

(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



(10) 国际公布号

WO 2021/088300 A1

(43) 国际公布日  
2021年5月14日 (14.05.2021)

(51) 国际专利分类号:  
G06K 9/00 (2006.01) G06N 3/02 (2006.01)  
G06K 9/46 (2006.01)

(21) 国际申请号: PCT/CN2020/080991

(22) 国际申请日: 2020年3月25日 (25.03.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:  
201911090619.5 2019年11月9日 (09.11.2019) CN

(71) 申请人: 北京工业大学(BEIJING UNIVERSITY OF TECHNOLOGY) [CN/CN]; 中国北京市朝阳区平乐园100号, Beijing 100124 (CN)。

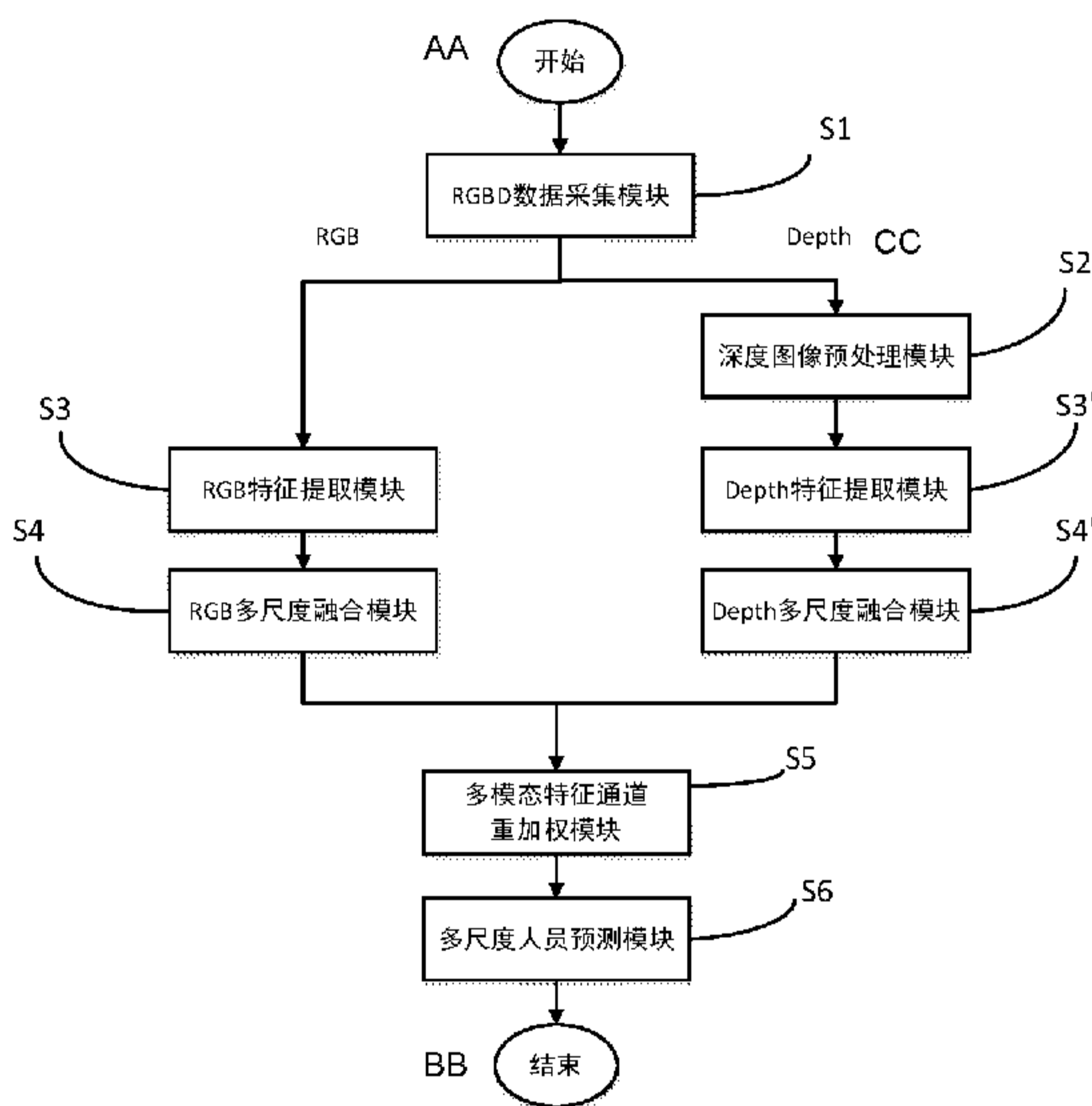
(72) 发明人: 张文利(ZHANG, Wenli); 中国北京市朝阳区平乐园100号, Beijing 100124 (CN)。 郭向(GUO, Xiang); 中国北京市朝阳区平乐园100号, Beijing 100124 (CN)。 杨堃(YANG, Kun); 中国北京市朝阳区平乐园100号, Beijing 100124 (CN)。 王佳琪(WANG, Jiaqi); 中国北京市朝阳区平乐园100号, Beijing 100124 (CN)。

(74) 代理人: 北京思海天达知识产权代理有限公司(BEIJING SIHAI TIANDA INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市朝阳区百子湾西里403号楼1011, Beijing 100124 (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,

(54) Title: RGB-D MULTI-MODE FUSION PERSONNEL DETECTION METHOD BASED ON ASYMMETRIC DOUBLE-STREAM NETWORK

(54) 发明名称: 一种基于非对称双流网络的RGB-D多模态融合人员检测方法



S1 RGBD DATA COLLECTION MODULE  
S2 DEPTH IMAGE PREPROCESSING MODULE  
S3 RGB FEATURE EXTRACTION MODULE  
S3' DEPTH FEATURE EXTRACTION MODULE  
S4 RGB MULTI-SCALE FUSION MODULE  
S4' DEPTH MULTI-SCALE FUSION MODULE  
S5 MULTI-MODE FEATURE CHANNEL REWEIGHTING MODULE  
S6 MULTI-SCALE PERSONNEL PREDICTION MODULE  
AA START  
BB END  
CC DEPTH

图 1

(57) Abstract: The present invention relates to the field of computer vision and image processing, and disclosed in the present invention is an RGB-D multi-mode fusion personnel detection method based on an asymmetric double-stream network. The method comprises the steps of RGBD image collection, depth image preprocessing, RGB feature extraction and Depth feature extraction, RGB multi-scale fusion and Depth multi-scale fusion, multi-mode feature channel reweighting, and multi-scale personnel prediction. According to the present invention, an asymmetric RGBD double-stream convolutional neural network model is designed to solve the problem that tra-

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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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))。

---

ditional symmetric RGBD double-stream networks are prone to causing depth feature loss. Multi-scale fusion structures are respectively designed for RGBD double-stream networks, so that multi-scale information complementation is achieved. A multi-mode reweighting structure is constructed, RGB and Depth feature maps are combined, and weighted assignment is performed on each combined feature channel to implement model automatic learning contribution proportion. Personnel classification and frame regression are performed by using multi-mode features, so that the accuracy of personnel detection is improved while the real-time performance is ensured, and the robustness of detection under low illumination at night and personnel shielding is enhanced.

**(57) 摘要:** 本发明公开了一种基于非对称双流网络的RGB-D多模态融合人员检测方法, 属于计算机视觉与图像处理领域。包含RGBD图像采集, 深度图像预处理, RGB特征提取和Depth特征提取, RGB多尺度融合和Depth多尺度融合, 多模态特征通道重加权以及多尺度人员预测; 本发明针对传统对称型RGBD双流网络易导致深度特征流失的问题, 设计非对称RGBD双流卷积神经网络模型。对RGBD双流网络分别设计多尺度融合结构, 实现多尺度信息互补。构建多模态重加权结构, 合并RGB和Depth特征图, 并对合并后的每个特征通道加权赋值, 实现模型自动学习贡献占比。利用多模态特征进行人员的分类和边框回归, 在保证实时性的同时, 提高人员检测的准确性, 并增强对夜间低照度以及人员遮挡下检测的鲁棒性。



## 一种基于非对称双流网络的 RGB-D 多模态融合人员检测方法

### 技术领域

本发明属于计算机视觉与图像处理领域，具体涉及一种基于非对称双流网络的 RGB-D 多模态融合人员检测方法。

### 背景技术

近年来，智慧家居、智慧建筑以及智能安防等领域得到了飞速发展，视频提取与分析技术的广泛应用成为推动其进步的关键动力，其中人员的检测与统计逐渐成为图像视频分析和人工智能领域的一个热门研究课题。在智慧家居方面，通过检测室内人员可以对人的位置进行定位，记录人员的行为习惯进行记录，进一步调节室内照明、空调等智能设备，为人们提供更为舒适智慧的家庭环境。在智慧建筑方面，人员检测技术可以应用于服务型机器人实现精准避障与办公文件传递，同时依据室内人员位置以及密集程度，可以自动化调节办公区舒适度，提高办公效率。在智能安防方面，安防监控视频中进行的人员检测可以用于身份核验，有效应对陌生人非法闯入，对可疑人员进行跟踪调查和异常行为分析，为智能安防体系提供核心的视频信息支撑。

目前人员检测主要有两种方式：基于 RGB 图像的人员检测和基于多模态图像融合的人员检测。

1) 基于 RGB 图像的人员检测方法是仅在 RGB 图像下进行人员的检测，典型的人员方法有通过基于 RGB 人脸的人员检测方法和基于 RGB 全身的人员检测方法。基于 RGB 人脸的人员检测方法通过在仅 RGB 图像下，对人脸所在区域进行人脸关键点标定、人脸特征编码等方式，提取人脸的通用特征表示，并采用机器学习或深度学习的方法训练人脸检测模型，通过模型预测输出的外接矩形框，在测试样本图像中框选并定位人物的人脸区域，从而达到人员检测的目的。基于 RGB 全身的人员检测方法不同于人脸检测，该方法是仅在 RGB 图像下，提取包含人物整个身体或具有辨识力的主要身体部位的图像区域进行特征表示，并训练基于全身图像的人员检测模型，通过模型预测输出的外接矩形框，框选并定位人物的全身区域，从而达到人员检测的目的。但该方法易受到场景的限制和图像成像分辨率的影响。由于可见光相机的光学成像原理，可见光相机捕获的 RGB 彩色图像对于光照条件变化的抗扰动性差，尤其在夜间、雨雪雾天等低照度场景下，相机实时拍摄捕获的图像呈现一片黑暗或相似背景，无法从图像中清



晰地分辨出的前景人员和背景信息，会很大程度上影响检测模型的训练收敛，降低人员检测的精确度。此外在对场景内的多个人物进行检测时，通常会产生人与物体之间的遮挡或者人与人之间的交叉遮挡，可见光相机无法获得场景中物体或人员的深度信息和热辐射信息，因此其捕获二维平面图像无法有效凸显被遮挡目标的边缘轮廓、纹理等解决人员遮挡问题的关键信息，甚至被相似背景信息淹没，导致人员检测的查准率和查全率大幅下降。

2) 基于多模态图像融合的人员检测方法不同于基于 RGB 图像的人员检测方法，其输入数据是来源于同一检测场景下的不同图像源的图像，例如 RGB 图像、深度图像、红外热图像，每一种图像源通过不同的相机设备捕获，图像本身具备不同的特性。多模态图像融合的检测方法主要是利用不同模态的图像交叉融合，以实现特征强化和互补关联。红外热图像和深度图像相比 RGB 彩色图像对光照变化的鲁棒性较好，能够在夜间等低照度条件下稳定成像，并且由于红外热相机、深度相机与可见光相机的成像原理不同，二者能够较好地捕获部分遮挡下的人员边缘轮廓等辅助线索，在一定程度上可以缓解部分遮挡的问题。现如今多采用深度学习的方法实现多模态信息的特征融合和关联建模，训练完成的模型对于多约束多场景条件下（例如夜间低照度、严重遮挡、远距离拍摄等）的人员检测具有更好的鲁棒性。但是现存方法对于多模态图像融合方式，多采用传统手工提取多模态特征融合以及利用 RGBT 或 RGBD（彩色图像+热红外图像，彩色图像+深度图像）双流神经网络进行附加四通道融合、单一尺度融合以及加权决策融合等简单融合方式。传统手工多模态融合方法需要人为设计并提取多模态特征，依赖于主观经验且费时费力，无法实现端到端的人员检测。而简单的双流神经网络多模态融合策略，无法充分有效地利用彩色图像的色彩、纹理等细粒度信息和深度图像提供的边缘、深度等语义信息，实现多模态数据之间的关联互补，甚至由于模型复杂度过高而产生过拟合现象，导致人员检测的查准率和查全率不升反降。而 RGB-T 人员检测由于红热成像相机价格昂贵，在实际应用中由于成本过高具有很大的局限性。

