用以平衡不同类型的损失，CrossEntropy用于表示交叉熵函数；$1_{i,j,k}^{obj}$用于表示网格 I,j 处的检测框 k 是否用来负责检测物体；x 用于表示函数的输入。

S2034：根据所述损失函数优化所述第二训练模型，得到所述行人检测模型。

通过权重连结层筛选出的特征还需要进一步的提取，最后才能输出特定的张量形状，

与真值标签做对比，构成损失。我们采用 3×3 卷积和 1×1 卷积配合的方式构成我们的检测与分类模块。

如表 2 所示，张量最后输出的形状所用于表示的意义是：将图像分成 7×7 的网格，每个网格预测 2 个检测框，每个检测框需要 6 个参数描述，其中包括描述检测框位置及大小的 4 个参数及区分背景和人物的 one-hot 编码 2 个。

不同方法的技术对比，可以看到，本实施例的方法参数量比原始 YOLO 少了非常多，这意味着本实施例的方法可以更高效的运行。与 YOLOv1 和 YOLOv2 相比，MNPB-YOLO 在网络结构上进行了较大的改动，采用深度可分离卷积及瓶颈层技术将参数量及计算量大大减少，同时提出的权重连结层可以提升网络性能，这让我们的网络同时具备速度及精度的优势。

表 2：分类及检测模块

输入大小	操作	扩展系数	卷积核数量	重复次数	步幅
7×7×768	3×3+1×1 卷积	-	512	1	1
7×7×512	3×3+1×1 卷积	-	256	1	1
7×7×256	3×3+1×1 卷积	-	128	1	1
7×7×128	3×3+1×1 卷积	-	64	1	1
7×7×64	3×3+1×1 卷积	-	32	1	1
7×7×32	3×3+1×1 卷积	-	12	1	1
输出:7×7×12					

不同方法的技术对比如表 3 所示，其中，YOLOv1 的网络结构受 GoogleLenet 启发，输入大小为 224×224×3 的图像，输出网格划分为 7×7，每个网格预测 2 个检测框，检测思想为利用卷积神经网络的映射能力直接构建图像到检测框参数的映射过程，检测框由物体中心位置(x, y)及物体高宽(h, w)共 4 个参数构成，同时加上网格中的某个检测框内是否有物体的置信度 c 及物体种类的 one-hot 编码一起构成网络的输出，整个网络相当于是从零开始学习如何标注出物体。事实上，对于一些特定场景的物体，我们有一些先验的知识可以帮助网络更好的辨识出物体，比如目标物体的高宽比信息，对于行人，一般在 3:1 左右。YOLOv2 就将这种先验知识融入进了网络预测当中，这就是我们说的先验框技术，对于 YOLOv2，输入大小调整为 448×448×3，网格划分为 13×13，每个网格预测 2 个检测框，先验框的大小由训练集中的标签大小通过聚类方法得到，此外，还将网络结构做了很大的改动，提出了 darknet 这个 backbone，参数量是原来的一半左右。YOLOv1 及 YOLOv2 的对比可以在表 3 中看到，由表 3 可知，本实施例中 MNPB-YOLO 的方法在批量标准化、先验框技术、卷积方式、权重连接层以及参数量方面，都有明显的优势：

表 3：不同方法的技术对比

方法名称	批量标准化	先验框技术	卷积方式	权重连接层	参数量
YOLOv1	×	×	标准卷积	×	93970430
YOLOv2	√	√	标准卷积	×	50962889
MNPB-YOLO	√	√	深度可分离卷积	√	6620213

S204：将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据；所述行人检测模型根据预设的深度可分离卷积方式训练得到；

在本实施例中 S204 与图 1 对应的实施例中 S102 的实现方式完全相同，具体可参考图 1 对应的实施例中的 S102 的相关描述，在此不再赘述。

S205：对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框。

请一并参阅之前引入的图 4，图 4 为本实施例提供的训练模型的训练过程和应用过程示意图，在根据历史图像及其扩充图像确定训练模型中的参数和权重之后，便可以将训练模型投入应用中，即通过训练模型来对待检测图像进行行人检测。在检测过程中，先通过车载摄像头来获取实时图像，即待检测图像，将获取到的实时图像输入训练好的训练模型中，检测得到实时图像中每个行人对应的至少一个边框，通过非极大值抑制的方式对各个边框进行计算和判定，确定待检测图像中的行人对应的行人边框，得到精确的检测结果。

在进行仿真实验时，原始 YOLO 是一种多目标检测模型，为了将 YOLO 与 MNPB-YOLO 的单目标检测进行对比，我们对其进行了一些改动，改动后的模型我们称之为 S-YOLOV1 和 S-YOLOV2，此外，我们还对比了两种传统的方法：基于 HAAR 特征和 Adaboost 分类器的行人检测算法，以及基于 HOG 特征和 SVM 分类器的行人检测算法。不同方法的检测结果对比如表 4 所示。其中，AP 用于表示平均精准度，具体为 PR 曲线与 XY 轴包围的面积占比，P 用于表示精准度，R 用于表示召回率，FPS 用于表示处理速度，即每秒处理多少帧。从表 4 可以看出，本实施例的方法不仅比其他方法准确，而且速度还快出很多，实验在 Intel-i7 6700K 和 GTX 1080 计算机平台上测得。

