

# (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  
2021年9月23日 (23.09.2021)

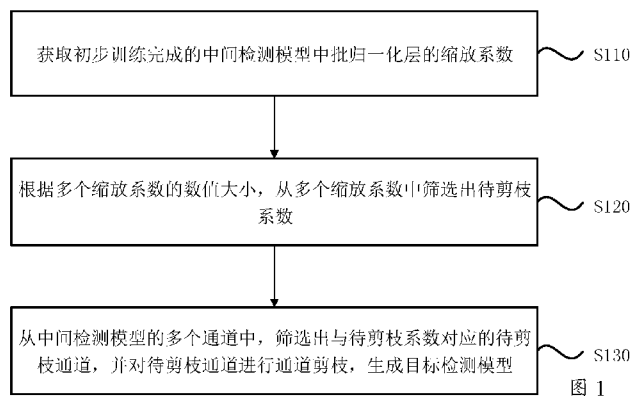


(10) 国际公布号  
**WO 2021/185121 A1**

- (51) 国际专利分类号:  
**G06K 9/62** (2006.01)
- (21) 国际申请号: PCT/CN2021/079690
- (22) 国际申请日: 2021年3月9日 (09.03.2021)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:  
202010188303.6 2020年3月17日 (17.03.2020) CN
- (71) 申请人: 北京京东乾石科技有限公司 (**BEIJING JINGDONG QIANSHI TECHNOLOGY CO., LTD.**) [CN/CN]; 中国北京市大兴区北京经济技术开发区科创十一街18号院2号楼19层A1905室, Beijing 100176 (CN)。
- (72) 发明人: 安耀祖 (**AN, Yaozu**); 中国北京市大兴区北京经济技术开发区科创十一街18号院2号楼19层A1905室, Beijing 100176 (CN)。 许新玉 (**XU, Xinyu**); 中国北京市大兴区北京经济技术开发区科创十一街18号院2号楼19层A1905室, Beijing 100176 (CN)。 孔旗 (**KONG, Qi**); 中国北京市大兴区北京经济技术开发区科创十一街18号院2号楼19层A1905室, Beijing 100176 (CN)。
- (74) 代理人: 北京品源专利代理有限公司 (**BEYOND ATTORNEYS AT LAW**); 中国北京市海淀区莲花池东路39号西金大厦6层, Beijing 100036 (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:** MODEL GENERATION METHOD AND APPARATUS, OBJECT DETECTION METHOD AND APPARATUS, DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 模型生成方法、目标检测方法、装置、设备及存储介质



- S110 Obtain scaling factors of a batch normalization layer in an intermediate detection model obtained by preliminary training
- S120 Filter, from among the multiple scaling factors according to numerical values of the multiple scaling factors, factors to be pruned
- S130 Filter, from among multiple channels of the intermediate detection model, channels to be pruned corresponding to the factors to be pruned, and prune the channel to be pruned to generate an object detection model

(57) **Abstract:** A model generation method and apparatus, an object detection method and apparatus, a device, and a storage medium. The model generation method comprises: obtaining multiple scaling factors of a batch normalization layer in an intermediate detection model obtained by preliminary training, wherein the intermediate detection model is obtained after training an original detection model on the basis of multiple training samples, and each training sample comprises a sample image and a sample annotation result of a known object in the sample image; filtering, from among the multiple scaling factors according to numerical values of the multiple scaling

WO 2021/185121 A1

GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, 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))。

---

factors, factors to be pruned; and filtering, from among multiple channels of the intermediate detection model, channels to be pruned corresponding to the factors to be pruned, and pruning the channel to be pruned to generate an object detection model.

**(57) 摘要:** 一种模型生成方法、目标检测方法、装置、设备及存储介质。该方法包括: 获取初步训练完成的中间检测模型中批归一化层的多个缩放系数, 其中, 中间检测模型是基于多个训练样本对原始检测模型进行训练后得到的, 每个训练样本包括样本图像和样本图像中已知目标的样本标注结果; 根据多个缩放系数的数值大小, 从多个缩放系数中筛选出待剪枝系数; 从中间检测模型的多个通道中, 筛选出与待剪枝系数对应的待剪枝通道, 并对待剪枝通道进行通道剪枝, 生成目标检测模型。

## 模型生成方法、目标检测方法、装置、设备及存储介质

本申请要求在 2020 年 3 月 17 日提交中国专利局、申请号为 202010188303.6 的中国专利申请的优先权，该申请的全部内容通过引用结合在本申请中。

### 技术领域

本申请实施例涉及计算机应用技术领域，例如涉及一种模型生成方法、目标检测方法、装置、设备及存储介质。

### 背景技术

目标检测(Object Detection)技术是很多计算机视觉任务的基础，其可用于判断待检测图像中是否存在感兴趣的待检测目标，并对待检测目标进行精准定位。而且，目标检测技术可与目标跟踪、目标重识别等技术相结合，应用于人工智能系统、车辆自动驾驶系统、智能机器人、智能物流等领域。

在实现本申请的过程中，发明人发现相关技术中存在以下情况：相关技术中的目标检测技术多是基于深度学习模型实现的，而深度学习模型因尺寸较大容易出现检测速度较慢的问题，这在计算资源受限的设备中尤其明显，这使得目标检测技术很难在实际项目上直接落地应用。

例如，在智能物流领域中，无人配送车的大量应用可以降低配送成本且提高配送效率，而基于视觉的目标检测技术是无人配送车感知周围环境所需的一种非常重要的技术手段。但是，出于量产和成本方面的考虑，无人配送车上的车载处理器多是基于计算资源相对有限的Xvaier平台组成的。由此，应用于这一车载处理器上的深度学习模型的检测速度相对较慢，这将直接影响无人配送车的环境感知能力，进而影响无人配送车的配送效率。因此，如何提升深度学习模型的检测速度，这对智能物流领域的发展至关重要。

### 发明内容

本申请实施例提供了一种模型生成方法、目标检测方法、装置、设备及存储介质，以实现通过压缩模型来提升模型的检测速度的效果。

第一方面，本申请实施例提供了一种模型生成方法，可以包括：

获取初步训练完成的中间检测模型中批归一化层的多个缩放系数，其中，中间检测模型是基于多个训练样本对原始检测模型进行训练后得到的，每个训练样本包括样本图像和样本图像中已知目标的样本标注结果；

根据多个缩放系数的数值大小，从多个缩放系数中筛选出待剪枝系数；

从中间检测模型的多个通道中，筛选出与待剪枝系数对应的待剪枝通道，并对待剪枝通道进行通道剪枝，生成目标检测模型。

第二方面，本申请实施例还提供了一种目标检测方法，可以包括：

获取待检测图像和按照上述任一项的方法生成的目标检测模型；

将待检测图像输入至目标检测模型中，根据目标检测模型的输出结果，得到待检测图像中待检测目标的目标检测结果。

第三方面，本申请实施例还提供了一种模型生成装置，可以包括：

第一获取模块，设置为获取初步训练完成的中间检测模型中批归一化层的多个缩放系数，其中，中间检测模型是基于多个训练样本对原始检测模型进行训练后得到的，每个训练样本包括样本图像和样本图像中已知目标的样本标注结果；

第一筛选模块，设置为根据多个缩放系数的数值大小，从多个缩放系数中筛选出待剪枝系数；

模型生成模块，设置为从中间检测模型的多个通道中，筛选出与待剪枝系数对应的待剪枝通道，并对待剪枝通道进行通道剪枝，生成目标检测模型。

第四方面，本申请实施例还提供了一种目标检测装置，可以包括：

第二获取模块，设置为获取待检测图像和按照上述任一项的方法生成的目标检测模型；

目标检测模块，设置为将待检测图像输入至目标检测模型中，根据目标检测模型的输出结果，得到待检测图像中待检测目标的目标检测结果。

第五方面，本申请实施例还提供了一种设备，可以包括：

至少一个处理器；

存储器，设置为存储至少一个程序；

当至少一个程序被至少一个处理器执行，使得至少一个处理器实现本申请任意实施例所提供的模型生成方法或是目标检测方法。

第六方面，本申请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实现本申请任意实施例所提供的模型生成方法或是目标检测方法。