现有代表性技术 1 项。

**(1) 发明名称：**一种基于 RGBD 的行人检测和身份识别方法及系统（申请号：201710272095）

本发明提供了一种基于 RGBD 的行人检测和身份识别方法，方法包括：输



入 RGB 和深度图像，并对图像进行预处理，转换颜色通道；然后构建 RGB 和深度图像的多通道特征，具体的，首先计算 RGB 图像的水平梯度和垂直梯度构建 RGB 梯度方向直方图特征，以及深度图像的水平梯度、垂直梯度和深度法向量方向，构建深度图像的梯度方向直方图，作为 RGBD 的多通道特征；计算深度图像每个像素点对应的尺度，对尺度进行量化，获取尺度列表；根据多通道特征，采用 Adaboost 算法训练行人检测分类器；采用检测分类器，搜索尺度列表对应的尺度空间，得到包含行人信息的外接矩形框，完成行人检测

但此方法需要手工提取传统的 RGBD 图像的梯度方向直方图作为图像特征，耗时费力且占用较大存储空间，无法端到端的实现行人检测；梯度方向直方图特征较为简单，难以提取 RGB 和深度图像中具有辨识力的特征进行行人检测；该方法采用 RGB 和深度图像特征的简单融合，难以充分有效地挖掘利用 RGB 图像的色彩、纹理等细粒度信息和深度图像提供的边缘、深度等语义信息，实现多模态数据之间的关联互补，在提升行人检测的精确度方面具有很大的局限性。

## 发明内容

针对现有技术中的缺陷，本发明提供了一种基于非对称双流网络的 RGBD 多模态融合人员检测方法，但不限于人员检测，也可以应用于目标检测、车辆检测等任务。

本发明提供的一种基于非对称双流网络的 RGBD 多模态融合人员检测方法代表图如图 1 所示，包含 RGBD 图像采集，深度图像预处理，RGB 特征提取和 Depth 特征提取，RGB 多尺度融合和 Depth 多尺度融合，多模态特征通道重加权以及多尺度人员预测，各步骤的具体功能如下：

### S1 RGBD 图像采集；

利用具有同时拍摄 RGB 图像和深度图像功能的相机获取原始 RGB 图像和深度图像（以下简称为 Depth 图像），并对 RGB 和 Depth 图像进行匹配分组，每组图像由一张 RGB 图像和同场景下捕获的 Depth 图像组成，输出分组匹配后的 RGB 和 Depth 图像。原始 RGB 图像和 Depth 图像也可以从公开 RGBD 数据集获取。

### S2 深度图像预处理；

从 S1 的 RGBD 图像采集获取分组匹配后的 Depth 图像，首先消除 Depth 图像的部分噪声，然后进行空洞填充，最后将单通道 Depth 图像重新编码为三个通道图像，并将三个通道的图像数值重新规范化到 0-255，输出编码规范化后的 Depth 图像。

S3 RGB 特征提取和 Depth 特征提取；

从所述 S1 的 RGBD 图像采集获取原始 RGB 图像，输入到 RGB 特征提取（非对称双流网络模型的 RGB 网络流），进行下采样特征提取，输出 RGB 图像的高、中、低分辨率特征图，分别记为 RGB\_FP\_H、RGB\_FP\_M、RGB\_FP\_L，代表 RGB 图像的低级色彩纹理、中级边缘轮廓和高级语义特征表示；从深度图像预处理获取编码规范化后的 Depth 图像，输入到 Depth 特征提取（非对称双流网络模型的 Depth 网络流），进行下采样特征提取，输出 Depth 图像的高、中、低分辨率特征图，分别记为 D\_FP\_H、D\_FP\_M、D\_FP\_L，代表 Depth 图像的低级色彩纹理、中级边缘轮廓和高级语义特征表示。RGB 网络流和 Depth 网络流是对称结构的，即 RGB 网络流和 Depth 网络流的结构完全相同。但 Depth 图像所包含的特征相对于 RGB 图像更简单，当采用与 RGB 网络相同深度的卷积网络结构提取 Depth 特征时，会由于网络传递过深而导致 Depth 特征消失，同时网络参数增加了过拟合的风险。基于上述原因，设计非对称双流卷积神经网络模型提取 RGB 图像和 Depth 图像特征。图 2-1 至图 2-4 为本方法设计的非对称双流卷积神经网络模型的一种具体实施例结构，但不限于图 2-1 至图 2-4 所示的结构。图 2-1 所述 DarkNet-53 和图 2-2 所述 MiniDepth-30 分别代表 RGB 网络流和 Depth 网络流，二者的网络结构具有非对称的特性。

S4 RGB 多尺度融合和 Depth 多尺度融合；

从 RGB 特征提取获取 RGB 特征图 RGB\_FP\_H、RGB\_FP\_M、RGB\_FP\_L 输入到 RGB 多尺度融合，首先将获取的 RGB\_FP\_L 通过上采样层拓展到与 RGB\_FP\_M 相同尺寸，然后与 RGB\_FP\_M 进行通道合并，实现 RGB 网络深层的高级语义特征与中间层的中级边缘轮廓特征的互补融合，输出通道合并后的新特征图 RGB\_FP\_M；然后对输出通道合并后的新特征图 RGB\_FP\_M，通过上采样层拓展到与 RGB\_FP\_H 相同尺寸，与 RGB\_FP\_H 进行通道合并，实现 RGB 网络深层的高级语义特征、中间层的中级边缘轮廓特征以及浅层的低级色彩纹理特征的互补融合，输出通道合并后的新特征图 RGB\_FP\_H；从 Depth 特征提取获



取 Depth 特征图  $D\_FP\_H$ 、 $D\_FP\_M$ 、 $D\_FP\_L$  输入到 Depth 多尺度融合，与 RGB 多尺度融合执行同样的操作。最终 Depth 多尺度融合的输出为原始输入  $RGB\_FP\_L$ 、通道合并后的新特征图  $RGB\_FP\_M$  和  $RGB\_FP\_H$ ；Depth 多尺度融合的输出为原始输入  $D\_FP\_L$ 、通道合并后的新特征图  $D\_FP\_M$  和  $D\_FP\_H$ 。

S5 多模态特征通道重加权；

从 RGB 多尺度融合获取 RGB 特征图  $RGB\_FP\_L$ 、 $RGB\_FP\_M$ 、 $RGB\_FP\_H$  和从 Depth 多尺度融合获取 Depth 特征图  $D\_FP\_L$ 、 $D\_FP\_M$ 、 $D\_FP\_H$ ，按照分辨率分组输入到多模态特征通道重加权中对应相同分辨率的通道重加权结构中，实现更有效的 RGB 与 Depth 的多模态特征融合，提高处理多种限制场景下的检测鲁棒性。具体做法以  $RGB\_FP\_L$  与  $D\_FP\_L$  通道重加权为例，从 RGB 多尺度融合获取  $RGB\_FP\_L$  以及从 Depth 多尺度融合获取  $D\_FP\_L$ ，首先进行通道合并，获得通道合并后的特征图记为  $Concat\_L$ ；然后应用通道重加权模块(以下简称为  $RW\_Module$ )对  $Concat\_L$  的特征通道进行线性加权，为每个特征通道赋予权重，输出通道重加权后的特征图记为  $RW\_L$ 。 $RGB\_FP\_M$  与  $D\_FP\_M$ ， $RGB\_FP\_H$  与  $D\_FP\_H$  的通道重加权采用与所述  $RGB\_FP\_L$  和  $D\_FP\_L$  相同方式完成。最终多模态特征通道重加权输出通道重加权后的低、中、高分辨率特征图，分别记为  $RW\_L$ 、 $RW\_M$ 、 $RW\_H$ 。

S6 多尺度人员预测；

从所述 S5 的多模态特征通道重加权获取通道重加权后的特征图  $RW\_L$ 、 $RW\_M$ 、 $RW\_H$ ，分别输入到多尺度人员预测中对应的预测分支中进行分类和边框坐标回归，获得较大、中等以及较小尺寸人员的预测结果。由于特征图分辨率不同，特征图上每个预测点对应的感受野也不同。 $RW\_L$  上的每个预测点具有较大感受野，用来预测图像中的较大目标； $RW\_M$  上的每个预测点具有中等感受野，用来预测图像中的中等目标； $RW\_H$  上的每个预测点具有较小感受野，用来预测图像中的较小目标。对上述三种不同尺度的预测结果进行汇总，采用非极大值抑制(以下简称 NMS)算法[1]剔除重叠目标边框，输出最终保留的人员检测结果，即人员的类别置信分数  $C_i$  和预测矩形边框  $(x_i^1, y_i^1, x_i^2, y_i^2)$ ， $i=1, 2, \dots, N$ 。在本实施例中， $i$  代表人员的 ID 编号， $N$  为在当前图像中保留的人员检测结果总数。

$x_i^1, y_i^1, x_i^2, y_i^2$  分别代表所有包含人员的矩形边框的左上角横坐标、左上角纵坐标、右下角横坐标以及右下角纵坐标。

与现有技术相比较，本发明针对传统对称型 RGBD 双流网络（RGB 网络流+Depth 网络流）由于 Depth 网络过深而易导致深度特征流失的问题，本发明设计非对称 RGBD 双流卷积神经网络模型，Depth 网络流通过对 RGB 网络流进行有效地模型剪枝获得，在降低参数的同时，能够降低模型过拟合的风险，提高检测精度。RGB 网络流和 Depth 网络流分别用于提取 RGB 和深度图像（以下简称为 Depth 图像）的高、中、低分辨率特征图，分别代表 RGB 和 Depth 图像的低级色彩纹理、中级边缘轮廓和高级语义特征表示；其次对 RGB 网络流和 Depth 网络流分别设计多尺度融合结构，实现低分辨率特征图所包含的高级语义特征与中、高分辨率特征图所包含的中级边缘轮廓、低级色彩纹理特征的多尺度信息互补；然后构建多模态特征通道加权结构，合并 RGB 与 Depth 特征图，并为合并后的每个特征通道进行加权赋值，使模型能够自动学习贡献比重，完成特征选择和去除冗余的功能，从而实现 RGB 和 Depth 特征在对应高、中、低分辨率下的多模态特征融合；最后，利用多模态特征进行人员的分类和边框回归，在保证实时性的同时，提高人员检测的准确性，并增强对夜间低照度以及人员遮挡下检测的鲁棒性。

## 附图说明

图 1 本发明提供的一种基于非对称双流网络的 RGBD 多模态融合人员检测方法的代表图

图 2-1 为一种 RGB 网络流——DarkNet-53 结构图，图 2-2 为一种 Depth 网络流——MiniDepth-30 结构图，图 2-3 为一种卷积块通用结构图，图 2-4 为一种残差卷积块通用结构图。

图 3 本发明实施例提供的一种基于非对称双流网络的 RGBD 多模态融合人员检测方法的流程图

图 4 本发明实施例提供的一种通道重加权模块的通用结构图

图 5 本发明实施例提供的 NMS 算法的流程图

## 具体实施方式



为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整的描述。显然，所描述的实施例是本发明的一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。下面通过具体实施例对本发明进行详细说明。