表 4：不同方法的检测结果对比

Method	AP	P	R	FPS
HAAR+Adaboost	35.6	64.7	50.2	-
HOG+SVM	43.5	71.3	54.7	-
S-YOLOv1	62.7	76.8	70.2	80.1
S-YOLOv2	80.8	85.1	84.1	81.7
MNPB-YOLO	86.6	88.0	89.3	151.9

请一并参阅图 12，图 12 为本实施例中的一些检测结果示例，通过图 12 可知，本实施例中的方式可以在雾霾天气或者环境可见度较低的情况下，检测到图像中的行人，并能精

确的确定行人的人像边框。

S206：检测所述人像边框对应的行人相对于当前车辆的方位、所述当前车辆与所述人像边框对应的行人之间的距离。

在确定了获取到的待检测图像中的人像边框之后，可以根据人像边框来确定当前视野中的行人，并进行对应的控制操作。

示例性地，当检测到车辆前方的行人对应的人像边框之后，即确定车辆前方有人，可以再次检测该行人相对于车辆的方位、以及当前车辆与该行人之间的距离。具体的，方位的确定可以根据行人在待处理图像中的区域来确定，例如根据人像边框在待处理图像中的右下，则判定该行人在当前车辆的右前方。距离的确定可以通过红外测距的方式来确定，通过红外测距装置来确定该行人的位置与当前车辆的距离。

S207：根据所述方位和所述距离，生成提醒信息，并进行播报；所述提醒信息用于提醒所述当前车辆的司机注意所述人像边框对应的行人。

在确定了行人相对于当前车辆的方位和距离之后，生成提醒信息，并进行播报，以提醒当前车辆的司机注意人像边框对应的行人。

进一步的，若是在无人驾驶的环境下使用本实施例中的方法，也可以在确定了行人相对于当前车辆的方位和距离之后，根据行人相对于当前车辆的方位和距离，生成车辆控制指令，以通过车辆控制指令来控制车辆行进速度和方向，避免撞向行人。

上述方案，通过实时获取待检测图像；获取包含行人的历史图像；根据预设的权重连接层构建训练模型，根据所述历史图像对所述训练模型进行训练，得到所述行人检测模型。将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据；所述行人检测模型根据预设的深度可分离卷积方式训练得到；对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框。通过对历史图像进行处理增加图像训练的基数，根据权重连接层构建训练模型并进行训练，得到所述行人检测模型，对待检测图像进行检测，并在得到检测结果之后，根据检测结果进行相应的处理，不仅提高了待检测图像的检测效率和检测精度，也提高了车辆驾驶和路人的安全性。

实施例 3：

参见图 13，图 13 是本申请实施例三提供的一种行人检测装置的示意图。行人检测装置 1300 可以为智能手机、平板电脑等移动终端。本实施例的行人检测装置 1300 包括的各单元用于执行图 1 对应的实施例中的各步骤，具体请参阅图 1 及图 1 对应的实施例中的相关描述，此处不赘述。本实施例的行人检测装置 1300 包括：

获取单元 1301，用于实时获取待检测图像；

识别单元 1302，用于将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据；所述行人检测模型根据预设的深度可分离卷积方式训练得到；

确定单元 1303，用于对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框。

进一步的，所述行人检测装置还包括：

历史获取单元，用于获取包含行人的历史图像；

训练单元，用于根据预设的权重连接层构建训练模型，根据所述历史图像对所述训练模型进行训练，得到所述行人检测模型。

进一步的，所述训练单元包括：

扩充单元，用于对所述历史图像进行图像增强处理，得到所述历史图像对应的至少两个扩充图像；

第一训练单元，用于根据预设的权重连接层构建第一训练模型，根据所述历史图像及其对应的所述扩充图像对所述第一训练模型进行训练，得到所述行人检测模型。

进一步的，所述训练单元包括：

第二训练单元，用于根据预设的权重连接层，基于预设的深度可分离卷积方式以及预设的线性瓶颈层技术，构建第二训练模型；

真值单元，用于将所述历史图像输入所述第二训练模型中，检测所述历史图像中的行人图像，并确定每个行人图像对应的真值标签；所述真值标签用于表示识别出的历史图像的先验框基于真值框的变换系数；