## 附图说明

图1是本申请实施例一中的一种模型生成方法的流程图；

图2是本申请实施例二中的一种模型生成方法的流程图；

图3是本申请实施例三中的一种目标检测方法的流程图；

图4a是本申请实施例三中的一种目标检测方法中模型压缩流程图；

图4b是本申请实施例三中的一种目标检测方法中模型剪枝流程图；

图5是本申请实施例四中的一种模型生成装置的结构框图；

图6是本申请实施例五中的一种目标检测装置的结构框图；

图7是本申请实施例六中的一种设备的结构示意图；

图8是本申请实施例九中的目标检测系统的结构示意图；

图9是本申请实施例十中的无人车的结构示意图。

## 具体实施方式

## 实施例一

图 1 是本申请实施例一中提供的一种模型生成方法的流程图。本实施例可适用于对目标检测技术中的深度学习模型进行压缩的情况。该方法可以由本申请实施例提供的模型生成装置来执行，该装置可以由软件和硬件中的至少一种方式实现，该装置可以集成在各种电子设备上。

参见图 1，本申请实施例的方法包括步骤 S110 至 S130。

S110、获取初步训练完成的中间检测模型中批归一化层的多个缩放系数，其中，中间检测模型是基于多个训练样本对原始检测模型进行训练后得到的，每个训练样本包括样本图像和样本图像中已知目标的样本标注结果。

其中，获取未经训练的原始检测模型，该原始检测模型是深度学习模型中一种用于视觉检测的模型，其可以分为 anchor-based、anchor-free 和两者融合类，它们的区别就在于是否利用 anchor 提取候选框。anchor 也可称为 anchor box，其是在模型训练前，在训练样本上基于聚类算法得到的一组矩形框。

anchor-based 中的原始检测模型包括 fasterRCNN、SSD(Single Shot MultiBox Detector, 单次多框检测器)、YoloV2、YoloV3 等等；anchor-free 中的原始检测模型包括 CornerNet、ExtremeNet、CenterNet、FCOS 等等；融合 anchor-based 分支和 anchor-free 分支的原始检测模型包括 FSAF、SFace、GA-RPN 等等。SSD 是一种单阶段 (one-stage) 检测模型，其没有候选区域生成 (region proposals) 阶段，直接生成待检测目标的类别概率和位置坐标，在检测速度上具有较大优势，可以更好的在无人配送车、移动终端上运行。由此，一个可选示例，原始检测模型可以是 SSD，在此基础上，SSD 的主干网络可以是 inception\_v3 结构。

由此，在基于多个训练样本对原始检测模型进行训练后，可以得到初步训练完成的中间检测模型，每个训练样本可以包括样本图像和样本图像中已知目

标的样本标注结果，该样本图像可以是一帧图像、视频序列等等，该样本标注结果可以是类别概率和位置坐标。

需要说明的是，在原始检测模型的每个卷积层的后面均紧邻一个批归一化（Batch Normalization, BN）层，BN层可以对每个卷积层的输出结果的尺度进行归一化，这可以在训练过程中避免出现梯度损失和梯度溢出的情况。BN层包括缩放系数（gamma系数）和偏移系数（beta系数），其中，在每个BN层中，缩放系数的个数与该BN层紧邻的卷积层中的通道个数是一致的，即每个缩放系数对应卷积层中的一个通道。例如，若某个BN层中有32个缩放系数，则与该BN层紧邻的卷积层中包括32个通道，且该BN层亦包括32个通道。而且，在原始检测模型的训练阶段和应用阶段，每个缩放系数是与卷积层中的对应通道进行乘法运算，即某个缩放系数是否存在，将直接影响与其对应的卷积层中的通道是否起到作用。因此，可以获取中间检测模型中批归一化层的多个缩放系数，并根据该多个缩放系数确定对中间检测模型中的哪些进行通道剪枝。

S120、根据多个缩放系数的数值大小，从多个缩放系数中筛选出待剪枝系数。

其中，根据多个缩放系数的数值大小，从多个缩放系数中筛选出待剪枝系数的实现方式有多种，比如，可以对多个缩放系数的数值大小进行排序，根据排序结果得到多个缩放系数的中值，并根据该中值从多个缩放系数中筛选出待剪枝系数；比如，可以计算出多个缩放系数的数值大小的均值，并根据该均值从多个缩放系数中筛选出待剪枝系数；再比如，可以根据多个缩放系数的数值大小和预设剪枝率，得出多个缩放系数的剪枝阈值，并根据剪枝阈值从多个缩放系数中筛选出待剪枝系数，该待剪枝系数可以是数值小于或等于剪枝阈值的缩放系数；等等。

S130、从中间检测模型的多个通道中，筛选出与待剪枝系数对应的待剪枝通道，并对待剪枝通道进行通道剪枝，生成目标检测模型。

其中，多个缩放系数和某个卷积层中的多个通道是一一对应的，且某个卷积层中的多个通道和与该卷积层紧邻的BN层中的多个通道也是一一对应的，由此，根据待剪枝系数可以从中间检测模型的多个通道中筛选出待剪枝通道，该待剪枝通道是那些重要性较低的通道，它们可能是某个卷积层中的通道，也可能是某个BN层中的通道。

可以对待剪枝通道进行通道剪枝，生成目标检测模型，由此实现了模型压缩的效果。其中，通道剪枝(channel pruning)是通过删减模型中冗余通道的方式来简化模型，是一种结构化压缩方式；而且，在对待剪枝通道进行通道剪枝后，与这些待剪枝通道对应的卷积核也会相应删除，因此通过通道剪枝也减小卷积的运算量。示例性的，如果某个卷积层是32个通道，则与其紧邻的BN层也是32个通道，BN层中的每个通道包括缩放系数和偏移系数，待剪枝系数是从缩放系数中筛选出来的，由此，根据待剪枝系数可以确定BN层中的哪些通道是待剪枝通道，相应的可以确定卷积层中的哪些通道是待剪枝通道。

可选的，上述通道剪枝的实现过程可以是：从中间检测模型的多个卷积层的多个通道中，筛选出与多个待剪枝系数中的当前剪枝系数对应的当前卷积层的输出通道，以及当前卷积层的下一层卷积层的输入通道，并将当前卷积层的输出通道和当前卷积层的下一层卷积层的输入通道作为待剪枝通道。这是因为，当前卷积层的输出通道是当前卷积层的下一层卷积层的输入通道，示例性的，若当前卷积层的输出通道是1-32，则当前卷积层的下一层卷积层的输入通道也是1-32，此时，若与当前剪枝系数对应的当前卷积层的输出通道17是待剪枝通道，则当前卷积层的下一层卷积层的输入通道17也是待剪枝通道。

本申请实施例的技术方案，通过获取初步训练完成的中间检测模型中批归一化层的多个缩放系数，可以根据多个缩放系数的数值大小，从多个缩放系数中筛选出待剪枝系数；因待剪枝系数和待剪枝通道具有对应关系，则可以从中间检测模型的多个通道中筛选出与待剪枝系数对应的待剪枝通道，并对待剪枝通道进行通道剪枝，生成目标检测模型。上述技术方案，将通道剪枝与中间检测模型相结合，可以根据初步训练完成的中间检测模型中的缩放系数，对中间检测模型进行通道剪枝，由此实现了通过压缩模型来提升模型的检测速度的效果。

一种可选的技术方案，上述模型生成方法还可以包括：从中间检测模型的多个卷积层中筛选出可剪枝卷积层，其中，可剪枝卷积层包括除  $1*1$  卷积层和/或分类回归分支中的卷积层之外的卷积层；从多个缩放系数中筛选出与可剪枝卷积层对应的缩放系数，所述可剪枝卷积层对应的缩放系数为目标缩放系数；相应的，根据多个缩放系数的数值大小，从多个缩放系数中筛选出待剪枝系数，可以包括：根据多个目标缩放系数的数值大小，从多个目标缩放系数中筛选出待剪枝系数。

其中，通常情况下，原始检测模型通常包括主干网络和分类回归分支这两部分，主干网络可以用于提取特征图，分类回归分支是从主干网络中分支出的一个分类分支和一个回归分支，可用于对已提取的特征图进行分类或是回归。由于分类回归的类别通常是固定的，因此可尽可能保持分类回归分支中卷积层的固定，这样可以确保输出维度的固定，且可以简化执行代码。由此，可以将  $1*1$  卷积层和分类回归分支中的卷积层中的至少之一之外的卷积层作为可剪枝卷积层，并从可剪枝卷积层中的多个目标缩放系数筛选出待剪枝系数。