本发明的实施例提供的方法示意图如图 3 所示，包括如下步骤：

**S1:** 利用具有同时拍摄 RGB 图像和深度图像功能的相机获取原始 RGB 图像和深度图像，对图像进行匹配和分组，输出分组匹配后的 RGB 和 Depth 图像。

步骤 S110: 利用具有同时拍摄 RGB 图像和深度图像功能的相机获取原始 RGB 图像，原始 RGB 图像也可以从公开 RGBD 数据集获取。

步骤 S120: 从所述步骤 S110 同步获取与 RGB 图像匹配的 Depth 图像，并对 RGB 和 Depth 图像进行分组，每组图像由一张 RGB 图像和同场景下捕获的深度图像组成，输出分组匹配后的 Depth 图像。

**S2:** 从所述步骤 S120 中获取的分组匹配后的 Depth 图像，对 Depth 图像进行去噪、空洞修复和编码规范化，输出编码预处理后的 Depth 图像。

从所述步骤 S120 获取的原始深度图像作为输入，首先消除 Depth 图像的部分噪声，然后进行空洞填充，最后将单通道 Depth 图像重新编码为三通道图像，并将三个通道的数值重新规范化到 0-255，输出编码规范化后的 Depth 图像。在本实施例中去噪采用 5x5 的高斯滤波器；空洞修复采用[2]提出的图像修复算法，提取 Depth 图像中的局部法线向量和遮挡边界，然后应用全局优化进行 Depth 图像空洞填充；Depth 图像编码采用 HHA 编码[3]（horizontal disparity, height above ground, and the angle the pixel），三个通道分别为水平视差，高于地面的高度以及表面法向量的角度。

**S3:** 从所述步骤 S110 获取原始 RGB 图像，采用非对称双流网络模型的 RGB 网络流在不同网络层级分别提取 RGB 图像的通用、低级、中级和高级特征之后，输出对应通用特征图以及高、中、低三种分辨率的 RGB 特征图，分别记为 RGB\_FP\_C、RGB\_FP\_H、RGB\_FP\_M、RGB\_FP\_L，并将 RGB\_FP\_H、RGB\_FP\_M、

RGB\_FP\_L 其输入到 S4。在本实施例中，非对称双流网络模型的 RGB 网络流采用 DarkNet-53[4]，DarkNet-53 的网络结构如图 2-1 所示。网络共包含 52 个卷积层，其中网络的 L1~L10 层用于提取 RGB 图像的通用特征，输出 RGB\_FP\_C；L11~L27 层用于提取 RGB 图像的低级色彩纹理特征，输出 RGB\_FP\_H；L28~L44 层用于提取 RGB 图像的中级边缘轮廓特征，输出 RGB\_FP\_M；L45~L52 层用于提取 RGB 图像的高级语义特征，输出 RGB\_FP\_L。值得注意的是，本实施例使用的 DarkNet-53 模型仅为所述非对称双流网络的 RGB 网络流的一个具体实施例，不限于上述 DarkNet-53 模型，下文仅以 DarkNet-53 为例进行方法论述。

步骤 S310：从所述 S110 获取原始 RGB 图像，经过 DarkNet-53 网络的 L1~L10 层提取 RGB 图像的通用特征，并将图像分辨率下采样 K 倍，输出 RGB 通用特征图 RGB\_FP\_C，其尺寸变为原始输入尺寸的 K 分之一。在本实施例中，K 取值为 8。L1~L10 层可以划分为 L1~L2、L3~L5 和 L6~L10 三个子采样层，每个子采样层将来自上一层的输入图像分辨率下采样 2 倍。第一子采样层包括 1 个步长为 1 的标准卷积块(记为 Conv0)和 1 个步长为 2 的池化卷积块(记为 Conv0\_pool)，其中卷积块的通用结构如图 2-3 所示，包括标准图像卷积层、批规范化层和 Leaky\_ReLU 激活层；第二子采样层包括一个残差卷积块(记为 Residual\_Block\_1)和 1 个所述池化卷积块(记为 Conv1\_pool)，其中残差卷积块的通用结构如图 2-4 所示，包含一个  $1 \times 1 \times M$  的标准卷积块、一个  $3 \times 3 \times N$  的标准卷积块以及一个将输入的恒等映射传递到输出的 Add 模块，M 代表输入特征通道数，N 表示输出特征通道数，此处 M,N 取值分别为 32；第三子采样层包括 2 个所述残差卷积块(记为 Residual\_Block\_2\_1 ~ 2\_2)和 1 个所述池化卷积块(记为 Conv2\_pool)。在本实施例中，K 取值为 8，M,N 的取值见图 3-1 的 L1~L10 层。

步骤 S320：从所述 S310 获取 RGB\_FP\_C，经过 DarkNet-53 网络的 L11~L27 层提取 RGB 图像的低级色彩纹理特征，并将图像分辨率下采样 K 倍，输出 RGB 高分辨率特征图 RGB\_FP\_H，其尺寸变为原始输入尺寸的 K 分之一。在本实施例中，L11~L27 由 8 个所述残差卷积块(记为 Residual\_Block\_3\_1 ~ 3\_8)和 1 个所述池化卷积块(Conv3\_pool)组成。K 取值为 2，M,N 的取值见图 3-1 的 L11~L27 层。

步骤 S330：从所述 S320 获取 RGB\_FP\_H，经过 DarkNet-53 网络的 L28~L44 层提取 RGB 图像的中级边缘轮廓特征，并将图像分辨率下采样 K 倍，输出 RGB



中分辨率特征图 RGB\_FP\_M, 其尺寸变为原始输入尺寸的 K 分之一。在本实施例中, L28~L44 由 8 个所述残差卷积块(记为 Residual\_Block\_4\_1 ~ 4\_8)和 1 个所述池化卷积块(Conv4\_pool)组成。K 取值为 2, M, N 的取值见图 3-1 的 L28~L44 层。

步骤 S340: 从所述 S320 获取 RGB\_FP\_M, 经过 DarkNet-53 网络的 L45~L52 层提取 RGB 图像的高级语义特征, 并将图像分辨率下采样 K 倍, 输出 RGB 低分辨率特征图 RGB\_FP\_L, 其尺寸变为原始输入尺寸的 K 分之一。在本实施例中, L45~L52 由 4 个所述残差卷积块(记为 Residual\_Block\_5\_1 ~ 5\_4)组成。K 取值为 2, M, N 的取值见图 3-1 的 L45~L52 层。

S3': 从所述 S2 获取编码规范化后的 Depth 图像, 采用非对称双流网络模型的 Depth 网络流在不同网络层级上分别提取 Depth 图像的通用、低级、中级和高级特征之后, 输出对应通用特征图以及高、中、低三种分辨率的 RGB 特征图, 分别记为 D\_FP\_C、D\_FP\_H、D\_FP\_M、D\_FP\_L, 并将 D\_FP\_H、D\_FP\_M、D\_FP\_L 输入到 S4'。在本实施例中, 非对称双流网络模型的 Depth 网络流是在 RGB 网络流 DarkNet-53 的基础上对模型进行剪枝获得, 下文简称之为 MiniDepth-30。MiniDepth-30 网络能更有效更清晰地提取深度图像的边缘轮廓等语义特征, 同时达到减少网络参数, 防止过拟合的效果。MiniDepth-30 的网络结构如图 2-2 所示。网络共包含 30 个卷积层, 其中网络的 L1~L10 层用于提取 Depth 图像的通用特征, 输出 D\_FP\_C; L11~L17 层用于提取 Depth 图像的低级色彩纹理特征, 输出 D\_FP\_H; L18~L24 层用于提取 Depth 图像的中级边缘轮廓特征, 输出 D\_FP\_M; L25~L30 层用于提取 Depth 图像的高级语义特征, 输出 D\_FP\_L。值得注意的是, 本实施例使用的 MiniDepth-30 模型仅为所述非对称双流网络的 Depth 网络流的一个具体实施例, 不限于上述 MiniDepth-30 模型, 下文仅以 MiniDepth-30 为例进行方法论述。

步骤 S310': 从所述 S2 获取编码规范化后的 Depth 图像, 经过 MiniDepth-30 网络的 L1~L10 层提取 RGB 图像的通用特征, 并将图像分辨率下采样 K 倍, 输出 Depth 通用特征图 D\_FP\_C, 其尺寸变为原始输入尺寸的 K 分之一。在本实施例中, MiniDepth-30 的 L1~L10 网络层与步骤 S310 中 DarkNet-53 的 L1~L10 网络层具有相同结构, K 取值为 8。

步骤 S320'：从所述步骤 S310' 获取 D\_FP\_C，经过 MiniDepth-30 网络的 L11~L17 层提取 Depth 图像的低级色彩纹理特征，并将图像分辨率下采样 K 倍，输出 Depth 高分辨率特征图 D\_FP\_H，其尺寸变为原始输入尺寸的 K 分之一。在本实施例中，L11~L17 由 3 个所述残差卷积块(记为 Residual\_Block\_D\_3\_1 ~ 3\_3) 和 1 个所述池化卷积块(Conv3\_D\_pool)组成。K 取值为 2，M, N 的取值见图 3-2 的 L11~L17 层。

步骤 S330'：从所述步骤 S320' 获取 D\_FP\_H，经过 MiniDepth-30 网络的 L18~L24 层提取 Depth 图像的中级边缘轮廓特征，并将图像分辨率下采样 K 倍，输出 Depth 中分辨率特征图 D\_FP\_M，其尺寸变为原始输入尺寸的 K 分之一。在本实施例中，L18~L24 由 3 个所述残差卷积块(记为 Residual\_Block\_D\_4\_1 ~ 4\_3)和 1 个所述池化卷积块(Conv4\_D\_pool)组成。K 取值为 2，M, N 的取值见图 3-1 的 L18~L24 层。

步骤 S340'：从所述步骤 S330' 获取 D\_FP\_M，经过 DarkNet-53 网络的 L25~L30 层提取 Depth 图像的高级语义特征，并将图像分辨率下采样 K 倍，输出 Depth 低分辨率特征图 D\_FP\_L，其尺寸变为原始输入尺寸的 K 分之一。在本实施例中，L25~L30 由 3 个所述残差卷积块(记为 Residual\_Block\_D\_5\_1 ~ 5\_3) 组成。K 取值为 2，M, N 的取值见图 3-1 的 L25~L30 层。

S4：从所述 S3 获取 RGB\_FP\_H、RGB\_FP\_M 和 RGB\_FP\_L，利用上采样拓展特征图尺寸，合并具有相同分辨率的 RGB 特征图的特征通道实现特征融合，输出特征融合之后的特征图 RGB\_FP\_H、RGB\_FP\_M 和 RGB\_FP\_L 到 S5。

步骤 S410：从所述步骤 S340 获取的 RGB\_FP\_L，上采样 M 倍后与所述步骤 S330 获取的 RGB\_FP\_M 进行通道合并，实现 RGB 网络深层的高级语义特征和中间层的中级边缘轮廓特征的互补融合，输出特征融合后的新特征图 RGB\_FP\_M。通道合并具体做法：RGB\_FP\_L 的通道数为 C1，RGB\_FP\_M 的通道数为 C2，二者通道合并 C1+C2 后获得 C3，C3 为特征融合后新特征图 RGB\_FP\_M 的通道数。在本实施例中 M 取值为 2，C1, C2, C3 取值分别为 256, 512, 768。

步骤 S420：从所述步骤 S410 获取特征融合后的新特征图 RGB\_FP\_M，上采样 M 倍后与所述步骤 S320 获取的 RGB\_FP\_H 进行通道合并，实现 RGB 网络深



层的高级语义特征、中间层的中级边缘轮廓特征以及浅层的低级色彩纹理特征的互补融合, 输出特征融合后的新特征图  $D\_FP\_H$ 。通道合并具体做法:  $RGB\_FP\_M$  的通道数为  $C1$ ,  $RGB\_FP\_H$  的通道数为  $C2$ , 二者通道合并  $C1+C2$  后获得  $C3$ ,  $C3$  为特征融合后新特征图  $RGB\_FP\_H$  的通道数。在本实施例中  $M$  取值为 2,  $C1$ ,  $C2$ ,  $C3$  取值分别为 128, 256, 384。