损失函数单元，用于根据所述真值标签，确定所述第二训练模型对应的损失函数；

优化单元，用于根据所述损失函数优化所述第二训练模型，得到所述行人检测模型。

进一步的，所述行人检测装置还包括：

定位单元，用于检测所述人像边框对应的行人相对于当前车辆的方位、所述当前车辆与所述人像边框对应的行人之间的距离；

提醒单元，用于根据所述方位和所述距离，生成提醒信息，并进行播报；所述提醒信息用于提醒所述当前车辆的司机注意所述人像边框对应的行人。

上述方案，通过实时获取待检测图像；获取包含行人的历史图像；根据预设的权重连接层构建训练模型，根据所述历史图像对所述训练模型进行训练，得到所述行人检测模型。将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据；所述行人检测模型根据预设的深度可分离卷积方式训练得到；对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框。通过对历史图像进行处理增加图像训练的基数，根据权重连接层构建训练模型并进行训练，得到所述行人检测模型，对待检测图像进行检测，并在得到检测结果之后，根据检测结果进行相应的处理，不仅提高了待检测图像的检测效率和检测精度，也提高了车辆驾驶和路人的安全性。

应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本申请实施例的实施过程构成任何限定。

实施例 4：

参见图 14，图 14 是本申请实施例五提供的一种行人检测装置的示意图。如图 14 所示的本实施例中的行人检测装置 1400 可以包括：处理器 1401、存储器 1402 以及存储在存储器 1402 中并可在处理器 1401 上运行的计算机程序 1403。处理器 1401 执行计算机程序 1403 时实现上述各个行人检测方法实施例中的步骤。存储器 1402 用于存储计算机程序，所述计算机程序包括程序指令。处理器 1401 用于执行存储器 1402 存储的程序指令。其中，处理器 1401 被配置用于调用所述程序指令执行以下操作：

处理器 1401 用于：

实时获取待检测图像；

将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据；所述行人检测模型根据预设的深度可分离卷积方式训练得到；

对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框。

进一步的，处理器 1401 具体用于：

获取包含行人的历史图像；

根据预设的权重连接层构建训练模型，根据所述历史图像对所述训练模型进行训练，得到所述行人检测模型。

进一步的，处理器 1401 具体用于：

对所述历史图像进行图像增强处理，得到所述历史图像对应的至少两个扩充图像；

根据预设的权重连接层构建第一训练模型，根据所述历史图像及其对应的所述扩充图像对所述第一训练模型进行训练，得到所述行人检测模型。

进一步的，处理器 1401 具体用于：

根据预设的权重连接层，基于预设的深度可分离卷积方式以及预设的线性瓶颈层技术，构建第二训练模型；

将所述历史图像输入所述第二训练模型中，检测所述历史图像中的行人图像，并确定每个行人图像对应的真值标签；所述真值标签用于表示识别出的历史图像的先验框基于真值框的变换系数；

根据所述真值标签，确定所述第二训练模型对应的损失函数；

根据所述损失函数优化所述第二训练模型，得到所述行人检测模型。

进一步的，处理器 1401 具体用于：

检测所述人像边框对应的行人相对于当前车辆的方位、所述当前车辆与所述人像边框对应的行人之间的距离；

根据所述方位和所述距离，生成提醒信息，并进行播报；所述提醒信息用于提醒所述当前车辆的司机注意所述人像边框对应的行人。

上述方案，通过实时获取待检测图像；获取包含行人的历史图像；根据预设的权重连接层构建训练模型，根据所述历史图像对所述训练模型进行训练，得到所述行人检测模型。将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据；所述行人检测模型根据预设的深度可分离卷积方式训练得到；对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框。通过对历史图像进行处理增加图像训练的基数，根据权重连接层构建训练模型并进行训练，得到所述行人检测模型，对待检测图像进行检测，并在得到检测结果之后，根据检测结果进行相应的处理，不仅提高了待检测图像的检测效率和检测精度，也提高了车辆驾驶和路人的安全性。

应当理解，在本申请实施例中，所称处理器 1401 可以是中央处理单元（Central Processing Unit，CPU），该处理器还可以是其他通用处理器、数字信号处理器（Digital Signal Processor，DSP）、专用集成电路（Application Specific Integrated Circuit，ASIC）、现成可编程门阵列（Field-Programmable Gate Array，FPGA）或者其他可编程逻辑器件、分立门或

者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

该存储器 1402 可以包括只读存储器和随机存取存储器，并向处理器 1401 提供指令和数据。存储器 1402 的一部分还可以包括非易失性随机存取存储器。例如，存储器 1402 还可以存储设备类型的信息。

具体实现中，本申请实施例中所描述的处理器 1401、存储器 1402、计算机程序 1403 可执行本申请实施例提供的行人检测方法的第一实施例和第二实施例中所描述的实现方式，也可执行本申请实施例所描述的终端的实现方式，在此不再赘述。

在本申请的另一实施例中提供一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令被处理器执行时实现：

实时获取待检测图像；

将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据；所述行人检测模型根据预设的深度可分离卷积方式训练得到；