一种可选的技术方案，在对待剪枝通道进行通道剪枝后，可以先生成剪枝

检测模型；然后，再对剪枝检测模型进行微调训练，生成目标检测模型。也就是说，可以对通道剪枝后的精简的剪枝检测模型进行微调训练，由此恢复检测效果，即在压缩模型的同时，尽可能保持了模型的原始性能。其中，微调训练的过程可以是：获取历史图像和历史图像中已知目标的历史标注结果，并将历史图像和历史标注结果作为一组历史样本；基于多个历史样本对剪枝检测模型进行训练，得到目标检测模型。需要说明的是，通常情况下，该历史样本和上述训练样本是相同的样本数据，即在微调训练时，历史图像和样本图像可以是相同的图像，且历史标注结果和样本标注结果也可以是相同的标注结果。

## 实施例二

图 2 是本申请实施例二中提供的一种模型生成方法的流程图。本实施例以上述各技术方案为基础进行细化。在本实施例中，可选的，上述模型生成方法，还可以包括：获取多个训练样本，并基于多个训练样本对原始检测模型进行基于批归一化层的稀疏化训练，得到中间检测模型。其中，与上述各实施例相同或相应的术语的解释在此不再赘述。

参见图 2，本实施例的方法可以包括步骤 S210 至 S230。

S210、获取多个训练样本，并基于多个训练样本对原始检测模型进行基于批归一化层的稀疏化训练，得到中间检测模型，其中，每个训练样本包括样本图像和样本图像中已知目标的样本标注结果。

其中，在对原始检测模型进行基于 BN 层的稀疏化训练时，可得到一个 BN 层稀疏化的中间检测模型，即在原始检测模型的稠密连接上引入稀疏化。示例性的，基于 BN 层的稀疏化训练的一个可选方案是，对原始检测模型中的每个缩放系数施加 L1 正则约束，使得原始检测模型朝着结构性稀疏的方向调整参数，此时 BN 层中的缩放系数（gamma 系数）所起的作用相当于信息流通道的开关

系数，控制着信息流通道的开关闭合。

这样设置的原因在于，在模型训练过程中，对缩放系数施加 L1 正则约束，可以将更多的缩放系数调整为 0。由此，在模型训练阶段和应用阶段，因缩放系数是与卷积层中的对应通道进行乘法运算，则当更多的缩放系数为 0 时，与其对应的卷积层中的通道将不再起到任何作用，即通过对缩放系数进行大幅压缩的方式也起到了通道剪枝的作用。在此基础上，在根据预设剪枝率筛选待剪枝通道时，若中间检测模型中数值为 0 的缩放系数越多，则将数值非 0 的缩放系数对应的通道剪掉的可能性越低，由此生成的目标检测模型的网络结构和中间检测模型的网络结构越一致，这样一来，两者的检测性能也越一致，即在保证检测性能的同时，实现了模型压缩的效果。

在此基础上，原始检测模型中的目标损失函数可以由原始损失函数和 L1 正则约束函数构成，L1 正则约束函数可以包括对多个缩放系数进行 L1 正则约束的损失函数。即，在原始损失函数的基础上，引入了 BN 层的缩放系数的 L1 正则约束项。这样一来，在训练过程中可以根据目标损失函数求解极小值，并根据求解结果调整模型中的多个参数值。

可选的，目标损失函数 L 可以通过如下公式表示：

$$L = \sum_{(x,y)} l(f(x,W), y) + \lambda \sum_{\gamma \in \Gamma} g(\gamma)$$

其中， $x$  是样本图像， $y$  是样本图像的样本标注结果， $W$  是原始检测模型中的参数值， $f(x,W)$  是样本图像中已知目标的样本预测结果， $\gamma$  是缩放系数， $\lambda$  是惩罚系数， $l()$  是原始损失函数， $g()$  是 L1 正则约束函数， $\Gamma$  表示所述原始检测模型中所有缩放系数的集合。而且，因为 L1 正则约束函数仅施加于 BN 层的缩放系数，因此在反向传播更新梯度时，缩放系数的梯度  $\gamma^{grad}$  需要加上一个 [缩放系数符

号 $\text{sign}(\gamma)$ 与惩罚系数 $\lambda$ 的乘积项], 其公式如下所示:

$$\gamma^{grad} = \gamma^{grad} + \lambda * \text{sign}(\gamma)$$

S220、获取中间检测模型中批归一化层的多个缩放系数, 并根据多个缩放系数的数值大小, 从多个缩放系数中筛选出待剪枝系数。

S230、从中间检测模型的多个通道中, 筛选出与待剪枝系数对应的待剪枝通道, 并对待剪枝通道进行通道剪枝, 生成目标检测模型。

本申请实施例的技术方案, 基于多个训练样本对原始检测模型进行基于BN层的稀疏化训练, 可以得到一个BN层稀疏化的中间检测模型; 在对中间检测模型进行通道剪枝后可生成的目标检测模型, 且二者的检测性能较为一致, 即在保证检测性能的同时, 实现了模型压缩的效果。

### 实施例三

图3是本申请实施例三中提供的一种目标检测方法的流程图。本实施例可适用于基于上述任意实施例所述的方法生成的目标检测模型, 对待检测图像进行目标检测的情况。该方法可以由本申请实施例提供的目标检测装置来执行, 该装置可以由软件和硬件中的至少之一的方式实现, 该装置可以集成在各种电子设备上。

参见图3, 本申请实施例的方法包括步骤S310至S320。

S310、获取待检测图像和按照上述任意实施例的方法生成的目标检测模型。

其中, 该待检测图像可以是一帧图像、视频序列等等, 该目标检测模型可以是按照上述任意实施例所述的方法生成的是视觉检测模型。示例性的, 目标检测模型的一种生成方法可以如图4a所示:

首先, 以原始SSD检测模型为基础, 以inception\_v3结构为主干网络, 在

模型训练期间,在保持原有参数设置的情况下,通过对与卷积层紧邻的BN层中的gamma系数施加L1正则约束,使得模型朝着结构性稀疏的方向调整参数,由此实现了BN层的稀疏化。其次,在完成基于BN层的稀疏化训练后,可以根据BN层的缩放系数,对初步训练完成的中间检测模型按照预设剪枝率裁剪对应的卷积层和BN层中的通道,由此能够精简模型并提升一定的检测速度。最后,对通道剪枝后的精简模型进行微调训练,恢复检测效果。

上述通道剪枝过程可以如图4b所示,首先,需要确定对哪些卷积层进行通道剪枝,在本申请实施例中有两个限制:1\*1卷积层和分类回归分支中的卷积层均不剪枝,这可以确保输出的维度不变。其次,统计其余的可剪枝卷积层对应的BN层中的gamma系数,并对统计出的全部gamma系数进行排序,按照预设剪枝率计算出gamma系数的剪枝阈值。再次,根据剪枝阈值对卷积层和BN层中的各个通道进行选取,保留那些数值大于剪枝阈值的gamma系数对应的通道,由此确定出BN层以及与BN层紧邻的卷积层中可保留的通道的掩模(MASK)。最后,根据MASK保留卷积层和BN层中对应的通道,并剪枝掉那些未保留的通道。

S320、将待检测图像输入至目标检测模型中,根据目标检测模型的输出结果,得到待检测图像中待检测目标的目标检测结果。

一个可选示例,以背景技术中的例子为例,上述目标检测方法可以应用于智能物流领域的无人配送车上的视觉目标检测。虽然无人配送车上的车载处理器多是基于计算资源相对有限的Xavier平台组成的,但是,上述目标检测方法中涉及到的目标检测模型的规模较小且检测速度较快,即使受限计算资源的约束下,依然可以满足无人配送车的真正意义上的脱人运行。而且,在通道级上实施结构化剪枝操作,由此生成的精简模型能够直接运行在成熟框架比如

Pytorch、MXnet、TensorFlow 等等，或是硬件平台上比如图形处理器（Graphics Processing Unit, GPU）、现场可编程逻辑门阵列（Field Programmable Gate Array, FPGA）等等，无需特殊算法库的支持，应用起来更加便捷。

