

(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 3 日 (03.12.2020)



(10) 国际公布号
WO 2020/237688 A1

- (51) 国际专利分类号:
G06N 3/04 (2006.01)
- (21) 国际申请号: PCT/CN2019/089697
- (22) 国际申请日: 2019 年 5 月 31 日 (31.05.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 深圳市大疆创新科技有限公司 (SZ DJI TECHNOLOGY CO., LTD.) [CN/CN]; 中国广东省深圳市南山区高新南区粤兴一道9号香港科大深圳产学研大楼6楼, Guangdong 518057 (CN)。
- (72) 发明人: 蒋阳(JIANG, Yang); 中国广东省深圳市南山区高新南区粤兴一道9号香港科大深圳产学研大楼6楼, Guangdong 518057 (CN)。 庞磊(PANG, Lei); 中国广东省深圳市南山区高新南区粤兴一

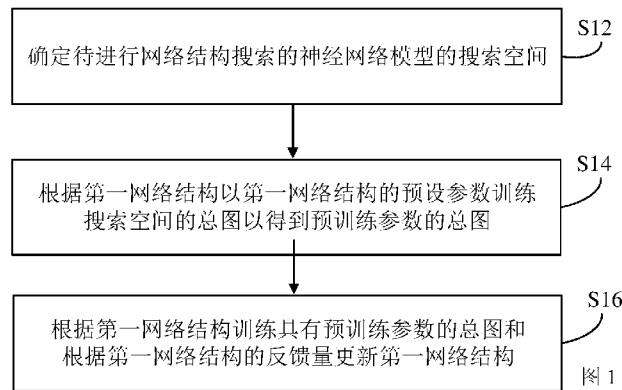
道9号香港科大深圳产学研大楼6楼, Guangdong 518057 (CN)。 胡湛(HU, Zhan); 中国广东省深圳市南山区高新南区粤兴一道9号香港科大深圳产学研大楼6楼, Guangdong 518057 (CN)。

(74) 代理人: 北京励诚知识产权代理有限公司 (BEIJING LISENG INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市海淀区北洼路45号2号楼3层301室, Beijing 100142 (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,

(54) Title: METHOD AND DEVICE FOR SEARCHING NETWORK STRUCTURE, COMPUTER STORAGE MEDIUM AND COMPUTER PROGRAM PRODUCT

(54) 发明名称: 网络结构搜索的方法及装置、计算机存储介质和计算机程序产品



S12 Determining the search space of the neural network model to be searched for the network structure

S14 Training the general map of the search space with the preset parameters of the first network structure according to the first network structure to obtain the general map with the pre-training parameters

S16 Training the general map with the pre-training parameters according to the first network structure and updating the first network structure according to the feedback amount of the first network structure

(57) Abstract: Provided is a method for searching network structure, including: the defining search space step: (step S12) determining the search space of the neural network model to be searched for the network structure, wherein the search space defines a variety of operations on the operation layer between every two nodes in the neural network model; the pre-training step: (step S14) training the general map of the search space with the preset parameters of the first network structure according to the first network structure to obtain the general map with the pre-training parameters, wherein the general map is composed of operations; the training step: (step

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))。

S16) training the general map with the pre-training parameters according to the first network structure and updating the first network structure according to the feedback amount of the first network structure. The present application also discloses a network structure search device, a computer storage medium and a computer program product.

(57) 摘要: 一种网络结构搜索的方法包括: 定义搜索空间步骤: (步骤S12)确定待进行网络结构搜索的神经网络模型的搜索空间, 搜索空间定义了神经网络模型中每两个节点之间的操作层上的多种操作; 预训练步骤: (步骤S14)根据第一网络结构以第一网络结构的预设参数训练搜索空间的总图以得到具有预训练参数的总图, 总图由操作构成; 训练步骤: (步骤S16)根据第一网络结构训练具有预训练参数的总图和根据第一网络结构的反馈量更新第一网络结构。本申请还公开了一种网络结构搜索的装置、计算机存储介质和计算机程序产品。