对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框。

进一步的，所述计算机程序被处理器执行时还实现：

获取包含行人的历史图像；

根据预设的权重连接层构建训练模型，根据所述历史图像对所述训练模型进行训练，得到所述行人检测模型。

进一步的，所述计算机程序被处理器执行时还实现：

对所述历史图像进行图像增强处理，得到所述历史图像对应的至少两个扩充图像；

根据预设的权重连接层构建第一训练模型，根据所述历史图像及其对应的所述扩充图像对所述第一训练模型进行训练，得到所述行人检测模型。

进一步的，所述计算机程序被处理器执行时还实现：

根据预设的权重连接层，基于预设的深度可分离卷积方式以及预设的线性瓶颈层技术，构建第二训练模型；

将所述历史图像输入所述第二训练模型中，检测所述历史图像中的行人图像，并确定每个行人图像对应的真值标签；所述真值标签用于表示识别出的历史图像的先验框基于真值框的变换系数；

根据所述真值标签，确定所述第二训练模型对应的损失函数；

根据所述损失函数优化所述第二训练模型，得到所述行人检测模型。

进一步的，所述计算机程序被处理器执行时还实现：

检测所述人像边框对应的行人相对于当前车辆的方位、所述当前车辆与所述人像边框对应的行人之间的距离；

根据所述方位和所述距离，生成提醒信息，并进行播报；所述提醒信息用于提醒所述当前车辆的司机注意所述人像边框对应的行人。

上述方案，通过实时获取待检测图像；获取包含行人的历史图像；根据预设的权重连接层构建训练模型，根据所述历史图像对所述训练模型进行训练，得到所述行人检测模型。

将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据；所述行人检测模型根据预设的深度可分离卷积方式训练得到；对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框。通过对历史图像进行处理增加图像训练的基数，根据权重连接层构建训练模型并进行训练，得到所述行人检测模型，对待检测图像进行检测，并在得到检测结果之后，根据检测结果进行相应的处理，不仅提高了待检测图像的检测效率和检测精度，也提高了车辆驾驶和路人的安全性。

所述计算机可读存储介质可以是前述任一实施例所述的终端的内部存储单元，例如终端的硬盘或内存。所述计算机可读存储介质也可以是所述终端的外部存储设备，例如所述终端上配备的插接式硬盘，智能存储卡（Smart Media Card，SMC），安全数字（Secure Digital，SD）卡，闪存卡（Flash Card）等。进一步地，所述计算机可读存储介质还可以既包括所述终端的内部存储单元也包括外部存储设备。所述计算机可读存储介质用于存储所述计算机程序及所述终端所需的其他程序和数据。所述计算机可读存储介质还可以用于暂时地存储已经输出或者将要输出的数据。

本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，上述描述的终端和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

在本申请所提供的几个实施例中，应该理解到，所揭露的终端和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另外，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口、装置或单元的间接耦合或通信连接，也可以是电的，机械的或其它的形式连接。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本申请实施例方案的目的。

另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以是两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分，或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备

（可以是个人计算机，服务器，或者网络设备等）执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（Read-Only Memory，ROM）、随机存取存储器（Random Access Memory，RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

以上所述，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到各种等效的修改或替换，这些修改或替换都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以权利要求的保护范围为准。

权利要求书

1. 一种行人检测方法，其特征在于，包括：

实时获取待检测图像；

将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据；所述行人检测模型根据预设的深度可分离卷积方式训练得到；

对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框。

2. 如权利要求 1 所述的行人检测方法，其特征在于，所述将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据之前，包括：

获取包含行人的历史图像；

根据预设的权重连接层构建训练模型，根据所述历史图像对所述训练模型进行训练，得到所述行人检测模型。

3. 如权利要求 2 所述的行人检测方法，其特征在于，所述根据预设的权重连接层构建训练模型，根据所述历史图像对所述训练模型进行训练，得到所述行人检测模型，包括：

对所述历史图像进行图像增强处理，得到所述历史图像对应的至少两个扩充图像；

根据预设的权重连接层构建第一训练模型，根据所述历史图像及其对应的所述扩充图像对所述第一训练模型进行训练，得到所述行人检测模型。

4. 如权利要求 2 所述的行人检测方法，其特征在于，所述根据预设的权重连接层构建训练模型，根据所述历史图像对所述训练模型进行训练，得到所述行人检测模型，包括：

根据预设的权重连接层，基于预设的深度可分离卷积方式以及预设的线性瓶颈层技术，构建第二训练模型；

将所述历史图像输入所述第二训练模型中，检测所述历史图像中的行人图像，并确定每个行人图像对应的真值标签；所述真值标签用于表示识别出的历史图像的先验框基于真值框的变换系数；

根据所述真值标签，确定所述第二训练模型对应的损失函数；

根据所述损失函数优化所述第二训练模型，得到所述行人检测模型。

5. 如权利要求 1-4 任一项所述的行人检测方法，其特征在于，所述对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框之后，还包括：