为验证上述目标检测方法的检测精度，将这一方法在加州大学伯克利 DeepDrive 深度学习自动驾驶产业联盟（Berkeley DeepDrive, BDD）数据集的 5 类子集（car, pedestrian, truck, bus, rider）上测试检测精度，其量化结果如下述两个表格所示。从表中数据可以得知，本申请实施例的结构化剪枝的目标检测方法，在保留部分卷积层和 BN 层不动的情况下，能够达到比较明显的压缩效果，同时其检测结果 mAP(5 类子集的平均值)只有很微小的下降。

预设剪枝率	模型参数量	模型计算量 (FLOPS)
base	28.29M	9.45G
30%	21.55M	5.97G
50%	16.76M	5.60G
80%	11.24M	4.94G

预设剪枝率	mAP	car	pedestrian	truck	bus	rider
base	37.56	56.66	22.95	44.56	44.37	19.28
80%	37.32	56.55	23.26	44.11	44.13	18.55

本申请实施例的技术方案，可以基于已生成的目标检测模型对待检测图像进行目标检测，因目标检测模型是模型压缩后的精简模型，这可有效提升待检测图像中待检测目标的检测速度，且可尽可能保持模型的原始性能。

#### 实施例四

图 5 为本申请实施例四提供的模型生成装置的结构框图，该装置设置为执行上述任意实施例所提供的模型生成方法。该装置与上述各实施例的模型生成方法属于同一个发明构思，在模型生成装置的实施例中未详尽描述的细节内容，可以参考上述模型生成方法的实施例。参见图 5，该装置可包括：第一获取模块 410、第一筛选模块 420 和模型生成模块 430。

其中，第一获取模块 410，设置为获取初步训练完成的中间检测模型中批归一化层的多个缩放系数，中间检测模型是基于多个训练样本对原始检测模型进行训练后得到的，每个训练样本包括样本图像和样本图像中已知目标的样本标注结果；

第一筛选模块 420，设置为根据多个缩放系数的数值大小，从多个缩放系数中筛选出待剪枝系数；

模型生成模块 430，设置为从中间检测模型的多个通道中，筛选出与待剪枝系数对应的待剪枝通道，并对待剪枝通道进行通道剪枝，生成目标检测模型。

可选的，第一筛选模块 420，可以设置为：

根据多个缩放系数的数值大小和预设剪枝率，得出多个缩放系数的剪枝阈值，并根据剪枝阈值从多个缩放系数中筛选出待剪枝系数。

可选的，在上述装置的基础上，该装置还可以包括：

第二筛选模块，设置为从中间检测模型的多个卷积层中筛选出可剪枝卷积层，可剪枝卷积层包括除  $1*1$  卷积层和分类回归分支中的卷积层中的至少之一

之外的卷积层；

第三筛选模块，设置为从多个缩放系数中筛选出与可剪枝卷积层对应的缩放系数，所述可剪枝卷积层对应的缩放系数为目标缩放系数；

第一筛选模块 420，可以设置为：

根据多个目标缩放系数的数值大小，从多个目标缩放系数中筛选出待剪枝系数。

可选的，模型生成模块 430，可以包括：

待剪枝通道筛选单元，设置为从中间检测模型的多个卷积层的多个通道中，筛选出与多个待剪枝系数中的当前剪枝系数对应的当前卷积层的输出通道，以及当前卷积层的下一层卷积层的输入通道，并将输出通道和输入通道作为待剪枝通道。

可选的，在上述装置的基础上，该装置还可以包括：

第三获取模块，设置为获取多个训练样本，并基于多个训练样本对原始检测模型进行基于批归一化层的稀疏化训练，得到中间检测模型。

可选的，原始检测模型中的目标损失函数由原始损失函数和 L1 正则约束函数构成，L1 正则约束函数包括对多个缩放系数进行 L1 正则约束的损失函数。

可选的，目标损失函数L通过如下公式表示：

$$L = \sum_{(x,y)} l(f(x,W), y) + \lambda \sum_{\gamma \in \Gamma} g(\gamma)$$

其中， $x$ 是样本图像， $y$ 是样本图像的样本标注结果， $W$ 是原始检测模型中的参数值， $f(x,W)$ 是样本图像中已知目标的样本预测结果， $\gamma$ 是缩放系数， $\lambda$ 是惩罚系数， $l()$ 是原始损失函数， $g()$ 是 L1 正则约束函数， $\Gamma$ 表示所述原始检测模型中所有缩放系数的集合。

可选的，模型生成模块 430，可以包括：

通道剪枝单元，设置为对待剪枝通道进行通道剪枝，得到剪枝检测模型；

微调训练单元，设置为对剪枝检测模型进行微调训练，生成目标检测模型。

本申请实施例四提供的模型生成装置，通过第一获取模块获取初步训练完成的中间检测模型中批归一化层的多个缩放系数；第一筛选模块可以根据多个缩放系数的数值大小，从多个缩放系数中筛选出待剪枝系数；模型生成模块因待剪枝系数和待剪枝通道具有对应关系，则可以从中间检测模型的多个通道中筛选出与待剪枝系数对应的待剪枝通道，并对待剪枝通道进行通道剪枝，生成目标检测模型。上述装置，将通道剪枝与中间检测模型相结合，可以根据初步训练完成的中间检测模型中的缩放系数，对中间检测模型进行通道剪枝，由此实现了通过压缩模型来提升模型的检测速度的效果。

本申请实施例所提供的模型生成装置可执行本申请任意实施例所提供的模型生成方法，具备执行方法相应的功能模块。

值得注意的是，上述模型生成装置的实施例中，所包括的各个单元和模块只是按照功能逻辑进行划分的，但并不局限于上述的划分，只要能够实现相应的功能即可；另外，各功能单元的名称也只是为了便于相互区分，并不用于限制本申请的保护范围。

#### 实施例五

图 6 为本申请实施例五提供的目标检测装置的结构框图，该装置设置为执行上述任意实施例所提供的目标检测方法。该装置与上述各实施例的目标检测方法属于同一个发明构思，在目标检测装置的实施例中未详尽描述的细节内容，可以参考上述目标检测方法的实施例。参见图 6，该装置可包括：第二获取模块 510 和目标检测模块 520。

其中，第二获取模块 510，设置为获取待检测图像和按照实施例一和实施例二中的任一项的方法生成的目标检测模型；

目标检测模块 520，设置为将待检测图像输入至目标检测模型中，根据目标检测模型的输出结果，得到待检测图像中待检测目标的目标检测结果。

本申请实施例五提供的目标检测装置，通过第二获取模块和目标检测模块相互配合，可以基于已生成的目标检测模型对待检测图像进行目标检测，因目标检测模型是模型压缩后的精简模型，这可有效提升待检测图像中待检测目标的检测速度，且可尽可能保持模型的原始性能。

本申请实施例所提供的目标检测装置可执行本申请任意实施例所提供的目标检测方法，具备执行方法相应的功能模块。

值得注意的是，上述目标检测装置的实施例中，所包括的各个单元和模块只是按照功能逻辑进行划分的，但并不局限于上述的划分，只要能够实现相应的功能即可；另外，各功能单元的名称也只是为了便于相互区分，并不用于限制本申请的保护范围。

#### 实施例六

图 7 为本申请实施例六提供的一种设备的结构示意图，如图 7 所示，该设备包括存储器 610、处理器 620、输入装置 630 和输出装置 640。设备中的处理器 620 的数量可以是至少一个，图 7 中以一个处理器 620 为例；设备中的存储器 610、处理器 620、输入装置 630 和输出装置 640 可以通过总线或其它方式连接，图 7 中以通过总线 650 连接为例。

存储器 610 作为一种计算机可读存储介质，可设置为存储软件程序、计算机可执行程序以及模块，如本申请实施例中的模型生成方法对应的程序指令/模块（例如，模型生成装置中的第一获取模块 410、第一筛选模块 420 和模型生成

模块 430)，或者，如本申请实施例中的目标检测方法对应的程序指令/模块（例如，目标检测装置中的第二获取模块 510 和目标检测模块 520）。处理器 620 通过运行存储在存储器 610 中的软件程序、指令以及模块，从而执行设备的各种功能应用以及数据处理，即实现上述的模型生成方法或是目标检测方法。