S4': 从所述 S3' 获取  $D\_FP\_H$ 、 $D\_FP\_M$ 、 $D\_FP\_L$ , 利用上采样拓展特征图尺寸, 合并具有相同分辨率的 Depth 特征图的特征通道实现特征融合, 输出特征融合之后的特征图  $D\_FP\_H$ 、 $D\_FP\_M$ 、 $D\_FP\_L$  到 S5。

步骤 S410': 从所述步骤 S340' 获取的  $D\_FP\_L$ , 上采样  $M$  倍后与所述步骤 S330' 获取的  $D\_FP\_M$  进行通道合并, 实现 Depth 网络深层的高级语义特征和中间层的中级边缘轮廓特征的互补融合, 输出特征融合后的新特征图  $D\_FP\_M$ 。通道合并具体做法:  $D\_FP\_L$  的通道数为  $C1$ ,  $D\_FP\_M$  的通道数为  $C2$ , 二者通道合并  $C1+C2$  后获得  $C3$ ,  $C3$  为特征融合后新特征图  $D\_FP\_M$  的通道数。在本实施例中  $M$  取值为 2,  $C1$ ,  $C2$ ,  $C3$  取值分别为 256, 512, 768。

步骤 S420': 从所述步骤 S410 获取特征融合后的新特征图  $D\_FP\_M$ , 上采样  $M$  倍后与所述步骤 S320' 获取的  $D\_FP\_H$  进行通道合并, 实现 Depth 网络深层的高级语义特征、中间层的中级边缘轮廓特征以及浅层的低级色彩纹理特征的互补融合, 输出特征融合后的新特征图  $D\_FP\_H$ 。通道合并具体做法:  $D\_FP\_M$  的通道数为  $C1$ ,  $D\_FP\_H$  的通道数为  $C2$ , 二者通道合并  $C1+C2$  后获得  $C3$ ,  $C3$  为特征融合后新特征图  $D\_FP\_H$  的通道数。在本实施例中  $M$  取值为 2,  $C1$ ,  $C2$ ,  $C3$  取值分别为 128, 256, 384。

S5: 从所述 S4 获取特征融合后新特征图  $RGB\_FP\_H$ 、 $RGB\_FP\_M$  和  $RGB\_FP\_L$ , 从 S4' 获取特征融合后新特征图  $D\_FP\_H$ 、 $D\_FP\_M$ 、 $D\_FP\_L$ , 在对应相等的分辨率上分别进行特征通道合并, 获得通道合并后的特征图, 分别记为  $Concat\_L$ 、 $Concat\_M$ 、 $Concat\_H$ , 然后应用通道重加权模块(以下简称为  $RW\_Module$ )分别对  $Concat\_L$ 、 $Concat\_M$ 、 $Concat\_H$  进行线性加权, 输出通道重加权后的高、中、低分辨率特征图, 分别记为  $RW\_H$ ,  $RW\_M$ ,  $RW\_L$ 。

步骤 S510: 从所述 S4 获取 RGB\_FP\_L 和 D\_FP\_L, 首先将 RGB\_FP\_L 和 D\_FP\_L 的特征通道进行合并获得 Concat\_L, 实现 RGB 和 Depth 在网络深层多模态信息的互补融合, 然后应用通道重加权模块 RW\_Module 对 Concat\_L 进行线性加权, 为每个特征通道赋予权重, 输出通道重加权后的特征图 RW\_L。以 RGB\_FP\_L 和 D\_FP\_L 的通道重加权为例, 本实施例提供的一种通道重加权模块的通用结构如图 4 所示。具体做法, RGB\_FP\_L 的通道数为 C1, D\_FP\_L 的通道数为 C2, 通道合并后的新特征图 Concat\_L 的通道数为 C3, 其中  $C3=C1+C2$ ; 然后对所述 Concat\_L 依次经过 1 个  $1\times 1$  的 Ave-Pooling 层、1 个由  $C3/s$  (s 为缩减步长) 个  $1\times 1$  卷积核组成的标准卷积层、1 个 C3 个  $1\times 1$  卷积核组成的标准卷积层和 1 个 Sigmoid 层, 获取 C3 个数值范围在 0~1 之间的权重值; 最后将获取的 C3 个权重值与所述 Concat\_L 的 C3 个特征通道相乘, 为每个特征通道赋予权重, 输出通道重加权后的 C3 个特征通道, 即 RW\_L。在本实施例中, C1、C2、C3 的取值分别为 1024, 1024, 2048, 缩减步长 s 的取值分别为 16。

步骤 S520: 从所述步骤 S410 获取 RGB\_FP\_M 和所述步骤 S410' 获取 D\_FP\_M, 首先将 RGB\_FP\_M 和 D\_FP\_M 的特征通道进行合并获得 Concat\_M, 实现 RGB 和 Depth 在网络中间层多模态信息的互补融合, 然后应用通道重加权模块 RW\_Module 对 Concat\_M 进行线性加权, 为每个特征通道赋予权重, 输出通道重加权后的特征图 RW\_M。在本实施例中, RGB\_FP\_M 和 D\_FP\_M 的通道重加权方式与所述步骤 S510 中 RGB\_FP\_L 和 D\_FP\_L 的通道重加权方式保持一致, 其中 C1、C2、C3 的取值分别为 512, 512, 1024, 缩减步长 s 的取值分别为 16。

步骤 S530: 从所述步骤 S420 获取 RGB\_FP\_H 和所述步骤 S420' 获取 D\_FP\_H, 首先将 RGB\_FP\_H 和 D\_FP\_H 的特征通道进行合并获得 Concat\_H, 实现 RGB 和 Depth 在网络浅层多模态信息的互补融合, 然后应用通道重加权模块 RW\_Module 对 Concat\_H 进行线性加权, 为每个特征通道赋予权重, 输出通道重加权后的特征图 RW\_H。在本实施例中, RGB\_FP\_H 和 D\_FP\_H 的通道重加权方式与所述步骤 S510 中 RGB\_FP\_L 和 D\_FP\_L 的通道重加权方式保持一致, 其中, C1、C2、C3 的取值分别为 256, 256, 512, 缩减步长 s 的取值分别为 16。

S6: 从所述 S5 获取通道重加权后的特征图 RW\_L, RW\_M, RW\_H, 分别进行分类和边框坐标回归, 获得较大、中等以及较小尺寸人员的预测结果, 对上述三种



不同尺度的预测结果进行汇总,采用非极大值抑制(以下简称 NMS)算法剔除重叠目标边框,输出最终保留的人员检测结果,即人员的类别置信分数 $C_i$ 和预测矩形边框 $(x_i^1, y_i^1, x_i^2, y_i^2)$ ,  $i=1, 2, \dots, N$ 。在本实施例中,  $i$  代表人员的 ID 编号,  $N$  为在当前图像中保留的人员检测结果总数。 $x_i^1, y_i^1, x_i^2, y_i^2$  分别代表所有包含人员的矩形边框的左上角横坐标、左上角纵坐标、右下角横坐标以及右下角纵坐标。

步骤 S610: 从所述步骤 S510 获取通道重加权后的低分辨率特征图  $RW\_L$ , 传输到 SoftMax 分类层和坐标回归层, 输出在低分辨率特征图下预测较大尺寸人员的类别置信分数 $C_{i_L}$ 和矩形边框的左上角、右下角坐标 $(x_i^1, y_i^1, x_i^2, y_i^2)_L$ , 其中下标  $L$  表示在低分辨率特征图下的预测结果。

步骤 S620: 从所述步骤 S520 获取通道重加权后的低分辨率特征图  $RW\_M$ , 传输到 SoftMax 分类层和坐标回归层, 输出在中分辨率特征图下预测中等尺寸人员的类别置信分数 $C_{i_M}$ 和矩形边框的左上角、右下角坐标 $(x_i^1, y_i^1, x_i^2, y_i^2)_M$ , 其中下标  $M$  表示在中分辨率特征图下的预测结果。

步骤 S630: 从所述步骤 S530 获取通道重加权后的高分辨率特征图  $RW\_H$ , 传输到 SoftMax 分类层和坐标回归层, 输出在高分辨率特征图下预测较小尺寸人员的类别置信分数 $C_{i_H}$ 和矩形边框的左上角、右下角坐标 $(x_i^1, y_i^1, x_i^2, y_i^2)_H$ , 其中下标  $H$  表示在高分辨率特征图下的预测结果。

步骤 S640: 从所述步骤 S610、S620 和 S630 获取较大、中等和较小尺寸人员的类别置信分数 $C_{i_L}, C_{i_M}, C_{i_H}$ 和矩形边框左上右下坐标 $(x_i^1, y_i^1, x_i^2, y_i^2)_L, (x_i^1, y_i^1, x_i^2, y_i^2)_M, (x_i^1, y_i^1, x_i^2, y_i^2)_H$ , 对三种尺度的预测结果进行汇总, 然后采用 NMS 算法去除重叠的目标边框, 输出最终保留的人员检测结果。即人员的类别置信分数 $C_i$ 和预测矩形边框 $(x_i^1, y_i^1, x_i^2, y_i^2)$ ,  $i=1, 2, \dots, N$ 。NMS 算法流程图如图 5 所示。

NMS 算法步骤如下:

步骤 S640-1: 从所述步骤 S610、S620 和 S630 获取较大、中等、较小尺寸的人员类别置信分数 $C_{i_L}, C_{i_M}, C_{i_H}$ 和矩形边框左上右下坐标 $(x_i^1, y_i^1, x_i^2, y_i^2)_L, (x_i^1, y_i^1, x_i^2, y_i^2)_M, (x_i^1, y_i^1, x_i^2, y_i^2)_H$ , 对三种尺度的预测结果进行汇总, 利用置信阈值对预测框进行筛选, 保留类别置信分数大于置信阈值的预测边框, 将其加入到预测列表中。在本实施例中置信阈值设置为 0.3。

步骤 S640-2: 从所述步骤 S640-1 获取的预测列表, 对预测列表中未处理的预测边框按照置信分数降序排列, 输出降序排列后的预测列表。

步骤 S640-3: 从所述步骤 S640-2 中获取降序排列后的预测列表, 选取最大置信分数对应的边框作为当前基准边框, 将当前基准边框的类别置信分数和边框坐标加入到最终结果列表中, 并将基准边框从预测列表中剔除, 其余所有预测边框与当前基准边框计算交并比(IoU)。

步骤 S640-4: 从所述步骤 S640-3 获取预测列表以及预测列表中所有边框与基准边框的 IoU 值, 若当前边框的 IoU 大于预设 NMS 阈值, 则认为其与基准边框为重复目标, 将其从预测边框列表中剔除, 否则保留当前边框。输出筛选后的预测列表。

步骤 S640-5: 从所述步骤 S640-4 获取筛选后的预测列表, 若预测列表中所有边框都处理完毕即预测边框为空, 则算法结束, 返回最终结果列表; 反之, 当前预测列表中仍存在未处理的边框, 则返回步骤 S640-2 重复算法流程。

步骤 S640-6: 对所述步骤 S640-5, 当预测列表中不存在未处理的预测边框时, 输出最终结果列表为最终保留的人员检测结果。

#### 参考文献:

- [1] Neubeck A, Gool L V. Efficient Non-Maximum Suppression[C]// International Conference on Pattern Recognition. 2006.
- [2] Zhang Y, Funkhouser T. Deep Depth Completion of a Single RGB-D Image[J]. 2018.
- [3] Gupta S, Girshick R, Arbeláez P, et al. Learning Rich Features from RGB-D Images for Object Detection and Segmentation[C]// 2014.
- [4] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. 2018.