检测所述人像边框对应的行人相对于当前车辆的方位、所述当前车辆与所述人像边框对应的行人之间的距离；

根据所述方位和所述距离，生成提醒信息，并进行播报；所述提醒信息用于提醒所述

当前车辆的司机注意所述人像边框对应的行人。

6. 一种行人检测装置，其特征在于，包括：

获取单元，用于实时获取待检测图像；

识别单元，用于将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据；所述行人检测模型根据预设的深度可分离卷积方式训练得到；

确定单元，用于对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框。

7. 如权利要求 6 所述的行人检测装置，其特征在于，所述行人检测装置还包括：

历史获取单元，用于获取包含行人的历史图像；

训练单元，用于根据预设的权重连接层构建训练模型，根据所述历史图像对所述训练模型进行训练，得到所述行人检测模型。

8. 如权利要求 6 所述的行人检测装置，其特征在于，所述训练单元包括：

扩充单元，用于对所述历史图像进行图像增强处理，得到所述历史图像对应的至少两个扩充图像；

第一训练单元，用于根据预设的权重连接层构建第一训练模型，根据所述历史图像及其对应的所述扩充图像对所述第一训练模型进行训练，得到所述行人检测模型。

9. 一种行人检测装置，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序，其特征在于，所述处理器执行所述计算机程序时实现如权利要求 1 至 5 任一项所述方法的步骤。

10. 一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现如权利要求 1 至 5 任一项所述方法的步骤。

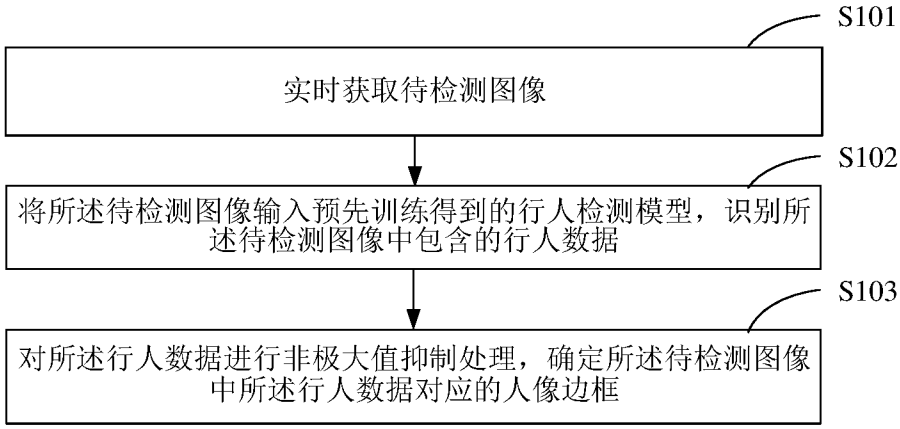

S101
实时获取待检测图像
S102
将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据
S103
对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框