存储器 610 可主要包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需的应用程序；存储数据区可存储根据设备的使用所创建的数据等。此外，存储器 610 可以包括高速随机存取存储器，还可以包括非易失性存储器，例如至少一个磁盘存储器件、闪存器件、或其他非易失性固态存储器件。在一些实例中，存储器 610 可包括相对于处理器 620 远程设置的存储器，这些远程存储器可以通过网络连接至设备。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

输入装置 630 可设置为接收输入的数字或字符信息，以及产生与装置的用户设置以及功能控制有关的键信号输入。输出装置 640 可包括显示屏等显示设备。

### 实施例七

本申请实施例七提供一种包含计算机可执行指令的存储介质，所述计算机可执行指令在由计算机处理器执行时用于执行一种模型生成方法，该方法包括：

获取初步训练完成的中间检测模型中批归一化层的多个缩放系数，其中，中间检测模型是基于多个训练样本对原始检测模型进行训练后得到的，每个训练样本包括样本图像和样本图像中已知目标的样本标注结果；

根据多个缩放系数的数值大小，从多个缩放系数中筛选出待剪枝系数；

从中间检测模型的多个通道中，筛选出与待剪枝系数对应的待剪枝通道，并对待剪枝通道进行通道剪枝，生成目标检测模型。

当然，本申请实施例所提供的一种包含计算机可执行指令的存储介质，其计算机可执行指令不限于如上所述的方法操作，还可以执行本申请任意实施例所提供的模型生成方法中的相关操作。

#### 实施例八

本申请实施例八提供一种包含计算机可执行指令的存储介质，所述计算机可执行指令在由计算机处理器执行时用于执行一种目标检测方法，该方法包括：

获取待检测图像和按照上述实施例一和实施例二中的任一项所述的方法生成的目标检测模型；

将待检测图像输入至目标检测模型中，根据目标检测模型的输出结果，得到待检测图像中待检测目标的目标检测结果。

#### 实施例九

本申请实施例还提供一种目标检测系统，参见图 8，该系统包括采集设备 710、计算设备 720 及存储设备 730，所述存储设备 730 存储有实施例一或实施例二中生成的目标检测模型，所述采集设备 710 设置为采集待检测图像，所述计算设备 720 设置为加载所述待检测图像和目标检测模型，并将所述待检测图像输入至目标检测模型中，根据目标检测模型的输出结果，得到所述待检测图像中待检测目标的目标检测结果。

#### 实施例十

本申请实施例还提供了一种无人车，参见图 9，该无人车包括驱动设备 810、路径规划设备 820 以及实施例七所述的设备 830，所述驱动设备 810 设置为驱动所述无人车按照所述路径规划设备 820 规划的路径运行，实施例六所述的设备 830 设置为实现对待检测图像中的待检测目标的检测，路径规划设备 820 设置为根据实施例六所述的设备对待检测图像中的待检测目标的检测结果对所述无人

车的路径进行规划。

通过以上关于实施方式的描述，所属领域的技术人员可以清楚地了解到，本申请可借助软件及必需的通用硬件来实现，当然也可以通过硬件实现。依据这样的理解，本申请的技术方案本质上或者说对相关技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品可以存储在计算机可读存储介质中，如计算机的软盘、只读存储器（Read-Only Memory, ROM）、随机存取存储器（Random Access Memory, RAM）、闪存（FLASH）、硬盘或光盘等，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）执行本申请各个实施例所述的方法。

1、一种模型生成方法，包括：

获取初步训练完成的中间检测模型中批归一化层的多个缩放系数，其中，所述中间检测模型是基于多个训练样本对原始检测模型进行训练后得到的，每个训练样本包括样本图像和所述样本图像中已知目标的样本标注结果；

根据多个所述缩放系数的数值大小，从多个所述缩放系数中筛选出待剪枝系数；

从所述中间检测模型的多个通道中，筛选出与所述待剪枝系数对应的待剪枝通道，并对所述待剪枝通道进行通道剪枝，生成目标检测模型。

2、根据权利要求1所述的方法，其中，所述根据多个所述缩放系数的数值大小，从多个所述缩放系数中筛选出待剪枝系数，包括：

根据多个所述缩放系数的数值大小和预设剪枝率，得出多个所述缩放系数的剪枝阈值，并根据所述剪枝阈值从多个所述缩放系数中筛选出待剪枝系数。

3、根据权利要求1所述的方法，还包括：

从所述中间检测模型的多个卷积层中筛选出可剪枝卷积层，其中，所述可剪枝卷积层包括除  $1 \times 1$  卷积层和分类回归分支中的卷积层中的至少之一之外的卷积层；

从多个所述缩放系数中筛选出与所述可剪枝卷积层对应的缩放系数，所述可剪枝卷积层对应的缩放系数为目标缩放系数；

所述根据多个所述缩放系数的数值大小，从多个所述缩放系数中筛选出待剪枝系数，包括：根据多个所述目标缩放系数的数值大小，从多个所述目标缩放系数中筛选出待剪枝系数。

4、根据权利要求1所述的方法，其中，所述待剪枝系数为多个，所述从所述中间检测模型的多个通道中，筛选出与所述待剪枝系数对应的待剪枝通道，包括：

从所述中间检测模型的多个卷积层的多个通道中，筛选出与多个所述待剪枝系数中的当前剪枝系数对应的当前卷积层的输出通道，以及所述当前卷积层

的下一层卷积层的输入通道，并将所述输出通道和所述输入通道作为待剪枝通道。

5、根据权利要求 1 所述的方法，还包括：

获取多个所述训练样本，并基于多个所述训练样本对所述原始检测模型进行基于所述批归一化层的稀疏化训练，得到所述中间检测模型。

6、根据权利要求 5 所述的方法，其中，所述原始检测模型中的目标损失函数由原始损失函数和 L1 正则约束函数构成，所述 L1 正则约束函数包括对多个所述缩放系数进行 L1 正则约束的损失函数。

7、根据权利要求 6 所述的方法，其中，所述目标损失函数L通过如下公式表示：

$$L = \sum_{(x,y)} l(f(x,W),y) + \lambda \sum_{\gamma \in \Gamma} g(\gamma)$$

其中， $x$ 是一样本图像， $y$ 是所述样本图像中的样本标注结果， $W$ 是所述原始检测模型中的参数值， $f(x,W)$ 是所述样本图像中已知目标的样本预测结果， $\gamma$ 是缩放系数， $\lambda$ 是惩罚系数， $l()$ 是所述原始损失函数， $g()$ 是所述 L1 正则约束函数， $\Gamma$ 表示所述原始检测模型中所有缩放系数的集合。

8、根据权利要求 1 所述的方法，其中，所述对所述待剪枝通道进行通道剪枝，生成目标检测模型，包括：

对所述待剪枝通道进行通道剪枝，得到剪枝检测模型；

对所述剪枝检测模型进行微调训练，生成目标检测模型。

9、根据权利要求 1 所述的方法，其中，所述原始检测模型包括单次多框检测器 SSD。

10、根据权利要求 9 所述的方法，其中，所述 SSD 的主干网络包括 inception\_v3 结构。

11、一种目标检测方法，包括：

获取待检测图像和按照权利要求 1-10 中任一项的方法生成的目标检测模

型；

将所述待检测图像输入至所述目标检测模型中，根据所述目标检测模型的输出结果，得到所述待检测图像中待检测目标的目标检测结果。

12、一种模型生成装置，包括：

第一获取模块，设置为获取初步训练完成的中间检测模型中批归一化层的多个缩放系数，其中，所述中间检测模型是基于多个训练样本对原始检测模型进行训练后得到的，每个训练样本包括样本图像和所述样本图像中已知目标的样本标注结果；

第一筛选模块，设置为根据多个所述缩放系数的数值大小，从多个所述缩放系数中筛选出待剪枝系数；

模型生成模块，设置为从所述中间检测模型的多个通道中，筛选出与所述待剪枝系数对应的待剪枝通道，并对待剪枝通道进行通道剪枝，生成目标检测模型。

13、一种目标检测装置，包括：

第二获取模块，设置为获取待检测图像和按照权利要求 1-10 中任一项的方法生成的目标检测模型；

目标检测模块，设置为将所述待检测图像输入至所述目标检测模型中，根据所述目标检测模型的输出结果，得到待检测图像中待检测目标的目标检测结果。

14、一种设备，包括：

至少一个处理器；

存储器，设置为存储至少一个程序；

当所述至少一个程序被所述至少一个处理器执行，使得所述至少一个处理器实现如权利要求 1-10 中任一所述的模型生成方法。

15、一种设备，包括：

至少一个处理器；

存储器，设置为存储至少一个程序；

当所述至少一个程序被所述至少一个处理器执行，使得所述至少一个处理器实现如权利要求 11 中所述的目标检测方法。

16、一种计算机可读存储介质，所述计算机可读存储介质上存储有计算机程序，所述计算机程序被处理器执行时，实现如权利要求 1-10 中任一所述的模型生成方法，或者如权利要求 11 中所述的目标检测方法。