1、一种基于非对称双流网络的 RGB-D 多模态融合人员检测方法，其特征在于：包含 RGBD 图像采集，深度图像预处理，RGB 特征提取和 Depth 特征提取，RGB 多尺度融合和 Depth 多尺度融合，多模态特征通道重加权以及多尺度人员预测。

2、根据权利要求 1 所述的一种基于非对称双流网络的 RGB-D 多模态融合人员检测方法，其特征在于：S1RGBD 图像采集；

利用具有同时拍摄 RGB 图像和深度图像功能的相机获取原始 RGB 图像和 Depth 图像，并对 RGB 和 Depth 图像进行匹配分组，每组图像由一张 RGB 图像和同场景下捕获的 Depth 图像组成，输出分组匹配后的 RGB 和 Depth 图像；原始 RGB 图像和 Depth 图像也能够从公开 RGBD 数据集获取。

3、根据权利要求 2 所述的一种基于非对称双流网络的 RGB-D 多模态融合人员检测方法，其特征在于：S2 深度图像预处理；

从 S1 的 RGBD 图像采集获取分组匹配后的 Depth 图像，首先消除 Depth 图像的部分噪声，然后进行空洞填充，最后将单通道 Depth 图像重新编码为三个通道图像，并将三个通道的图像数值重新规范化到 0-255，输出编码规范化后的 Depth 图像。

4、根据权利要求 3 所述的一种基于非对称双流网络的 RGB-D 多模态融合人员检测方法，其特征在于：S3RGB 特征提取和 Depth 特征提取；

从所述 S1 的 RGBD 图像采集获取原始 RGB 图像，输入到 RGB 特征提取，进行下采样特征提取，输出 RGB 图像的高、中、低分辨率特征图，分别记为 RGB\_FP\_H、RGB\_FP\_M、RGB\_FP\_L，代表 RGB 图像的低级色彩纹理、中级边缘轮廓和高级语义特征表示；从深度图像预处理获取编码规范化后的 Depth 图像，输入到 Depth 特征提取，进行下采样特征提取，输出 Depth 图像的高、中、低分辨率特征图，分别记为 D\_FP\_H、D\_FP\_M、D\_FP\_L，代表 Depth 图像的低级色彩纹理、中级边缘轮廓和高级语义特征表示；RGB 网络流和 Depth 网络流是对称结构的，即 RGB 网络流和 Depth 网络流的结构完全相同；设计非对称双流卷积神经网络模型提取 RGB 图像和 Depth 图像特征；DarkNet-53 和 MiniDepth-30 分别代表 RGB 网络流和 Depth 网络流，DarkNet-53 和 MiniDepth-30 的网络结构具有非对称的特性。

5、根据权利要求4所述的一种基于非对称双流网络的RGB-D多模态融合人员检测方法，其特征在于：S4RGB多尺度融合和Depth多尺度融合；

从RGB特征提取获取RGB特征图RGB\_FP\_H、RGB\_FP\_M、RGB\_FP\_L输入到RGB多尺度融合，首先将获取的RGB\_FP\_L通过上采样层拓展到与RGB\_FP\_M相同尺寸，然后与RGB\_FP\_M进行通道合并，实现RGB网络深层的高级语义特征与中间层的中级边缘轮廓特征的互补融合，输出通道合并后的新特征图RGB\_FP\_M；然后对输出通道合并后的新特征图RGB\_FP\_M，通过上采样层拓展到与RGB\_FP\_H相同尺寸，与RGB\_FP\_H进行通道合并，实现RGB网络深层的高级语义特征、中间层的中级边缘轮廓特征以及浅层的低级色彩纹理特征的互补融合，输出通道合并后的新特征图RGB\_FP\_H；从Depth特征提取获取Depth特征图D\_FP\_H、D\_FP\_M、D\_FP\_L输入到Depth多尺度融合，与RGB多尺度融合执行同样的操作；最终Depth多尺度融合的输出为原始输入RGB\_FP\_L、通道合并后的新特征图RGB\_FP\_M和RGB\_FP\_H；Depth多尺度融合的输出为原始输入D\_FP\_L、通道合并后的新特征图D\_FP\_M和D\_FP\_H。

6、根据权利要求5所述的一种基于非对称双流网络的RGB-D多模态融合人员检测方法，其特征在于：S5多模态特征通道重加权；

从RGB多尺度融合获取RGB特征图RGB\_FP\_L、RGB\_FP\_M、RGB\_FP\_H和从Depth多尺度融合获取Depth特征图D\_FP\_L、D\_FP\_M、D\_FP\_H，按照分辨率分组输入到多模态特征通道重加权中对应相同分辨率的通道重加权结构中，实现更有效的RGB与Depth的多模态特征融合，提高处理多种限制场景下的检测鲁棒性；具体做法以RGB\_FP\_L与D\_FP\_L通道重加权为例，从RGB多尺度融合获取RGB\_FP\_L以及从Depth多尺度融合获取D\_FP\_L，首先进行通道合并，获得通道合并后的特征图记为Concat\_L；然后应用通道重加权模块简称为RW\_Module对Concat\_L的特征通道进行线性加权，为每个特征通道赋予权重，输出通道重加权后的特征图记为RW\_L；RGB\_FP\_M与D\_FP\_M，RGB\_FP\_H与D\_FP\_H的通道重加权采用与上述RGB\_FP\_L和D\_FP\_L相同方式完成；最终多模态特征通道重加权输出通道重加权后的低、中、高分辨率特征图，分别记为RW\_L, RW\_M, RW\_H。

7、根据权利要求6所述的一种基于非对称双流网络的RGB-D多模态融合人员检测方法，其特征在于：S6多尺度人员预测；



从所述 S5 的多模态特征通道重加权获取通道重加权后的特征图 RW\_L, RW\_M, RW\_H, 分别输入到多尺度人员预测中对应的预测分支中进行分类和边框坐标回归, 获得较大、中等以及较小尺寸人员的预测结果; 由于特征图分辨率不同, 特征图上每个预测点对应的感受野也不同; RW\_L 上的每个预测点具有较大感受野, 用来预测图像中的较大目标; RW\_M 上的每个预测点具有中等感受野, 用来预测图像中的中等目标; RW\_H 上的每个预测点具有较小感受野, 用来预测图像中的较小目标; 对上述三种不同尺度的预测结果进行汇总, 采用非极大值抑制算法剔除重叠目标边框, 输出最终保留的人员检测结果, 即人员的类别置信分数  $C_i$  和预测矩形边框  $(x_i^1, y_i^1, x_i^2, y_i^2)$ ,  $i=1, 2, \dots, N$ ;  $i$  代表人员的 ID 编号,  $N$  为在当前图像中保留的人员检测结果总数;  $x_i^1, y_i^1, x_i^2, y_i^2$  分别代表所有包含人员的矩形边框的左上角横坐标、左上角纵坐标、右下角横坐标以及右下角纵坐标。

8、根据权利要求 2 所述的一种基于非对称双流网络的 RGB-D 多模态融合人员检测方法, 其特征在于:

步骤 S110: 利用具有同时拍摄 RGB 图像和深度图像功能的相机获取原始 RGB 图像, 原始 RGB 图像也可以从公开 RGBD 数据集获取;

步骤 S120: 从所述步骤 S110 同步获取与 RGB 图像匹配的 Depth 图像, 并对 RGB 和 Depth 图像进行分组, 每组图像由一张 RGB 图像和同场景下捕获的深度图像组成, 输出分组匹配后的 Depth 图像。

9、根据权利要求 2 所述的一种基于非对称双流网络的 RGB-D 多模态融合人员检测方法, 其特征在于: 获取较大、中等和较小尺寸人员的类别置信分数  $C_{i_L}, C_{i_M}, C_{i_H}$  和矩形边框左上右下坐标

$$(x_i^1, y_i^1, x_i^2, y_i^2)_L, (x_i^1, y_i^1, x_i^2, y_i^2)_M, (x_i^1, y_i^1, x_i^2, y_i^2)_H,$$

对三种尺度的预测结果进行汇总, 然后采用 NMS 算法去除重叠的目标边框, 输出最终保留的人员检测结果; 即人员的类别置信分数  $C_i$  和预测矩形边框

$$(x_i^1, y_i^1, x_i^2, y_i^2), i=1, 2, \dots, N。$$

10、根据权利要求 9 所述的一种基于非对称双流网络的 RGB-D 多模态融合人员检测方法, 其特征在于:

NMS 算法步骤如下:

步骤 S640-1: 从获取较大、中等、较小尺寸的人员类别置信分数  $C_{i_L}, C_{i_M}, C_{i_H}$  和矩形边框左上右下坐标  $(x_i^1, y_i^1, x_i^2, y_i^2)_L, (x_i^1, y_i^1, x_i^2, y_i^2)_M, (x_i^1, y_i^1, x_i^2, y_i^2)_H$ , 对

三种尺度的预测结果进行汇总，利用置信阈值对预测框进行筛选，保留类别置信分数大于置信阈值的预测边框，将其加入到预测列表中；置信阈值设置为 0.3；

步骤 S640-2：从所述步骤 S640-1 获取的预测列表，对预测列表中未处理的预测边框按照置信分数降序排列，输出降序排列后的预测列表；

步骤 S640-3：从所述步骤 S640-2 中获取降序排列后的预测列表，选取最大置信分数对应的边框作为当前基准边框，将当前基准边框的类别置信分数和边框坐标加入到最终结果列表中，并将基准边框从预测列表中剔除，其余所有预测边框与当前基准边框计算交并比 IoU；

步骤 S640-4：从所述步骤 S640-3 获取预测列表以及预测列表中所有边框与基准边框的 IoU 值，若当前边框的 IoU 大于预设 NMS 阈值，则认为其与基准边框为重复目标，将其从预测边框列表中剔除，否则保留当前边框；输出筛选后的预测列表；

步骤 S640-5：从所述步骤 S640-4 获取筛选后的预测列表，若预测列表中所有边框都处理完毕即预测边框为空，则算法结束，返回最终结果列表；反之，当前预测列表中仍存在未处理的边框，则返回步骤 S640-2 重复算法流程；