网络结构搜索的方法及装置、计算机存储介质和计算机程序产品

技术领域

本申请涉及机器学习领域，特别涉及一种网络结构搜索的方法及装置、计算机存储介
5 质和计算机程序产品。

背景技术

相关技术中，机器学习算法尤其深度学习算法近年来得到快速发展与广泛应用。随着
应用场景和模型结构变得越来越复杂，在应用场景中得到最优模型的难度越来越大，其中，
10 可以使用基于权值分享的高效网络结构搜索（Efficient Neural Architecture Search via
Parameter Sharing, ENAS）来提高网络结构搜索（Neural Architecture Search, NAS）的效
率。然而，通过 ENAS 搜索到的网络结构经常出现较大的偏见（bias），即，通过 ENAS 搜
索到的网络结构总会倾向于卷积核尺寸（kernel size）较大的操作。这导致搜索到的模型参
数较大，不易调试训练。另外，控制器带有偏见就意味着控制器收敛到的局部最优解，无
15 法充分探索（explore）搜索空间。这样的控制器不具有较高的可信度，不能保证搜索到的
模型就是全局最优。

发明内容

本申请的实施方式提供一种网络结构搜索的方法及装置、计算机存储介质和计算机程
20 序产品。

本申请实施方式的网络结构搜索的方法包括：

定义搜索空间步骤：确定待进行网络结构搜索的神经网络模型的搜索空间，所述搜索
空间定义了所述卷积神经网络中每两个节点之间的操作层上的多种操作；

预训练步骤：根据第一网络结构以所述第一网络结构的预设参数训练所述搜索空间的
25 总图以得到具有预训练参数的所述总图，所述总图由所述操作构成；

训练步骤：根据所述第一网络结构训练具有所述预训练参数的所述总图和根据所述第
一网络结构的反馈量更新所述第一网络结构。

本申请实施方式的网络结构搜索的装置包括处理器和存储器，所述存储器存储有一个
或多个程序，所述处理器用于定义搜索空间：确定待进行网络结构搜索的神经网络模型的
30 搜索空间，所述搜索空间定义了所述卷积神经网络中每两个节点之间的操作层上的多种操
作；及用于预训练：根据第一网络结构以所述第一网络结构的预设参数训练所述搜索空间
的总图以得到具有预训练参数的所述总图，所述总图由所述操作构成；以及用于训练：根

据所述第一网络结构训练具有所述预训练参数的所述总图和根据所述第一网络结构的反馈量更新所述第一网络结构。

本申请实施方式的计算机存储介质，其上存储有计算机程序，所述计算机程序被计算机执行时使得，所述计算机执行上述的方法。

5 本申请实施方式的包含指令的计算机程序产品，所述指令被计算机执行时使得计算机执行上述的方法。

10 本申请实施方式的网络结构搜索的方法及装置、计算机存储介质和计算机程序产品，在对总图和第一网络结构进行优化前，先以固定的第一网络结构的预设参数对总图进行预训练，使得具有预训练参数的总图得到充分的训练。预训练完成后，放开第一网络结构的参数对总图和第一网络结构进行训练，从而对网络结构和第一网络结构进行优化，避免从头开始训练时优化第一网络结构导致的偏见，提高第一网络结构的可信度，保证搜索到的模型是全局最优。