图 1

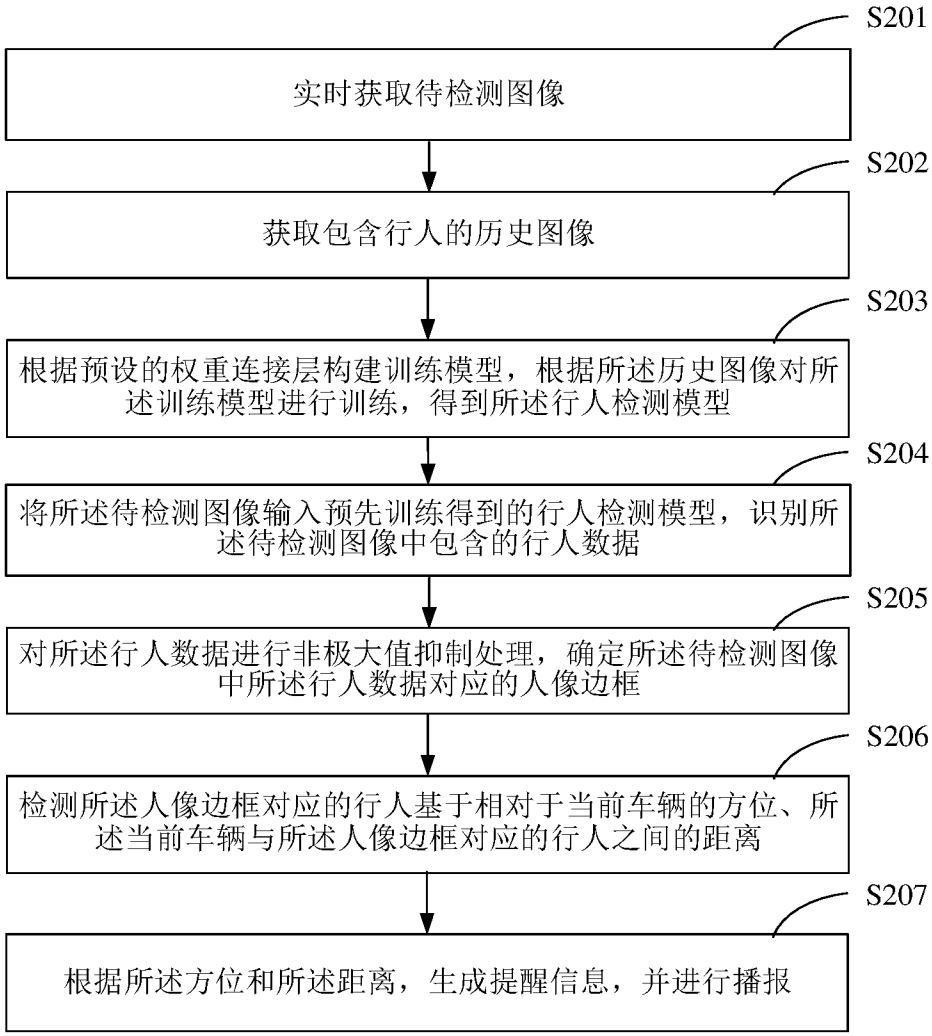
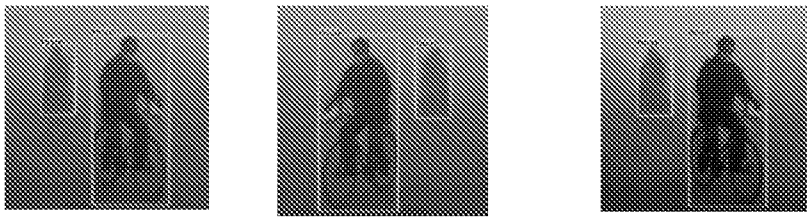
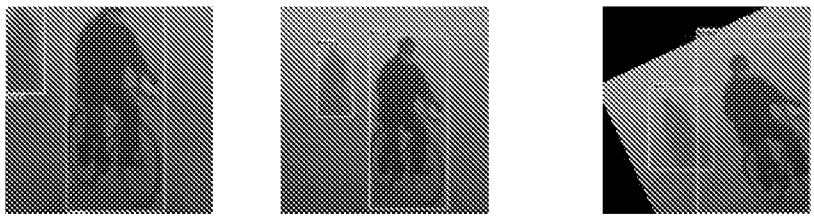

S201
实时获取待检测图像
S202
获取包含行人的历史图像
S203
根据预设的权重连接层构建训练模型，根据所述历史图像对所述训练模型进行训练，得到所述行人检测模型
S204
将所述待检测图像输入预先训练得到的行人检测模型，识别所述待检测图像中包含的行人数据
S205
对所述行人数据进行非极大值抑制处理，确定所述待检测图像中所述行人数据对应的人像边框
S206
检测所述人像边框对应的行人基于相对于当前车辆的方位、所述当前车辆与所述人像边框对应的行人之间的距离
S207
根据所述方位和所述距离，生成提醒信息，并进行播报