17、一种无人车，包括驱动设备、路径规划设备以及权利要求 15 所述的设备，所述驱动设备设置为驱动所述无人车按照所述路径规划设备规划的路径运行，权利要求 15 所述的设备设置为实现待检测图像中的待检测目标的检测，所述路径规划设备设置为根据权利要求 15 所述的设备对待检测图像中的待检测目标的检测结果对所述无人车的路径进行规划。

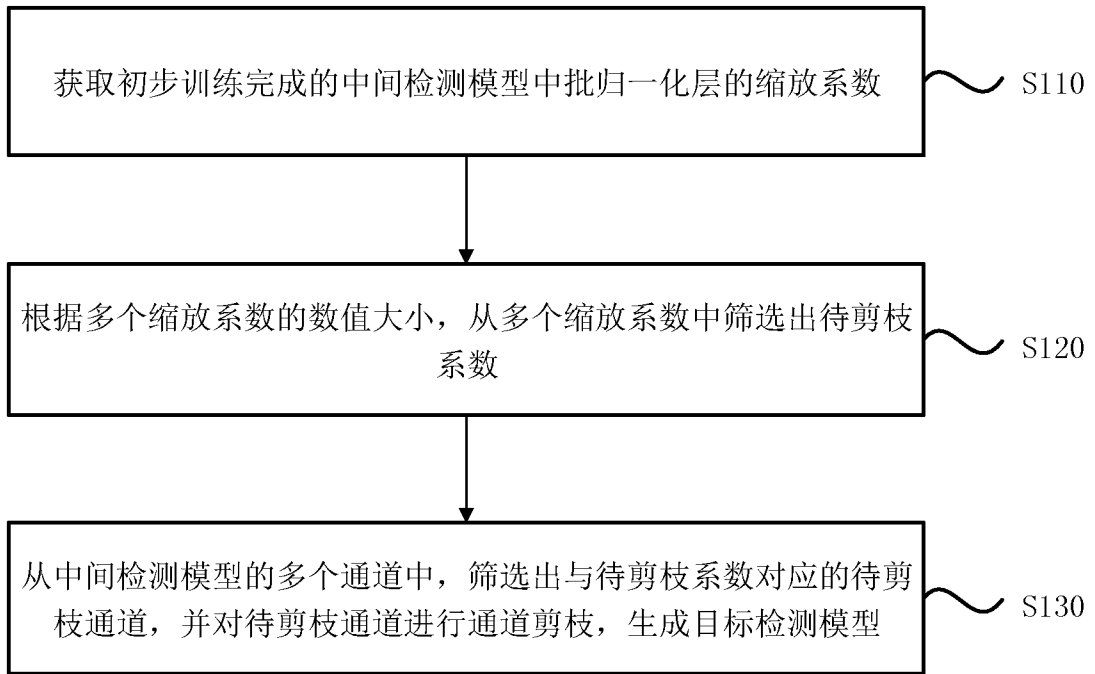


图 1

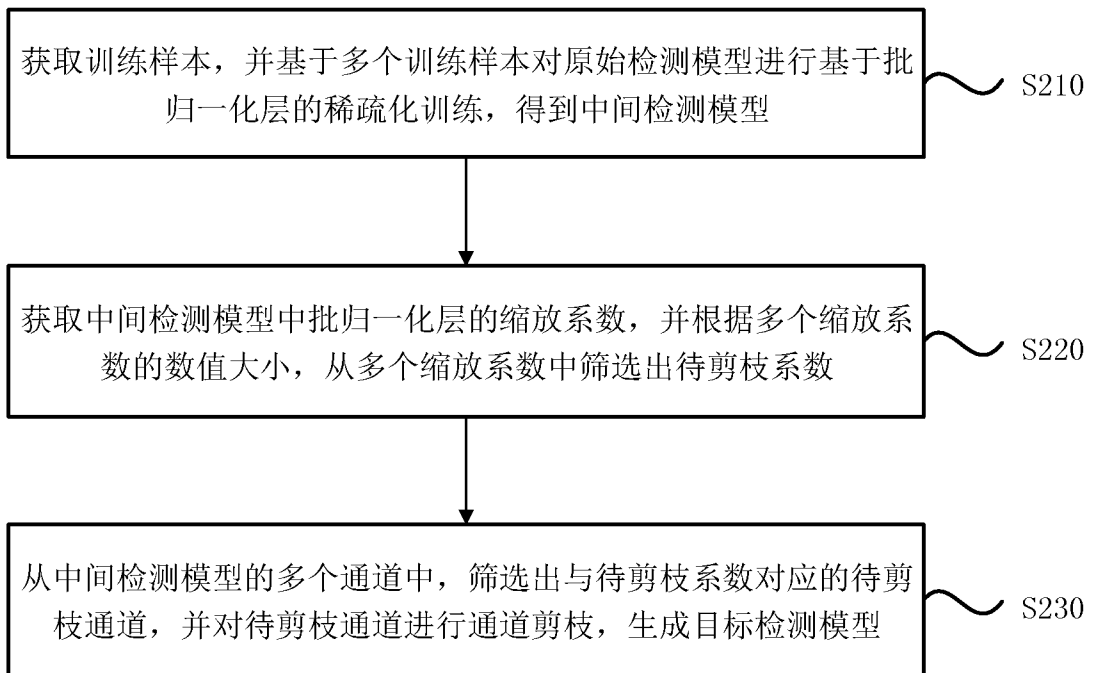


图 2

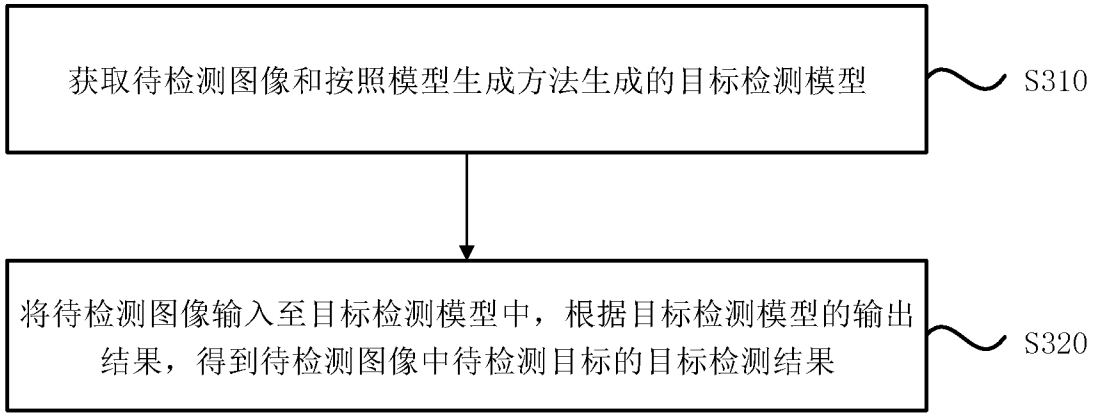


图 3

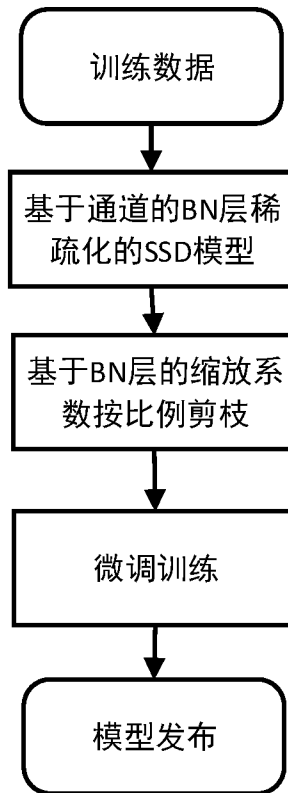


图 4a

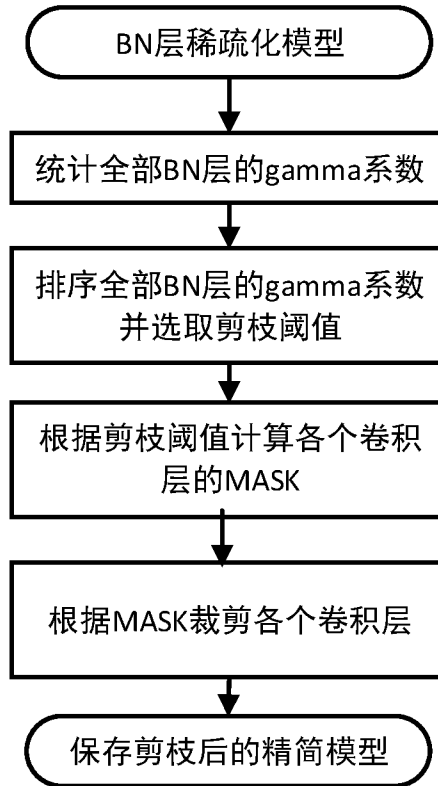


图 4b

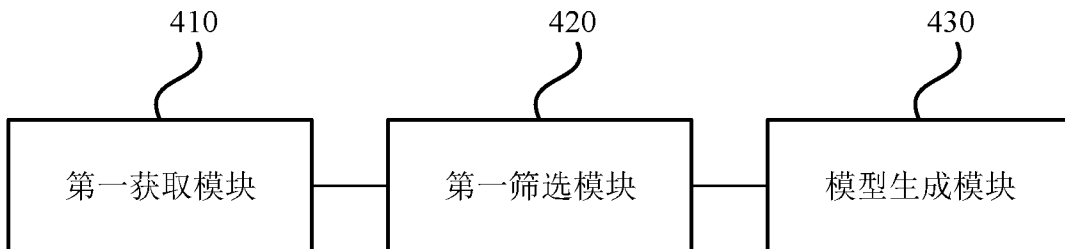


图 5

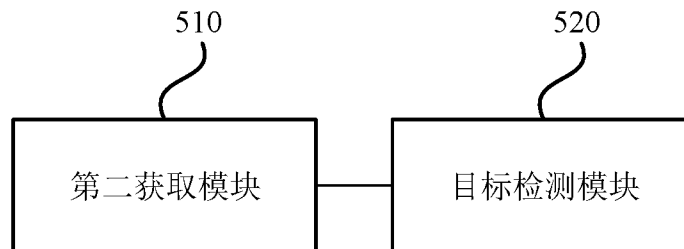


图 6

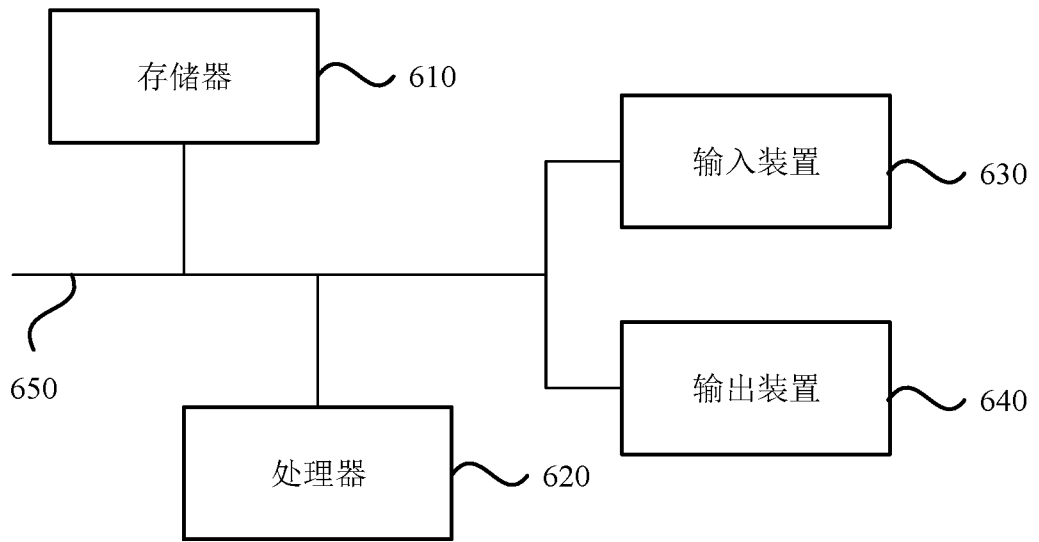


图 7

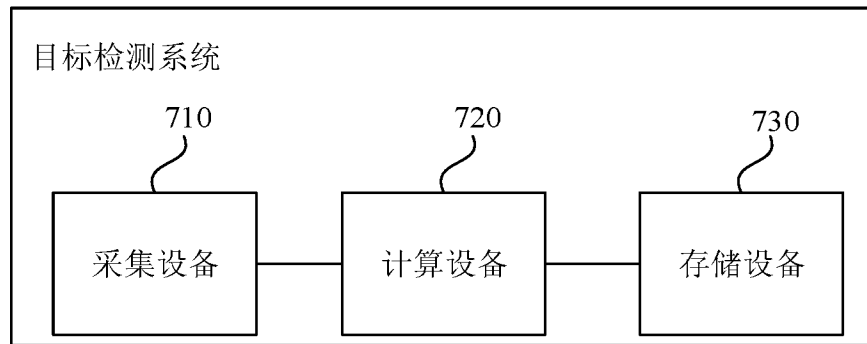


图 8

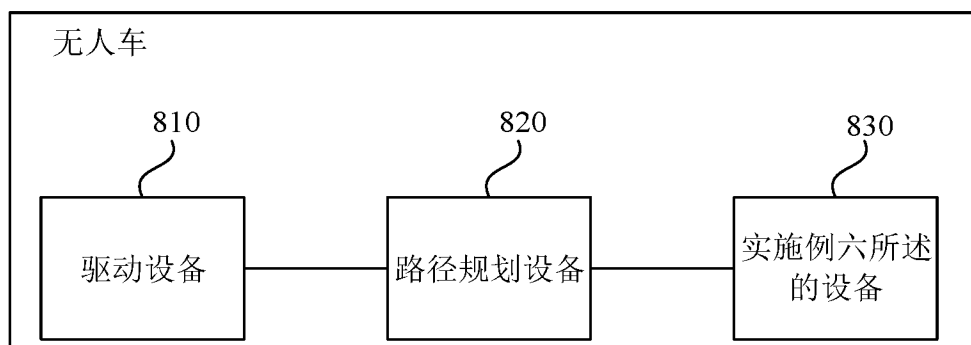


图 9

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/079690

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>		
G06K 9/62(2006.01)i		
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)		
G06K G06F		
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)		
CNKI, CNPAT, WPI, EPODOC: 规划, 无人车, 无人驾驶, 路径, 检测, 图像, 驱动, 通道, 剪枝, 模型, 网络, 缩放参数, 压缩, 缩放因子, 缩放系数, 裁剪, tailoring, prune, channel, model, path, image		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 111325342 A (ZICT TECHNOLOGY CO., LTD.) 23 June 2020 (2020-06-23) claims 1-10, description paragraphs 0005-0029, 0042-0097	1-10, 12, 14, 16
PY	CN 111325342 A (ZICT TECHNOLOGY CO., LTD.) 23 June 2020 (2020-06-23) claims 1-10, description paragraphs 0005-0029, 0042-0097	11, 13, 15, 16, 17
PX	CN 111062382 A (BEIJING JIAOTONG UNIVERSITY) 24 April 2020 (2020-04-24) claims 1-8, description paragraphs 0006-0085	1-10, 12, 14, 16
PY	CN 111062382 A (BEIJING JIAOTONG UNIVERSITY) 24 April 2020 (2020-04-24) claims 1-8, description paragraphs 0006-0085	11, 13, 15, 16, 17
X	宋非洋等 (SONG, Feiyang et al.). "基于MobileNetV3的结构性剪枝优化 (Optimization of Structural Pruning Based on MobileNetV3)" <i>自动化与信息工程 (Automation &amp; Information Engineering)</i> , Vol. 40, No. 6, 15 December 2019 (2019-12-15), pp. 20-25	1-10, 12, 14, 16
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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
17 May 2021		08 June 2021
Name and mailing address of the ISA/CN		Authorized officer
<b>China National Intellectual Property Administration (ISA/CN)</b> <b>No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088</b> <b>China</b>		
Facsimile No. (86-10)62019451		Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	宋非洋 等 (SONG, Feiyang et al.). "基于MobileNetV3的结构性剪枝优化 (Optimization of Structural Pruning Based on MobileNetV3)" <i>自动化与信息工程 (Automation &amp; Information Engineering)</i> , Vol. 40, No. 6, 15 December 2019 (2019-12-15), pp. 20-25	11, 13, 15, 16, 17