本申请的实施方式的附加方面和优点将在下面的描述中部分给出，部分将从下面的描述中变得明显，或通过本申请的实施方式的实践了解到。

15

附图说明

本申请的上述和/或附加的方面和优点从结合下面附图对实施方式的描述中将变得明显和容易理解，其中：

图 1 是本申请实施方式的网络结构搜索的方法的流程示意图；

20 图 2 是本申请实施方式的网络结构搜索的装置的模块示意图；

图 3 是相关技术的网络结构搜索的方法的原理示意图；

图 4 是本申请实施方式的网络结构搜索的方法的总图示意图；

图 5 是本申请另一实施方式的网络结构搜索的方法的流程示意图；

图 6 是本申请另一实施方式的网络结构搜索的方法的流程示意图；

25 图 7 是本申请又一实施方式的网络结构搜索的方法的流程示意图；

图 8 是本申请又一实施方式的网络结构搜索的方法的原理示意图；

图 9 是本申请再一实施方式的网络结构搜索的方法的流程示意图；

图 10 是本申请再一实施方式的网络结构搜索的方法的流程示意图；

图 11 是本申请再一实施方式的网络结构搜索的方法的流程示意图；

30 图 12 是本申请实施方式的网络结构搜索的方法的惩罚效果的示意图。

主要元件符号说明：

网络结构搜索的装置 10、存储器 102、处理器 104、通信接口 106。

具体实施方式

下面详细描述本申请的实施方式，所述实施方式的示例在附图中示出，其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施方式是示例性的，仅用于解释本申请，而不能理解为对本申请的限制。

5 在本申请的描述中，需要理解的是，术语“第一”、“第二”仅用于描述目的，而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此，限定有“第一”、“第二”的特征可以明示或者隐含地包括一个或者更多个所述特征。在本申请的描述中，“多个”的含义是两个或两个以上，除非另有明确具体的限定。

10 在本申请的描述中，需要说明的是，除非另有明确的规定和限定，术语“安装”、“相连”、“连接”应做广义理解，例如，可以是固定连接，也可以是可拆卸连接，或一体地连接；可以是机械连接，也可以是电连接或可以相互通信；可以是直接相连，也可以通过中间媒介间接相连，可以是两个元件内部的连通或两个元件的相互作用关系。对于本领域的普通技术人员而言，可以根据具体情况理解上述术语在本申请中的具体含义。

15 下文的公开提供了许多不同的实施方式或例子用来实现本申请的不同结构。为了简化本申请的公开，下文中对特定例子的部件和设置进行描述。当然，它们仅仅为示例，并且目的不在于限制本申请。此外，本申请可以在不同例子中重复参考数字和/或参考字母，这种重复是为了简化和清楚的目的，其本身不指示所讨论各种实施方式和/或设置之间的关系。

20 下面详细描述本申请的实施方式，所述实施方式的示例在附图中示出，其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施方式是示例性的，仅用于解释本申请，而不能理解为对本申请的限制。

请参阅图 1 和图 2，本申请实施方式提供一种网络结构搜索的方法及装置 10。

本申请实施方式的网络结构搜索的方法包括：

定义搜索空间步骤 S12：确定待进行网络结构搜索的神经网络模型的搜索空间，搜索空间定义了神经网络模型中每两个节点之间的操作层上的多种操作；

25 预训练步骤 S14：根据第一网络结构以第一网络结构的预设参数训练搜索空间的总图（whole graph）以得到预训练参数的总图，总图由操作构成；

训练步骤 S16：根据第一网络结构训练具有预训练参数的总图和根据第一网络结构的反馈量（ACC）更新第一网络结构。

30 本申请实施方式的网络结构搜索的装置 10 包括处理器 104 和存储器 102，存储器 102 存储有一个或多个程序，在程序被处理器执行的情况下，使得处理器 104 用于执行定义搜索空间步骤：确定待进行网络结构搜索的神经网络模型的搜索空间，搜索空间定义了神经网络模型中每两个节点之间的操作层上的多种操作；及用于执行预训练步骤：根据第一网

络结构以第一网络结构的预设参数训练搜索空间的总图以得到预训练参数的总图，总图由操作构成；以及用于执行训练步骤：根据第一网络结构训练具有预训练参数的总图和根据第一网络结构的反馈量更新第一网络结构。

也就是说，本申请实施方式的网络结构搜索的方法可由本申请实施方式的网络结构搜索的装置 10 实现。具体地，步骤 S12、步骤 S14 和步骤 S16 可以由处理器 104 实现。