图 2



(a) 原图 (b) 随机翻转 (c) 随机对比度改变



(d) 随机裁剪 (e) 随机颜色改变 (f) 随机仿射变换

图 3

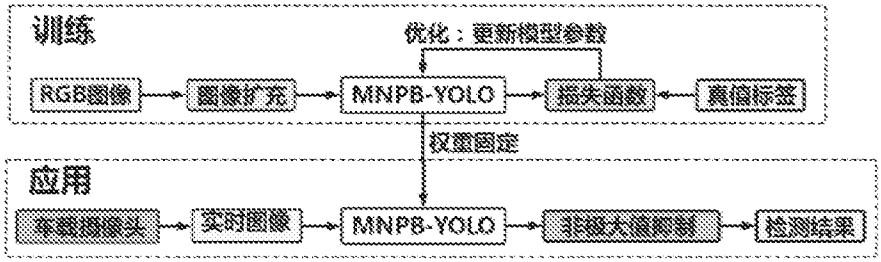


图 4

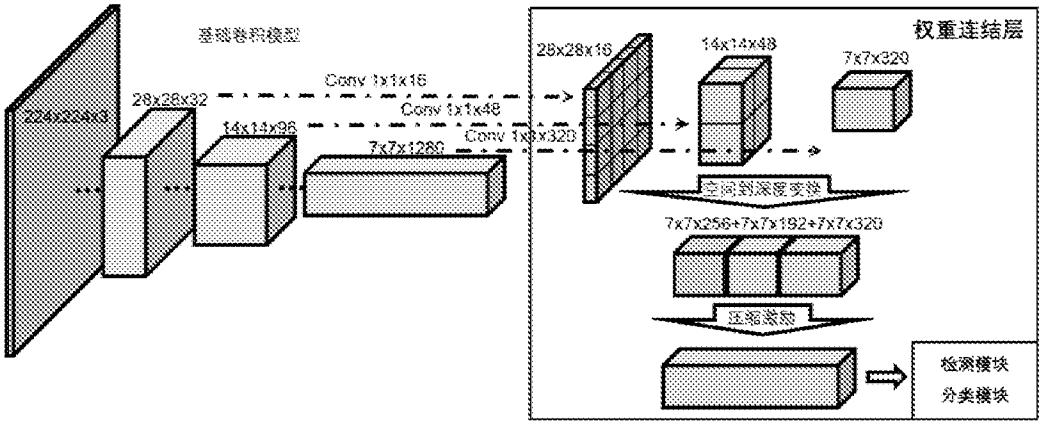


图 5

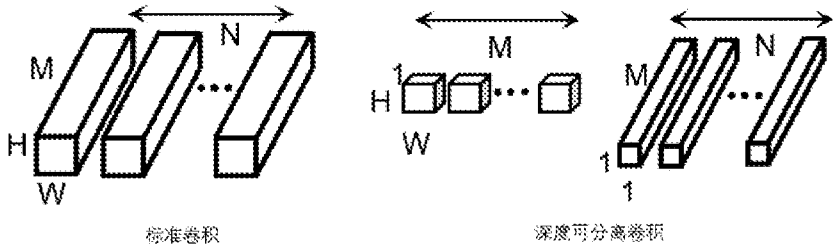


图 6

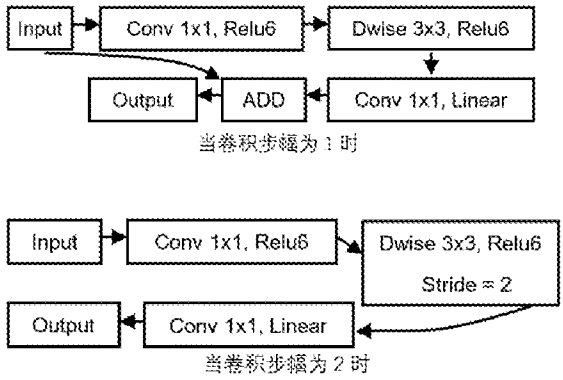

Input
Conv 1x1, Relu6
Dwise 3x3, Relu6
Output
ADD
Conv 1x1, Linear
当卷积步幅为 1 时
Input
Conv 1x1, Relu6
Dwise 3x3, Relu6
Stride = 2
Output
Conv 1x1, Linear
当卷积步幅为 2 时

图 7

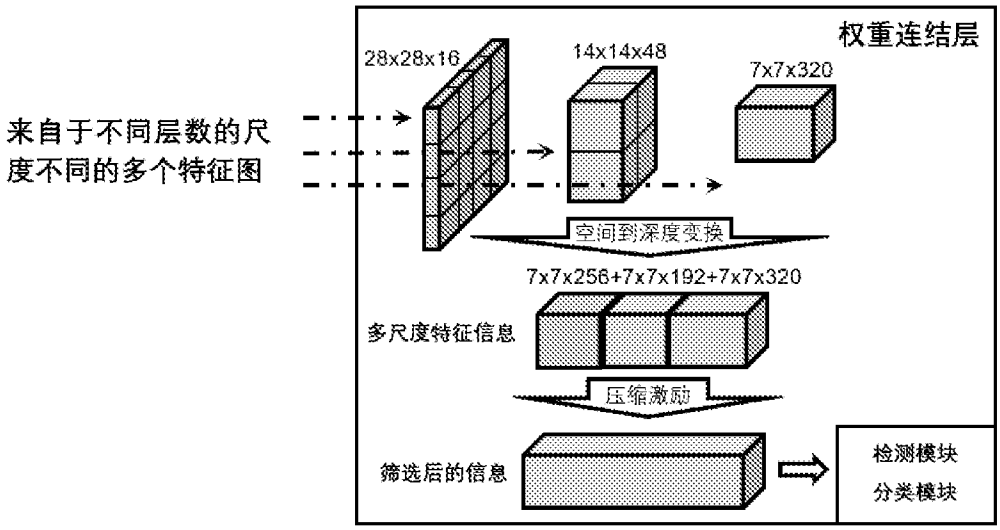

来自于不同层数的尺度不同的多个特征图
28x28x16
14x14x48
7x7x320
权重连结层
空间到深度变换
7x7x256+7x7x192+7x7x320
多尺度特征信息
压缩激励
筛选后的信息
检测模块
分类模块