Y	CN 109344921 A (HUNAN G MAX INTELLIGENT TECHNOLOGY CO., LTD.) 15 February 2019 (2019-02-15) description paragraphs 0019-0044	11, 13, 15, 16, 17
Y	CN 107728620 A (JIANGSU KAWEI AUTOMOTIVE INDUSTRY GROUP CO., LTD.) 23 February 2018 (2018-02-23) description, paragraphs [0031]-[0054]	17
A	CN 110263841 A (NANJING UNIVERSITY OF INFORMATION SCIENCE & TECHNOLOGY) 20 September 2019 (2019-09-20) entire document	1-17
A	CN 107368885 A (BEIJING ICETECH SCIENCE & TECHNOLOGY CO., LTD.) 21 November 2017 (2017-11-21) entire document	1-17

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

International application No.

**PCT/CN2021/079690**

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 111325342 A	23 June 2020	None	
CN 111062382 A	24 April 2020	None	
CN 109344921 A	15 February 2019	None	
CN 107728620 A	23 February 2018	None	
CN 110263841 A	20 September 2019	None	
CN 107368885 A	21 November 2017	None	

国际检索报告

国际申请号

PCT/CN2021/079690

<p><b>A. 主题的分类</b></p> <p>G06K 9/62 (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>G06K G06F</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>CNKI, CNPAT, WPI, EPDOC: 规划, 无人车, 无人驾驶, 路径, 检测, 图像, 驱动, 通道, 剪枝, 模型, 网络, 缩放参数, 压缩, 缩放因子, 缩放系数, 裁剪, tailoring, prune, channel, model, path, image</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>PX</td> <td>CN 111325342 A (深圳中兴网信科技有限公司) 2020年 6月 23日 (2020 - 06 - 23) 权利要求第1-10, 说明书第0005-0029, 0042-0097段</td> <td>1-10, 12, 14, 16</td> </tr> <tr> <td>PY</td> <td>CN 111325342 A (深圳中兴网信科技有限公司) 2020年 6月 23日 (2020 - 06 - 23) 权利要求第1-10, 说明书第0005-0029, 0042-0097段</td> <td>11, 13, 15, 16, 17</td> </tr> <tr> <td>PX</td> <td>CN 111062382 A (北京交通大学) 2020年 4月 24日 (2020 - 04 - 24) 权利要求1-8, 说明书第0006-0085段</td> <td>1-10, 12, 14, 16</td> </tr> <tr> <td>PY</td> <td>CN 111062382 A (北京交通大学) 2020年 4月 24日 (2020 - 04 - 24) 权利要求1-8, 说明书第0006-0085段</td> <td>11, 13, 15, 16, 17</td> </tr> <tr> <td>X</td> <td>宋非洋等. “基于MobileNetV3的结构性剪枝优化” 自动化与信息工程, 第40卷, 第6期, 2019年 12月 15日 (2019 - 12 - 15), 第20-25页</td> <td>1-10, 12, 14, 16</td> </tr> <tr> <td>Y</td> <td>宋非洋等. “基于MobileNetV3的结构性剪枝优化” 自动化与信息工程, 第40卷, 第6期, 2019年 12月 15日 (2019 - 12 - 15), 第20-25页</td> <td>11, 13, 15, 16, 17</td> </tr> <tr> <td>Y</td> <td>CN 109344921 A (湖南极点智能科技有限公司) 2019年 2月 15日 (2019 - 02 - 15) 说明书第0019-0044段</td> <td>11, 13, 15, 16, 17</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	PX	CN 111325342 A (深圳中兴网信科技有限公司) 2020年 6月 23日 (2020 - 06 - 23) 权利要求第1-10, 说明书第0005-0029, 0042-0097段	1-10, 12, 14, 16	PY	CN 111325342 