本申请实施方式的网络结构搜索的方法及装置 10，在对网络结构和第一网络结构进行优化前，先以固定的第一网络结构的预设参数对总图进行预训练，使得具有预训练参数的总图得到充分的训练。预训练完成后，放开第一网络结构的参数对总图和第一网络结构进行训练，从而对网络结构和第一网络结构进行优化，避免从头开始训练时优化第一网络结构导致的偏见，提高第一网络结构的可信度，保证搜索到的模型是全局最优。

需要说明的是，处理器 104 的数量可以是一个。处理器 104 的数量也可以是多个，例如 2 个、3 个、5 个或其他数量。在处理器 104 的数量为多个的情况下，本申请网络结构搜索的方法的步骤流程可以由不同的处理器 104 分别执行，例如可以由不同的处理器 104 分别执行步骤 S14 和步骤 S16。

可选地，网络结构搜索的装置 10 还可以包括通信接口 106，用于将网络结构搜索的装置 10 处理完成的数据输出，和/或，从外部设备输入网络结构搜索的装置 10 将要处理的数据。例如，处理器 104 用于控制通信接口 106 输入和/输出数据。

近年来，机器学习算法，尤其是深度学习算法，得到了快速发展与广泛应用。随着模型性能不断地提高，模型结构也越来越复杂。在非自动化机器学习算法中，这些结构需要机器学习专家手工设计和调试，过程非常繁复。而且，随着应用场景和模型结构变得越来越复杂，在应用场景中得到最优模型的难度也越来越大。在这种情况下，自动化机器学习算法（AutoMachine Learning, AutoML）受到学术界与工业界的广泛关注，尤其是网络结构搜索（Neural Architecture Search, NAS）。

具体地，网络结构搜索是一种利用算法自动化设计神经网络模型的技术。网络结构搜索就是要搜索出神经网络模型的结构。在本申请实施方式中，待进行网络结构搜索的神经网络模型为卷积神经网络（Convolutional Neural Networks, CNN）。

网络结构搜索要解决的问题就是确定神经网络模型中的节点之间的操作。节点之间的操作的不同组合对应不同的网络结构。进一步地，神经网络模型中的节点可以理解为神经网络模型中的特征层。两个节点之间的操作指的是，其中一个节点上的特征数据变换为另一个节点上的特征数据所需的操作。本申请提及的操作可以为卷积操作、池化操作、或全连接操作等其他神经网络操作。可以认为两个节点之间的操作构成这两个节点之间的操作

层。通常，两个节点之间的操作层上具有多个可供搜索的操作，即具有多个候选操作。网络结构搜索的目的就是在每个操作层上确定一个操作。

例如，将 conv3*3, conv5*5, depthwise3*3, depthwise5*5, maxpool3*3, average pool3*3 等定义为搜索空间。也即是说，网络结构的每一层操作是在这六个选择中采样。

5 请参阅图 3，NAS 在建立搜索空间后，通常利用第一网络结构在搜索空间中采样到第二网络结构，然后将第二网络结构训练到收敛，以确定反馈量，最后利用反馈量更新第一网络结构。

具体地，NAS 的思想是，通过一个第一网络结构在搜索空间中得到一个网络结构，然后根据该网络结构得到准确率 R，将准确率 R 作为反馈以更新第一网络结构，第一网络结构继续优化得到另一个网络结构，如此反复进行直到得到最佳的结果。在图 3 的示例中，
10 第一网络结构通过循环神经网络（Recurrent Neural Network, RNN）构建，当然，第一网络结构也可以通过卷积神经网络（Convolutional Neural Networks, CNN）或长短期记忆人工神经网络（Long-Short Term Memory, LSTM）构建。在此不对第一网络结构构建的具体方式进行限定。

15 然而，将第二网络结构训练到收敛比较耗时。对此，相关技术出现了多种解决 NAS 高效性的方法，例如，基于网络变换的高效网络结构搜索（Efficient Architecture Search by Network Transformation），以及基于权值分享的高效网络结构搜索（Efficient Neural Architecture Search via Parameter Sharing, ENAS）。其中，基于权值分享的高效网络结构搜索应用较为广泛。