步骤 S640-6：对所述步骤 S640-5，当预测列表中不存在未处理的预测边框时，输出最终结果列表为最终保留的人员检测结果。



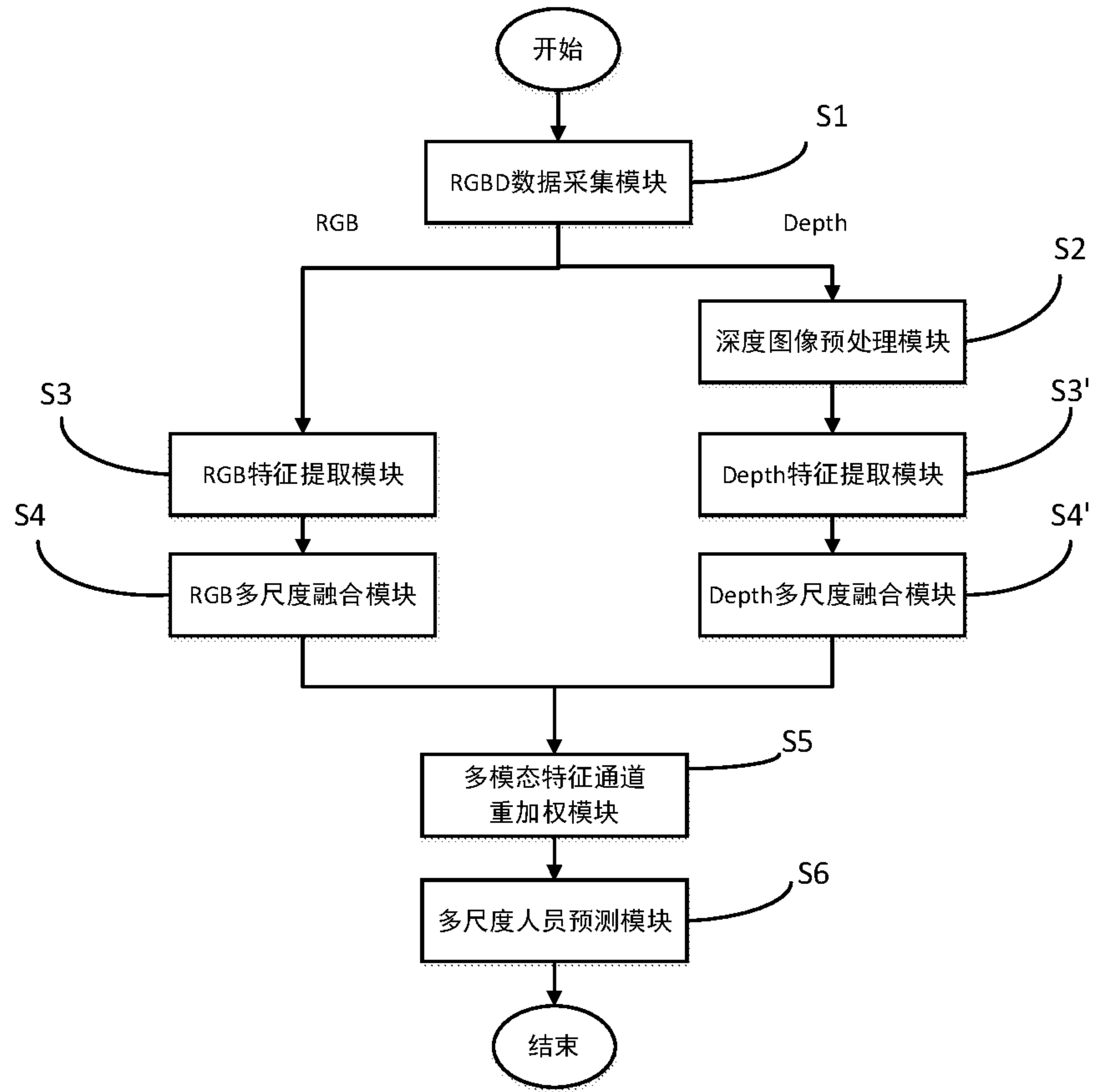


图 1

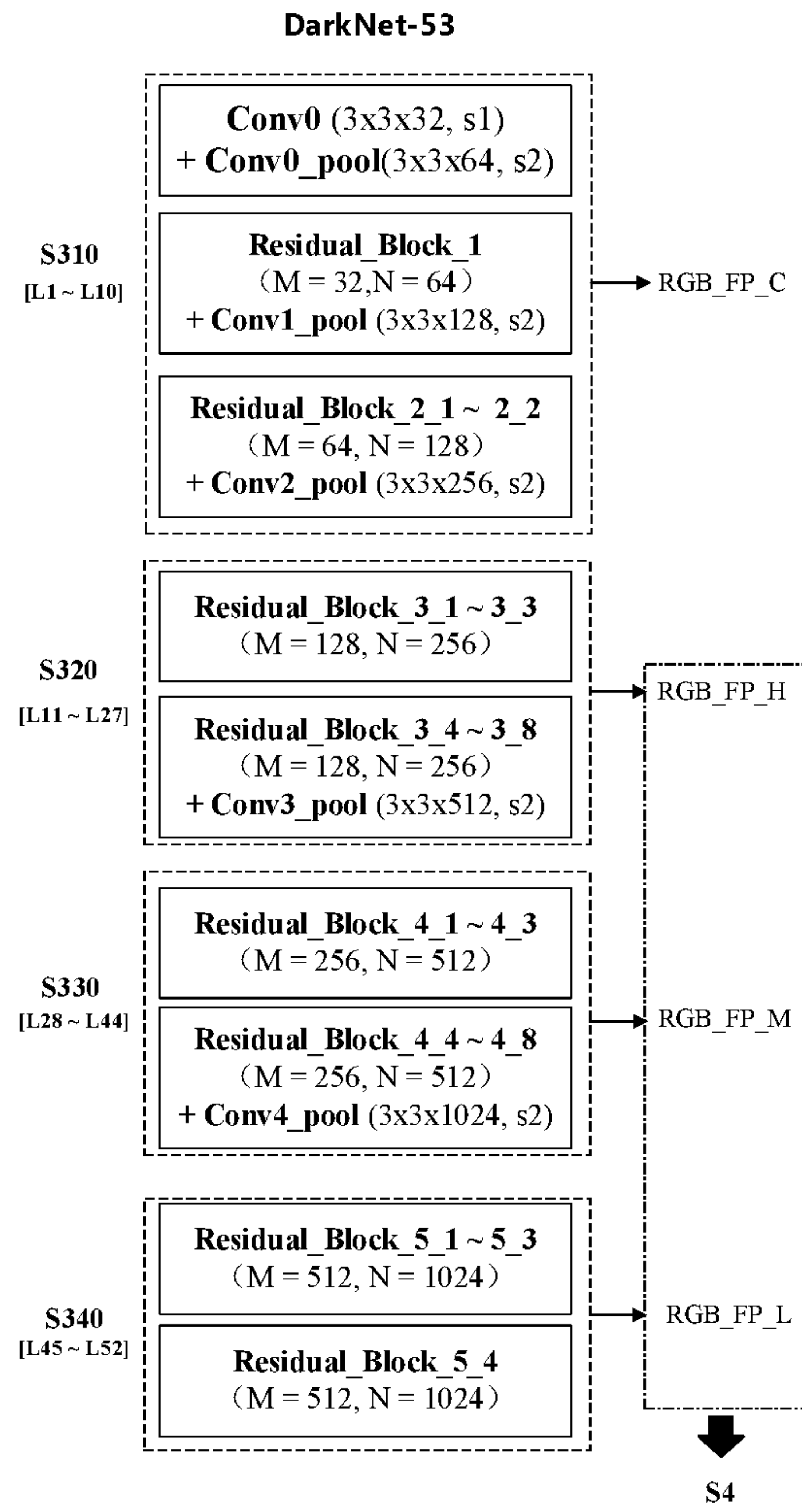


图 2-1



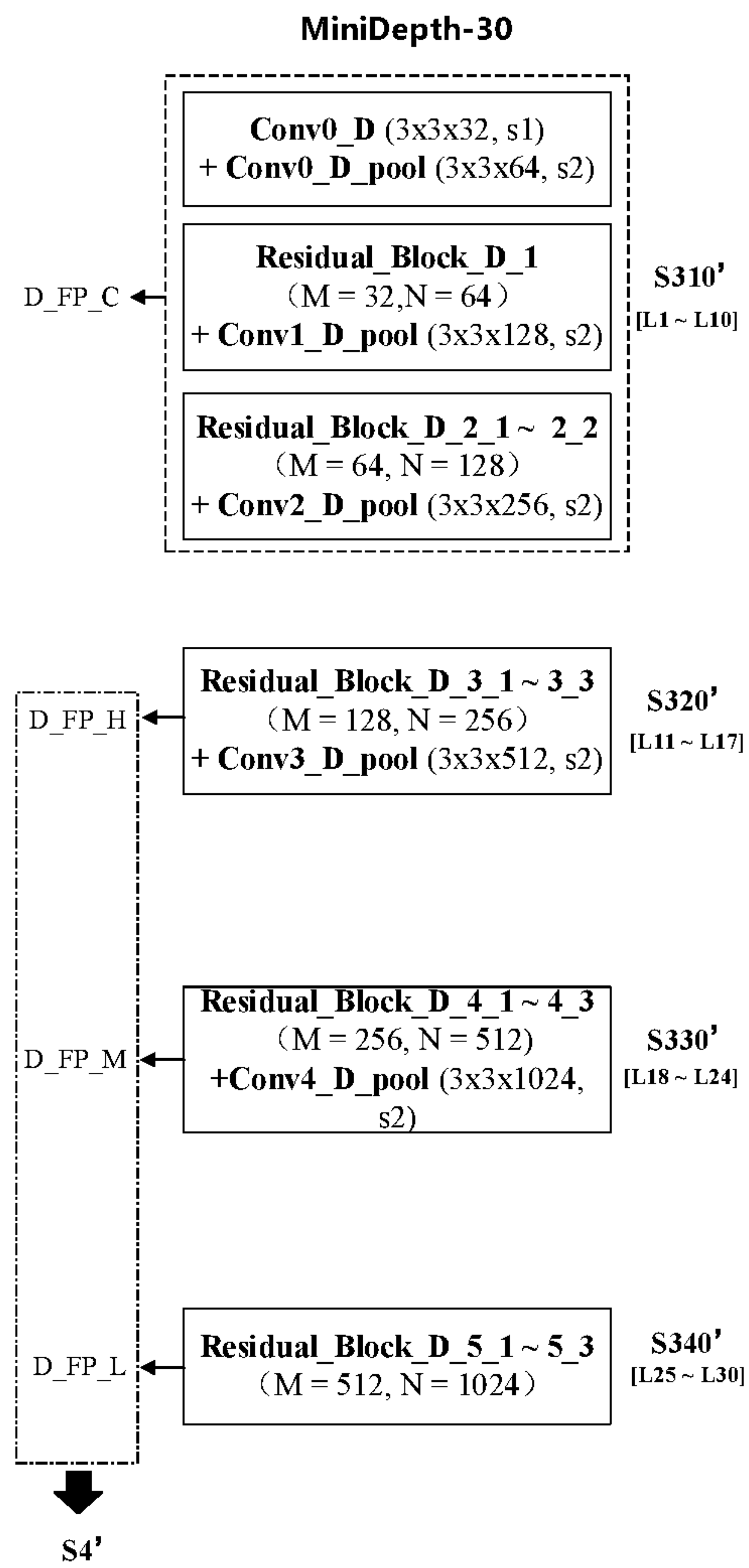


图 2-2

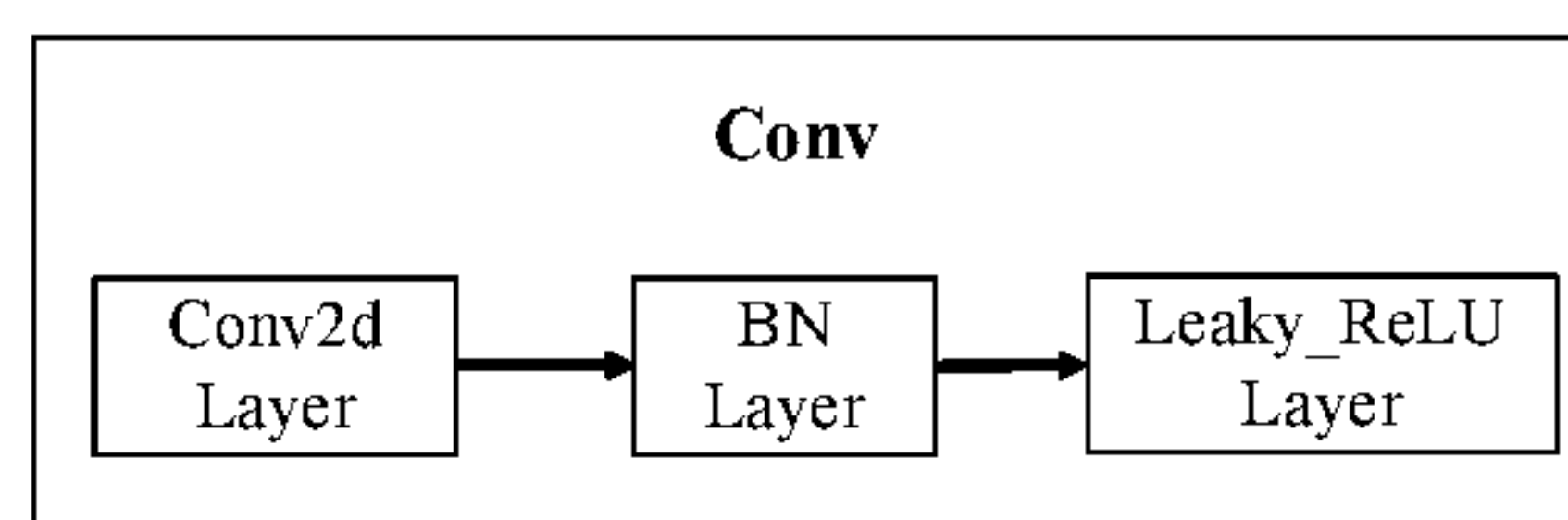


图 2-3