图 8

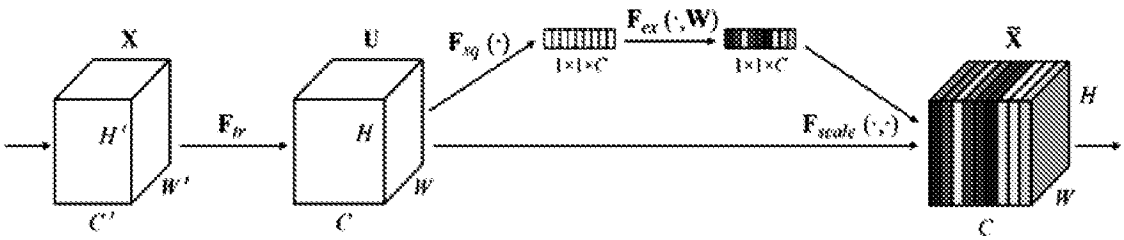

X
H'
W'
C'
F_tr
U
H
W
C
F_sq (·)
1×1×C
F_ex (·,W)
1×1×C
F_scale (·,·)
X̃
H
W
C

图 9

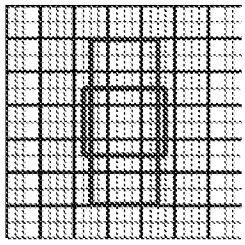


图 10

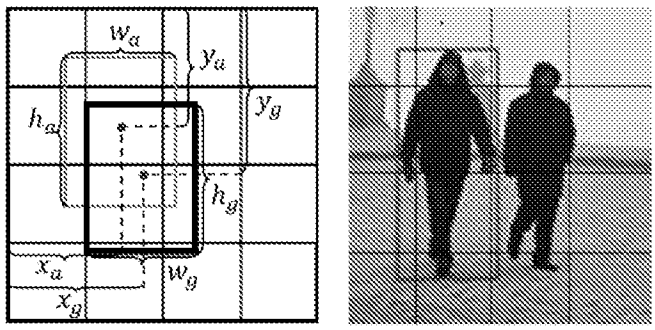


图 11



图 12



图 13

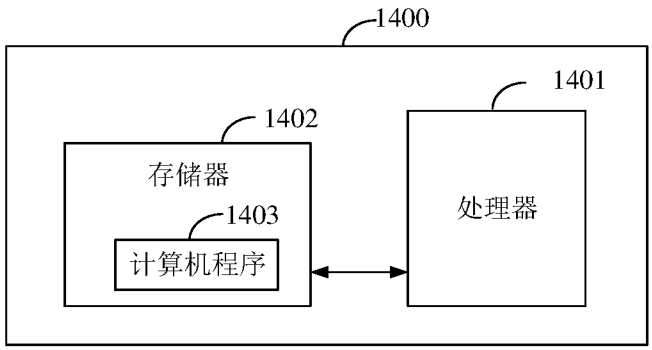


图 14

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/093024

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/62(2006.01)i; G06N 3/04(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; G06N

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)

CNABS; CNTXT; VEN; USTXT; WOTXT; EPTXT; CNKI: 深度可分离, 线性瓶颈, 行人检测, YOLO, Mobilenet v2, depth separable convolution, depthwise separable convolution, linear bottleneck, pedestrian detection

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 109815886 A (NANJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS) 28 May 2019 (2019-05-28) claims 1-10, and figures 1-8	1-10
Y	CN 109086678 A (TIANJIN UNIVERSITY) 25 December 2018 (2018-12-25) claims 1-8	1-10
A	CN 109472315 A (JIANGSU MUMENG INTELLIGENT TECHNOLOGY CO., LTD.) 15 March 2019 (2019-03-15) entire document	1-10
A	US 2019147318 A1 (GOOGLE LLC) 16 May 2019 (2019-05-16) entire document	1-10

☐ 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	Date of mailing of the international search report
08 March 2020	**25 March 2020**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration** **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/CN2019/093024

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 109815886 A	28 May 2019	None	
CN 109086678 A	25 December 2018	None	
CN 109472315 A	15 March 2019	None	
US 2019147318 A1	16 May 2019	WO 2019099074 A1	23 May 2019

国际检索报告

国际申请号 PCT/CN2019/093024

A. 主题的分类

G06K 9/62(2006.01)i; G06N 3/04(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; G06N

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）)

CNABS;CNTXT;VEN;USTXT;WOTXT;EPTXT;CNKI: 深度可分离，线性瓶颈，行人检测，YOLO，Mobilenet v2，depth separable convolution，depthwise separable convolution，linear bottleneck，pedestrian detection

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
Y	CN 109815886 A (南京邮电大学) 2019年 5月 28日 (2019-05-28) 权利要求1-10、图1-8	1-10
Y	CN 109086678 A (天津大学) 2018年 12月 25日 (2018-12-25) 权利要求1-8	1-10
A	CN 109472315 A (江苏木盟智能科技有限公司) 2019年 3月 15日 (2019-03-15) 全文	1-10
A	US 2019147318 A1 (GOOGLE LLC) 2019年 5月 16日 (2019-05-16) 全文	1-10

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 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	国际检索报告邮寄日期
2020年 3月 8日	2020年 3月 25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	受权官员 廖小丽 电话号码 86-(20)-28950434

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/093024

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	109815886	A	2019年 5月 28日	无			
CN	109086678	A	2018年 12月 25日	无			
CN	109472315	A	2019年 3月 15日	无			
US	2019147318	A1	2019年 5月 16日	WO	2019099074	A1	2019年 5月 23日