20 具体地，请参阅图 4，基于权值分享的高效网络结构搜索在使用过程中，将操作连接成总图，搜索到的最终最优结构是总图中的其中一个子图。在图 4 的示例中，总图由节点间的操作连接而成。图 4 中加粗的带边最优结构的连接方式是总图的一个子图。

ENAS 采用权值分享策略，在每次采样到一个网络结构后，不再将其直接训练至收敛，而是训练一个批（batch），迭代多次后，总图最终可以收敛。请注意，图的收敛并不相当于
25 网络结构的收敛。

在训练完所述总图后，可以将总图的参数固定（fix）住，然后训练第一网络结构。具体地，可以对总图进行采样以得到第二网络结构，根据第二网络结构进行预测（prediction）以训练第一网络结构的参数并得到第一网络结构的反馈量，从而更新第一网络结构。

可以理解，基于权值分享的高效网络结构搜索，由于在每次搜索网络结构时，分享了
30 可以分享的参数，可以节约时间，从而提高网络结构搜索的效率。例如，在图 4 的示例中，如果在搜索到节点 1、节点 3 和节点 6 并对搜索到的网络结构进行训练之后，本次搜索到节点 1、节点 2、节点 3 和节点 6，那么，搜索到节点 1、节点 3 和节点 6 时训练的网络结

构的相关参数可以应用到对本次搜索到的网络结构的训练中。这样，就可以实现通过权值分享提高效率。

ENAS 可以将 NAS 的效率提升 1000 倍以上，但是，在实际使用的过程中，会出现如下问题：搜索的网络经常出现较大的偏见（bias），即，通过 ENAS 搜索到的网络结构总会
5 倾向于卷积核尺寸（kernel size）较大的操作。

以上述搜索空间为例，将 conv3*3, conv5*5, depthwise3*3, depthwise5*5, maxpool3*3, average pool3*3 等定义为搜索空间，那么第一网络结构总会搜索到 conv5*5。这将导致两个较为严重的问题：第一，搜索到的模型参数较大。这样即使计算力足够，对模型参数没有限制，参数较多的模型也易于过拟合，泛化能力下降，不易调试训练。第二，第一网络结构
10 带有偏见就意味着第一网络结构收敛到的是局部最优解，无法充分探索（explore）搜索空间。这样的第一网络结构不具有较高的可信度，不能保证我们搜索到的模型就是全局最优。

上述问题较为严重，需要十分精细的调节第一网络结构使其不会收敛至局部最优解。否则，该问题将使 ENAS 框架失效。但是精细调节第一网络结构将使 ENAS 变得十分繁琐，
15 失去的 AutoML 的初衷，变得不易使用。而且精细调节也不能保证得到较好的第一网络结构。

基于此，本申请实施方式的网络结构搜索的方法及装置 10，在对总图和第一网络结构进行优化前，先以固定的第一网络结构的预设参数对总图进行预训练，使得具有预训练参数的总图得到充分的训练。预训练完成后，放开第一网络结构的参数对总图和
20 第一网络结构进行训练，从而对网络结构和第一网络结构进行优化，避免从头开始训练时优化第一网络结构导致的偏见，提高第一网络结构的可信度，保证搜索到的模型是全局最优。

可以理解，传统的 CNN 中，由网络结构搜索得到的神经网络结构一般可以通过样本数据进行训练和验证。其中，样本数据包括验证样本和训练样本，验证样本可以用于验证网络结构好不好，训练样本可以用于训练网络结构搜索方法搜索到的网络结构。
25

在本实施方式中，训练样本还可以包括训练集（train）和测试集（valid）。

也即是说，在本实施方式中，训练样本可以划分为训练集和测试集。具体地，训练集和测试集可以由用户进行划分，或者由网络结构搜索的装置 10 进行划分。其中，训练集由网络结构搜索的装置 10 进行划分时，处理器 104 可以用于将训练样本分为训练集和测试集。