A (深圳中兴网信科技有限公司) 2020年 6月 23日 (2020 - 06 - 23) 权利要求第1-10, 说明书第0005-0029, 0042-0097段	11, 13, 15, 16, 17	PX	CN 111062382 A (北京交通大学) 2020年 4月 24日 (2020 - 04 - 24) 权利要求1-8, 说明书第0006-0085段	1-10, 12, 14, 16	PY	CN 111062382 A (北京交通大学) 2020年 4月 24日 (2020 - 04 - 24) 权利要求1-8, 说明书第0006-0085段	11, 13, 15, 16, 17	X	宋非洋等. “基于MobileNetV3的结构性剪枝优化” 自动化与信息工程, 第40卷, 第6期, 2019年 12月 15日 (2019 - 12 - 15), 第20-25页	1-10, 12, 14, 16	Y	宋非洋等. “基于MobileNetV3的结构性剪枝优化” 自动化与信息工程, 第40卷, 第6期, 2019年 12月 15日 (2019 - 12 - 15), 第20-25页	11, 13, 15, 16, 17	Y	CN 109344921 A (湖南极点智能科技有限公司) 2019年 2月 15日 (2019 - 02 - 15) 说明书第0019-0044段	11, 13, 15, 16, 17
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																								
PX	CN 111325342 A (深圳中兴网信科技有限公司) 2020年 6月 23日 (2020 - 06 - 23) 权利要求第1-10, 说明书第0005-0029, 0042-0097段	1-10, 12, 14, 16																								
PY	CN 111325342 A (深圳中兴网信科技有限公司) 2020年 6月 23日 (2020 - 06 - 23) 权利要求第1-10, 说明书第0005-0029, 0042-0097段	11, 13, 15, 16, 17																								
PX	CN 111062382 A (北京交通大学) 2020年 4月 24日 (2020 - 04 - 24) 权利要求1-8, 说明书第0006-0085段	1-10, 12, 14, 16																								
PY	CN 111062382 A (北京交通大学) 2020年 4月 24日 (2020 - 04 - 24) 权利要求1-8, 说明书第0006-0085段	11, 13, 15, 16, 17																								
X	宋非洋等. “基于MobileNetV3的结构性剪枝优化” 自动化与信息工程, 第40卷, 第6期, 2019年 12月 15日 (2019 - 12 - 15), 第20-25页	1-10, 12, 14, 16																								
Y	宋非洋等. “基于MobileNetV3的结构性剪枝优化” 自动化与信息工程, 第40卷, 第6期, 2019年 12月 15日 (2019 - 12 - 15), 第20-25页	11, 13, 15, 16, 17																								
Y	CN 109344921 A (湖南极点智能科技有限公司) 2019年 2月 15日 (2019 - 02 - 15) 说明书第0019-0044段	11, 13, 15, 16, 17																								
<p><input checked="" 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, 但为了理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>2021年 5月 17日</p>		<p>国际检索报告邮寄日期</p> <p>2021年 6月 8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) 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>刘宇儒</p> <p>电话号码 86-(10)-53961294</p>																								

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 107728620 A (江苏卡威汽车工业集团股份有限公司) 2018年 2月 23日 (2018 - 02 - 23) 说明书第0031-0054段	17
A	CN 110263841 A (南京信息工程大学) 2019年 9月 20日 (2019 - 09 - 20) 全文	1-17
A	CN 107368885 A (北京智芯原动科技有限公司) 2017年 11月 21日 (2017 - 11 - 21) 全文	1-17

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2021/079690

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 111325342 A	2020年 6月 23日	无	
CN 111062382 A	2020年 4月 24日	无	
CN 109344921 A	2019年 2月 15日	无	
CN 107728620 A	2018年 2月 23日	无	
CN 110263841 A	2019年 9月 20日	无	
CN 107368885 A	2017年 11月 21日	无	