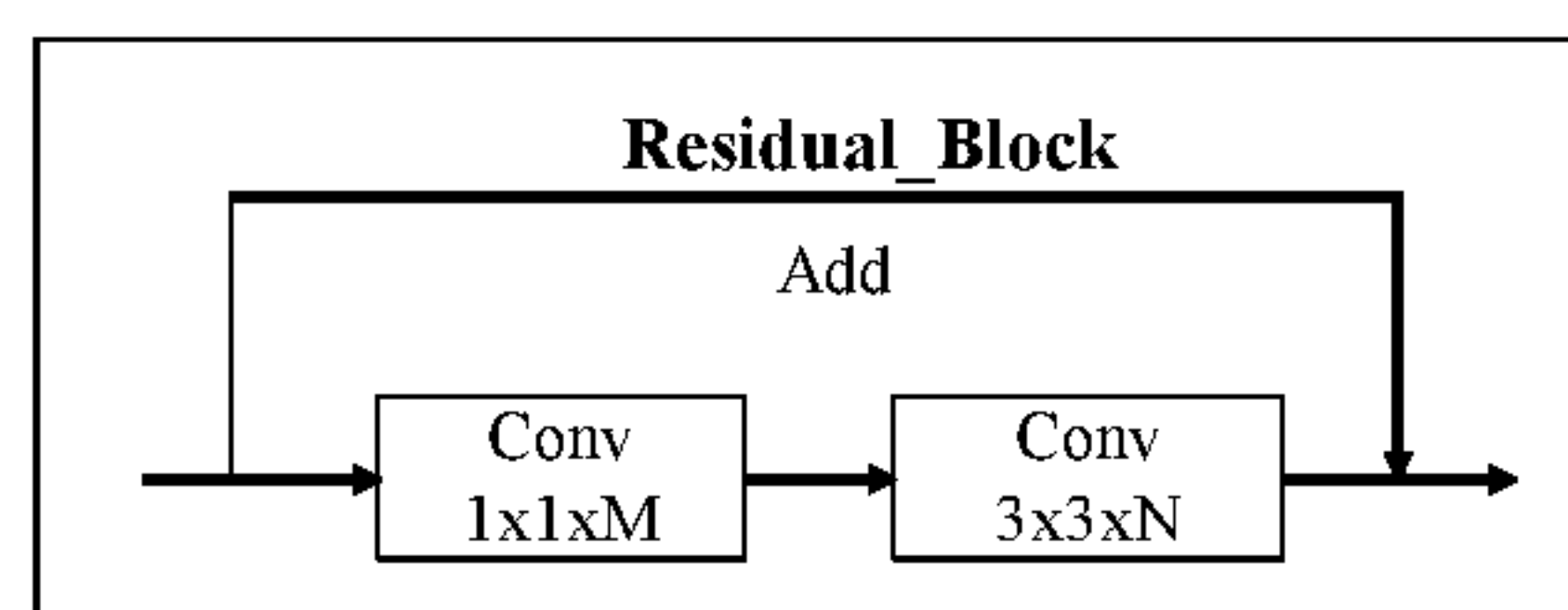


图 2-4

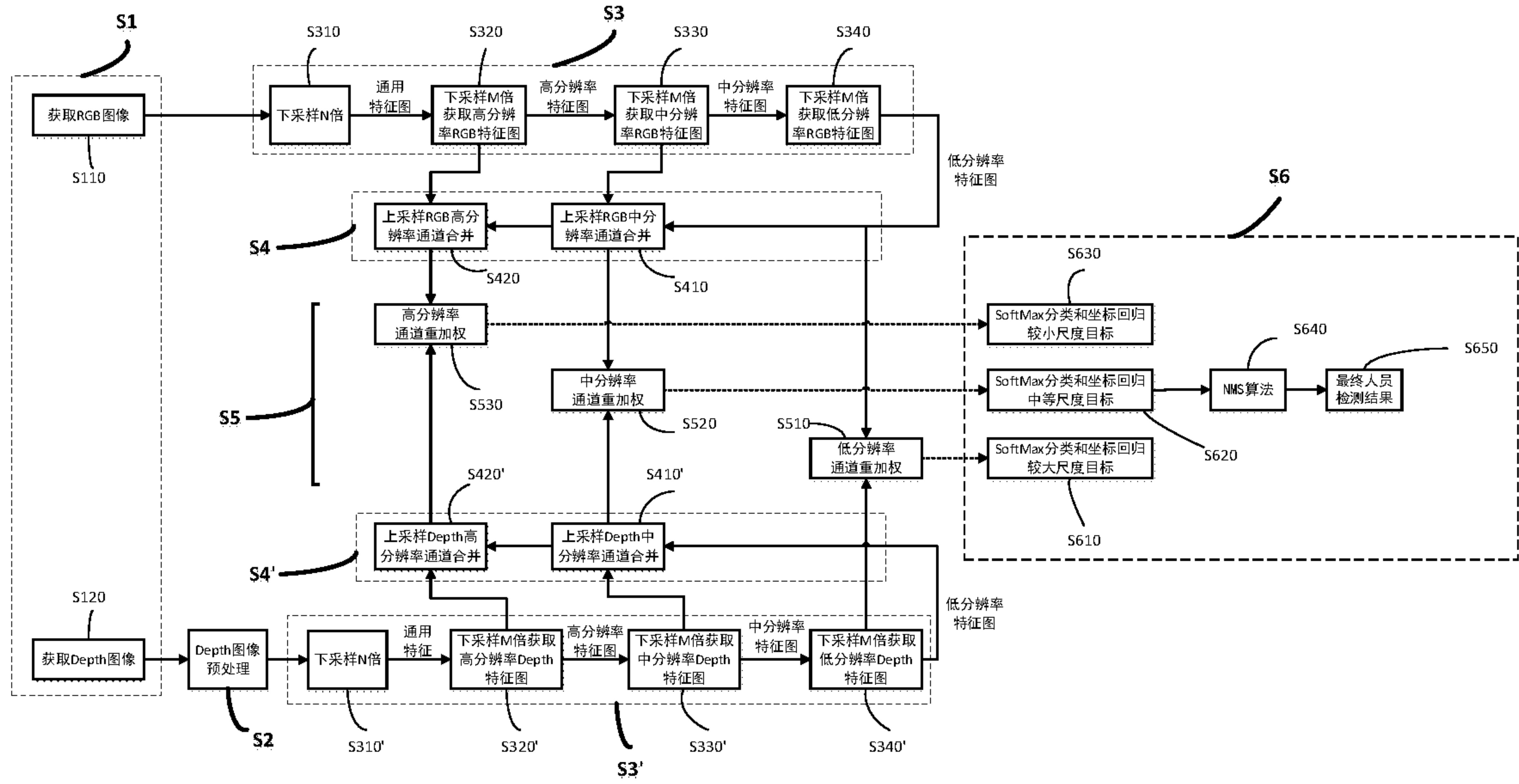


图 3

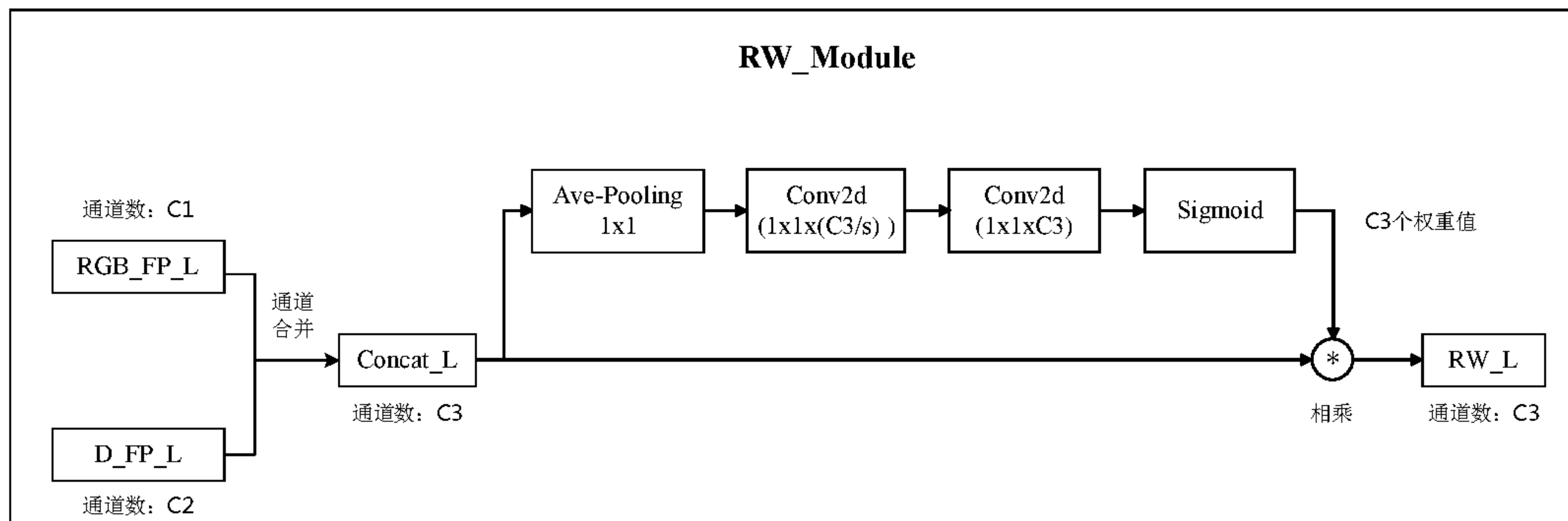


图 4



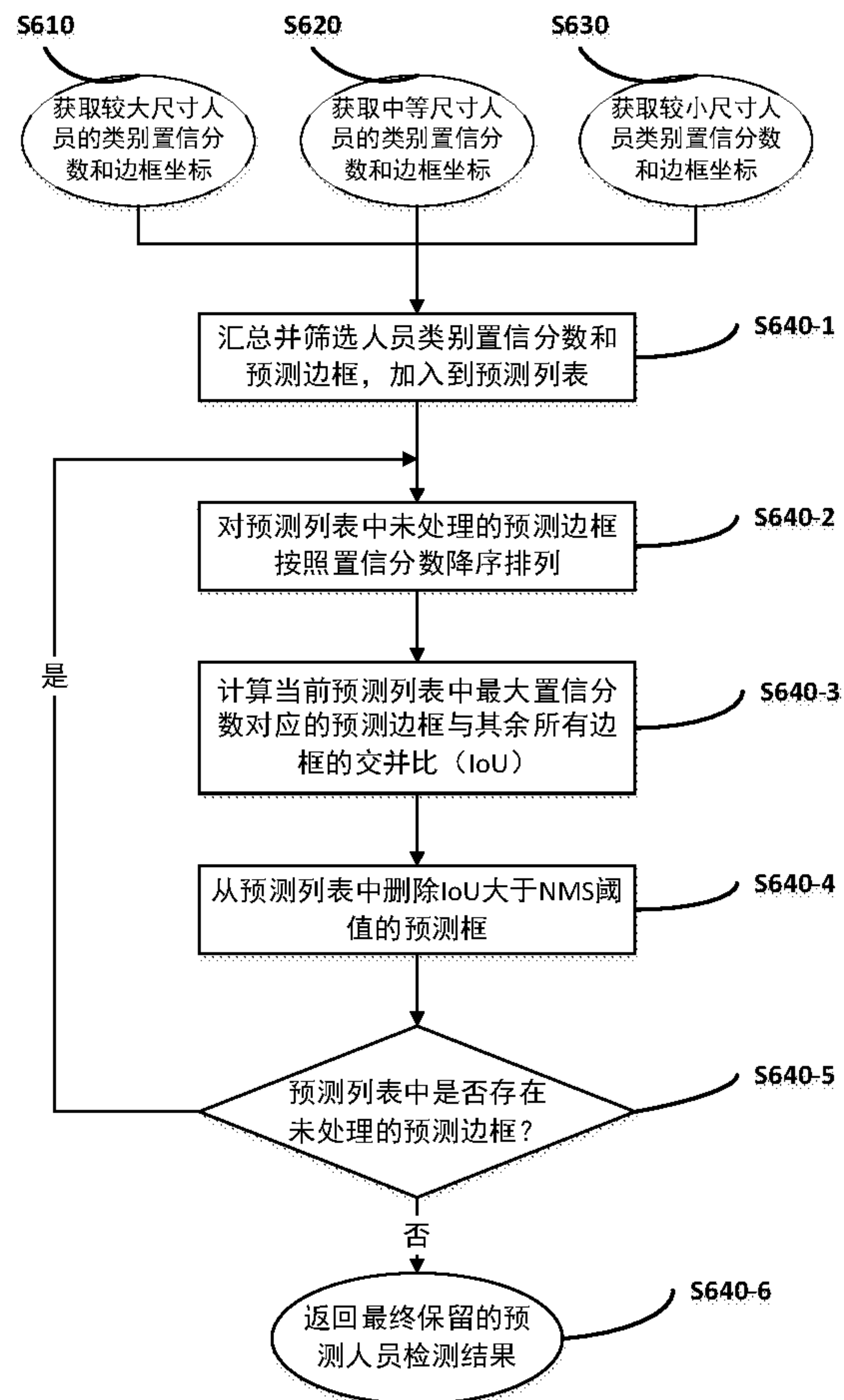


图 5 NMS 算法流程图

## INTERNATIONAL SEARCH REPORT

International application No.