在本实施方式中，第一网络结构通过 LSTM 构建。在搜索网络结构时，训练集用于训练搜索结构的参数，如通过 conv3*3, sep5*5 计算出的结构的参数。而 Cell 用于搜索网络结构，为看搜索到的网络结构的泛化能力，会把 LSTM 的参数在测试集上训练。也即是说，
30

训练集用于训练搜索结构的参数，测试集用于训练 LSTM 的参数。而验证样本用于验证训练后搜索到的网络结构好不好。

在一个例子中，训练样本的数量为 10 个，将训练样本划分为数量为 8 个的训练集和数量为 2 个的测试集，数量为 8 个的训练集用于训练搜到的网络结构，数量为 2 个的测试集用于训练 LSTM。

需要说明的是，与传统的 CNN 一样，本实施方式可以由验证样本用于验证网络结构好不好，也即是说，本实施方式的验证样本可以保持不变。训练样本中训练集可以用于预训练步骤 S14 和训练步骤 S16，也即是说，预训练步骤 S14 和训练步骤 S16 可以在同一个训练集上进行训练。

可选地，请参阅图 5，步骤 S14 包括：

步骤 S142：将第一网络结构的参数设置为第一网络结构的预设参数。

具体地，步骤 S142 可以由处理器 104 实现，也即是说，处理器 104 可以用于将第一网络结构的参数设置为第一网络结构的预设参数。

如此，实现预训练过程中的第一网络结构的参数设置，此时，预训练过程中的第一网络结构的参数是固定的。

可选地，步骤 S14 还包括：

步骤 S144：根据第一网络结构以第一网络结构的预设参数在搜索空间的每个操作层采样一个操作以得到总图的一个子图；和

步骤 S146：利用训练集的一批训练数据训练总图的子图以得到具有预训练参数的总图。

具体地，步骤 S144 和步骤 S146 可以由处理器 104 实现，也即是说，处理器 104 可以用于根据第一网络结构以第一网络结构的预设参数在搜索空间的每个操作层采样一个操作以得到总图的一个子图，以及用于利用训练集的一批训练数据训练总图的子图以得到具有预训练参数的总图。

如此，实现对总图的预训练。在本实施方式中，预训练时，第一网络结构以固定的预设参数采样操作，可以使得采样到每种操作的概率呈均匀分布。也即是说，在预训练过程中，第一网络结构以固定的预设参数采样相应的操作时，采样到每种操作的概率是相等的，卷积核尺寸不会影响第一网络结构的采样到各个操作的概率。

可选地，步骤 S144 和步骤 S146 迭代进行至完成第一迭代总次数。

可以理解，步骤 S144 和步骤 S146 每次可对一个子图进行预训练，即每次预训练后具有预训练参数的总图中总有一个子图的参数发生变化，因此，步骤 S144 和步骤 S146 迭代进行可以使得多个子图可以得到预训练，通过控制预训练的第一迭代总次数，可以使得总图中每个子图都能够被采样到，从而总图可以得到充分的预训练，也即得到具有预训练参

数的总图。

在一个例子中，搜索空间为 5 层，每层有 4 个可选用的操作，相当于 4X5 的图。网络结构搜索需要在每层选一个操作，相当于在图上进行路径优化。初始时，步骤 S144 在搜索空间的每个操作层随机采样一个操作，然后把采样到的操作连起来，得到一个网络结构，
5 这个网络结构为搜索空间的总图中的一个子图，步骤 S146 在训练集的一批测试数据上训练这个网络结构；接着，重复步骤 S144 再在每层随机采样一个操作得到另一个网络结构，步骤 S146 再在训练集的另一批测试数据上训练这个网络结构。

其中，训练集可以划分为多批训练数据，不断重复步骤 S144 和步骤 S146，直到训练集的多批训练数据使用完毕，也即是一个迭代 (epoch)。在一个 epoch 后，随机采样到了
10 总图中的部分子图，也对相应的子图进行了预训练，接着，以相同的方式训练总图完成第二个 epoch……如此，进行多次迭代，即可实现对总图充分的预训练。

可以理解，通过设置合适的第一迭代总次数，可以使得对总图进行预训练后，总图中每个子图都已经受到过训练，得到具有预训练参数的总图。

在本实施方式中，第一迭代总次数可以是 310 次。可以理解，在其他的实施方式中，
15 第二迭代总次数的数值可为 100 次、200 次或其他数值。

可选地，请参阅图 6，步骤 S16 包括：

训练总图步骤 S162，根据第一网络结构训练具有预训练参数的总图；和

训练第一网络结构步骤 S164，确定第一网络结构的反馈量并根据第一网络结构的反馈量更新第一网络结构。

具体地，步骤 S162 和步骤 S164 可以由处理器 104 实现，也即是说，处理器 104 可以用于根据第一网络结构训练具有预训练参数的总图，以及用于确定反馈量并根据反馈量更新第一网络结构。
20

如此，在预训练的基础上实现对网络结构的优化。

可选地，步骤 S162 和步骤 S164 迭代进行至完成第二迭代总次数。

如此，实现将总图和第一网络结构交替进行优化。具体地，第二迭代总次数可以使得总图和第一网络结构得到充分训练。
25

可选地，请参阅图 7，步骤 S162 包括：

步骤 S1622：根据第一网络结构在搜索空间的每个操作层采样一个操作以得到具有预训练参数的总图的一个子图；和

步骤 S1624：利用训练集的一批训练数据训练采样到的子图。
30

具体地，步骤 S1622 和步骤 S1624 可以由处理器 104 实现，也即是说，处理器 104 可以用于根据第一网络结构在搜索空间的每个操作层采样一个操作以得到具有预训练参数的

总图的一个子图，以及用于利用训练集的一批训练数据训练采样到的子图。

如此，实现对总图的训练。在本实施方式中，ENAS 采用权值分享策略，在每次采样到一个网络结构后，不再将其直接训练至收敛，而是训练一个批，然后进行第一网络结构的训练。请注意，图的收敛并不相当于网络结构的收敛。

5 由于总图已经在预训练过程中受到过训练，此时，可以放开第一网络结构的参数，即步骤 S1622 中可以不以第一网络结构的预设参数进行采样，相应地，在具有预训练参数的总图的操作层中，采样到各个操作的概率可以不相同。

10 相应地，步骤 S1622 和步骤 S1624 可以循环进行，在一个例子中，初始时，每层采样一个操作，然后把采样到的操作连起来，得到一个子图，在训练集的一批数据上训练这个子图；接着，再每层采样一个操作得到另一个子图，再在训练集的另一批数据上训练这个子图；接着，继续每层采样到又一个子图并在训练集的又一批数据上训练这个子图……直到训练集中的数据使用完毕，也即是完成一个 epoch。

当训练总图步骤 S162 完成一个 epoch 后，固定总图的参数，进入训练第一网络结构步骤 S164 从而实现总图和第一网络结构交替进行优化。

15 接着，以相同的方式训练总图完成第二个 epoch，固定总图的参数，然后训练第一网络结构。

接着，以相同的方式训练总图完成第三个 epoch，固定总图的参数，然后训练第一网络结构……如此迭代，直至完成第二迭代总次数，以将总图和第一网络结构交替进行优化。

20 需要说明的是，在执行训练总图步骤 S162 中，总图的参数随着训练的进行而变化，也就是说，每次迭代中具有预训练参数的总图在训练的过程中总图参数持续地更新直至完成一个 epoch。