**PCT/CN2020/080991**

|   |  |   |
|---|--|---|
| <b>A. CLASSIFICATION OF SUBJECT MATTER</b><br>G06K 9/00(2006.01)i; G06K 9/46(2006.01)i; G06N 3/02(2006.01)i<br><br>According to International Patent Classification (IPC) or to both national classification and IPC  |  |   |
| <b>B. FIELDS SEARCHED</b>   |  |   |
| Minimum documentation searched (classification system followed by classification symbols)<br>G06K9/; 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)<br>CNABS; CNTXT; CNKI; DWPI; SIPOABS: 人员, 行人, 人体, 检测, RGBD, RGB-D, RGB, 彩色, 深度, 立体, 多尺度, 融合, 高, 中, 低, 特征, human, pedestrain, detect, RGBD, RGB-D, red green blue depth, RGB, red green blue, depth, multiscale, merge, combine, high, middle, low, feature, character  |  |   |
| <b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>   |  |   |
| Category*   | Citation of document, with indication, where appropriate, of the relevant passages   | Relevant to claim No.   |
| Y   | CN 109543697 A (NORTHWESTERN POLYTECHNICAL UNIVERSITY) 29 March 2019 (2019-03-29)<br>description paragraphs 21-31, figures 1-3 | 1-10  |
| Y   | CN 107045630 A (HANGZHOU SILANMU TECHNOLOGY CO., LTD.) 15 August 2017 (2017-08-15)<br>description paragraphs 52-91, figure 1   | 1-10  |
| Y   | CN 109598301 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 09 April 2019 (2019-04-09)<br>description, paragraphs 42-117            | 4-7   |
| Y   | CN 108734210 A (ZHEJIANG UNIVERSITY OF TECHNOLOGY) 02 November 2018 (2018-11-02)<br>description paragraphs 58-61, figures 1-2  | 1-10  |
| A   | WO 2019162241 A1 (ROBERT BOSCH GMBH) 29 August 2019 (2019-08-29)<br>entire document  | 1-10  |
| <input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.   |  |   |
| <p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&amp;” document member of the same patent family</p> |  |   |
| Date of the actual completion of the international search<br><b>22 July 2020</b>  |  | Date of mailing of the international search report<br><b>03 August 2020</b> |
| Name and mailing address of the ISA/CN<br><b>China National Intellectual Property Administration (ISA/CN)<br/>No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing<br/>100088<br/>China</b>   |  | Authorized officer  |
| Facsimile No. (86-10)62019451   |  | Telephone No.   |



**INTERNATIONAL SEARCH REPORT**

International application No.

**PCT/CN2020/080991**

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                     | Relevant to claim No. |
|-----------|--|-----------------------|
| A         | US 2014107842 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 17 April 2014 (2014-04-17)<br>entire document | 1-10                  |

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

|   |
|---|
| International application No.<br><b>PCT/CN2020/080991</b> |
|---|

| Patent document cited in search report |            |    | Publication date (day/month/year) | Patent family member(s) | Publication date (day/month/year) |
|--|------------|----|-----------------------------------|-------------------------|-----------------------------------|
| CN                                     | 109543697  | A  | 29 March 2019                     | None                    |                                   |
| CN                                     | 107045630  | A  | 15 August 2017                    | None                    |                                   |
| CN                                     | 109598301  | A  | 09 April 2019                     | None                    |                                   |
| CN                                     | 108734210  | A  | 02 November 2018                  | None                    |                                   |
| WO                                     | 2019162241 | A1 | 29 August 2019                    | TW                      | 201944283 A 16 November 2019      |
| US                                     | 2014107842 | A1 | 17 April 2014                     | KR                      | 20140049152 A 25 April 2014       |
|  |            |    |                                   | US                      | 9201425 B2 01 December 2015       |



国际检索报告

国际申请号

PCT/CN2020/080991

| <p><b>A. 主题的分类</b></p> <p>G06K 9/00(2006.01) i; G06K 9/46(2006.01) i; G06N 3/02(2006.01) i</p> <p>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</p>  |   |   |     |                   |         |   |   |      |   |  |      |   |   |     |   |   |      |   |  |      |   |   |      |
|--|---|---|-----|-------------------|---------|---|---|------|---|--|------|---|---|-----|---|---|------|---|--|------|---|---|------|
| <p><b>B. 检索领域</b></p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>G06K9/; G06N</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>CNABS;CNTXT;CNKI;DWPI;SIPOABS:人员, 行人, 人体, 检测, RGBD, RGB-D, RGB, 彩色, 深度, 立体, 多尺度, 融合, 高, 中, 低, 特征, human, pedestrain, detect, RGBD, RGB-D, red green blue depth, RGB, red green blue, depth, multiscale, merge, combine, high, middle, low, feature, character</p>   |   |   |     |                   |         |   |   |      |   |  |      |   |   |     |   |   |      |   |  |      |   |   |      |
| <p><b>C. 相关文件</b></p> <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>Y</td> <td>CN 109543697 A (西北工业大学) 2019年 3月 29日 (2019 - 03 - 29)<br/>说明书第21-31段, 图1-3</td> <td>1-10</td> </tr> <tr> <td>Y</td> <td>CN 107045630 A (杭州司兰木科技有限公司) 2017年 8月 15日 (2017 - 08 - 15)<br/>说明书第52-91段, 图1</td> <td>1-10</td> </tr> <tr> <td>Y</td> <td>CN 109598301 A (腾讯科技深圳有限公司) 2019年 4月 9日 (2019 - 04 - 09)<br/>说明书第42-117段</td> <td>4-7</td> </tr> <tr> <td>Y</td> <td>CN 108734210 A (浙江工业大学) 2018年 11月 2日 (2018 - 11 - 02)<br/>说明书第58-61段, 图1-2</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>WO 2019162241 A1 (BOSCH GMBH ROBERT) 2019年 8月 29日 (2019 - 08 - 29)<br/>全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>US 2014107842 A1 (KOREA ELECTRONICS TELECOMM) 2014年 4月 17日 (2014 - 04 - 17)<br/>全文</td> <td>1-10</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 | Y | CN 109543697 A (西北工业大学) 2019年 3月 29日 (2019 - 03 - 29)<br>说明书第21-31段, 图1-3 | 1-10 | Y | CN 107045630 A (杭州司兰木科技有限公司) 2017年 8月 15日 (2017 - 08 - 15)<br>说明书第52-91段, 图1 | 1-10 | Y | CN 109598301 A (腾讯科技深圳有限公司) 2019年 4月 9日 (2019 - 04 - 09)<br>说明书第42-117段 | 4-7 | Y | CN 108734210 A (浙江工业大学) 2018年 11月 2日 (2018 - 11 - 02)<br>说明书第58-61段, 图1-2 | 1-10 | A | WO 2019162241 A1 (BOSCH GMBH ROBERT) 2019年 8月 29日 (2019 - 08 - 29)<br>全文 | 1-10 | A | US 2014107842 A1 (KOREA ELECTRONICS TELECOMM) 2014年 4月 17日 (2014 - 04 - 17)<br>全文 | 1-10 |
| 类型*  | 引用文件, 必要时, 指明相关段落   | 相关的权利要求                                     |     |                   |         |   |   |      |   |  |      |   |   |     |   |   |      |   |  |      |   |   |      |
| Y  | CN 109543697 A (西北工业大学) 2019年 3月 29日 (2019 - 03 - 29)<br>说明书第21-31段, 图1-3         | 1-10  |     |                   |         |   |   |      |   |  |      |   |   |     |   |   |      |   |  |      |   |   |      |
| Y  | CN 107045630 A (杭州司兰木科技有限公司) 2017年 8月 15日 (2017 - 08 - 15)<br>说明书第52-91段, 图1      | 1-10  |     |                   |         |   |   |      |   |  |      |   |   |     |   |   |      |   |  |      |   |   |      |
| Y  | CN 109598301 A (腾讯科技深圳有限公司) 2019年 4月 9日 (2019 - 04 - 09)<br>说明书第42-117段           | 4-7   |     |                   |         |   |   |      |   |  |      |   |   |     |   |   |      |   |  |      |   |   |      |
| Y  | CN 108734210 A (浙江工业大学) 2018年 11月 2日 (2018 - 11 - 02)<br>说明书第58-61段, 图1-2         | 1-10  |     |                   |         |   |   |      |   |  |      |   |   |     |   |   |      |   |  |      |   |   |      |
| A  | WO 2019162241 A1 (BOSCH GMBH ROBERT) 2019年 8月 29日 (2019 - 08 - 29)<br>全文          | 1-10  |     |                   |         |   |   |      |   |  |      |   |   |     |   |   |      |   |  |      |   |   |      |
| A  | US 2014107842 A1 (KOREA ELECTRONICS TELECOMM) 2014年 4月 17日 (2014 - 04 - 17)<br>全文 | 1-10  |     |                   |         |   |   |      |   |  |      |   |   |     |   |   |      |   |  |      |   |   |      |
| <p><input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p>  |   |   |     |                   |         |   |   |      |   |  |      |   |   |     |   |   |      |   |  |      |   |   |      |
| <p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“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)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“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</p> <p>“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了解发明之理论或原理的在后文件</p> <p>“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</p> <p>“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</p> <p>“&amp;” 同族专利的文件</p>   |   |   |     |                   |         |   |   |      |   |  |      |   |   |     |   |   |      |   |  |      |   |   |      |
| <p>国际检索实际完成的日期</p> <p>2020年 7月 22日</p>   |   | <p>国际检索报告邮寄日期</p> <p>2020年 8月 3日</p>        |     |                   |         |   |   |      |   |  |      |   |   |     |   |   |      |   |  |      |   |   |      |
| <p>ISA/CN的名称和邮寄地址</p> <p>中国国家知识产权局(ISA/CN)<br/>中国北京市海淀区蓟门桥西土城路6号 100088</p> <p>传真号 (86-10)62019451</p>   |   | <p>授权官员</p> <p>许菲菲</p> <p>电话号码 62411752</p> |     |                   |         |   |   |      |   |  |      |   |   |     |   |   |      |   |  |      |   |   |      |

国际检索报告  
关于同族专利的信息

国际申请号  
PCT/CN2020/080991

| 检索报告引用的专利文件 |            |    | 公布日<br>(年/月/日) | 同族专利 |             |    | 公布日<br>(年/月/日) |
|-------------|------------|----|----------------|------|-------------|----|----------------|
| CN          | 109543697  | A  | 2019年 3月 29日   | 无    |             |    |                |
| CN          | 107045630  | A  | 2017年 8月 15日   | 无    |             |    |                |
| CN          | 109598301  | A  | 2019年 4月 9日    | 无    |             |    |                |
| CN          | 108734210  | A  | 2018年 11月 2日   | 无    |             |    |                |
| WO          | 2019162241 | A1 | 2019年 8月 29日   | TW   | 201944283   | A  | 2019年 11月 16日  |
| US          | 2014107842 | A1 | 2014年 4月 17日   | KR   | 20140049152 | A  | 2014年 4月 25日   |
|             |            |    |                | US   | 9201425     | B2 | 2015年 12月 1日   |