在上述讨论的实施方式中，训练总图步骤 S162 每完成一个 epoch，再固定总图的参数，然后进入训练第一网络结构步骤 S164，当然，在其他实施方式中，训练总图步骤 S162 可以是每完成 10 个 epoch，再固定总图的参数，然后进入训练第一网络结构步骤 S164。上述
25 列出步骤 S162 每次完成 epoch 的数量仅作为示例，不能解释为对本申请的限制，可以理解，步骤 S162 每次完成 epoch 的数量可以根据实际情况进行设置，在此不做具体限定。

在本实施方式中，第二迭代总次数为 310 次。可以理解，在其他的实施方式中，第二迭代总次数的数值可为 100 次、200 次或其他数值。

30 需要说明的是，在本实施方式中，第一迭代总次数和第二迭代总次数是相同的，而在其他实施方式中，第一迭代总次数和第二迭代总次数可以是不相同的，在此不做具体限定。

请参阅图 8，搜索空间的每个操作层对应于长短期记忆人工神经网络（LSTM）的一个时间步（timestep），对于每个时间步，长短期记忆人工神经网络的细胞（Cell）输出一个隐

状态 (hidden state), 步骤 S1622 包括:

将隐状态映射为特征向量, 特征向量的维度与每个操作层上的操作数量相同;

根据特征向量在每个操作层采样一个操作以得到总图的子图。

具体地, 处理器 104 可以用于将隐状态映射为特征向量, 以及用于根据特征向量在每
5 个操作层采样一个操作以得到总图的子图。其中, 特征向量的维度与每个操作层上的操作
数量相同。

如此, 实现在搜索空间的每个操作层采样一个操作以得到总图的子图。例如, 一共要
搜索一个 20 层的网络, 不考虑跳线, 需要 20 个时间步。

在图 8 的示例中, 实线箭头表示时间步 (timestep), 时间 1 表示 LSTM 的第一个 Cell,
10 时间 2 表示 LSTM 的第二个 Cell……以此类推。方块 conv3*3 表示在模型中该层的操作,
圆形表示操作层与操作层之间的连接关系。

可以理解, 由于网络结构的计算有先后顺序, 将计算先后的关系的逻辑关系映射到
LSTM 上, 就是图 8 中一个小方块从左往右, 对应的每一个时间的 LSTM 的 cell 的状态。

具体地, 在时间 1 下, cell 输出的隐状态经过计算得到卷积 conv3×3, conv3×3 作为
15 时间 2 下 cell 的输入层, 时间 1 下 cell 输出的隐状态也作为时间 2 下 cell 的输入, 计算得
到圆圈 1。

同理, 圆圈 1 作为时间 3 下 cell 的输入, 时间 2 下 cell 输出的隐状态也作为时间 3 的
输入, 计算得到卷积 sep5×5……以此类推。

进一步地, 根据特征向量在每个操作层采样一个操作以得到网络结构的步骤包括:

20 将特征向量进行归一化 (softmax) 以得到每个操作层的每个操作的概率;

根据概率在每个操作层采样一个操作以得到网络结构。

如此, 实现根据特征向量在每个操作层采样一个操作以得到网络结构。具体地, 在图
8 所示的例子中, 对 LSTM 的 cell 输出的隐状态进行编码 (encoding) 操作, 将其映射维度
为 6 的向量 (vector), 该向量经过归一化指数函数 (softmax), 变为概率分布, 依据此概
25 率分布进行采样, 得到当前层的操作。以此类推最终得到一个网络结构。可以理解, 在此
例子中, 只有一个输入, 一共包含六种操作 (3×3 卷积, 5×5 卷积, 3×3depthwise-separable
卷积, 5×5depthwise-separable 卷积, max pooling, 3×3average pooling), 向量的维度与搜
索空间对应, 6 是指搜索空间有 6 个操作可选择。

请参阅图 9, 步骤 S164 中, 确定反馈量的步骤包括:

30 步骤 S1642: 根据第一网络结构在搜索空间的每个操作层采样一个操作以得到第二网
络结构;

步骤 S1644: 利用测试集的一批测试数据预测第二网络结构以确定第一网络结构的反

馈量。

对应地，步骤 S1642 和步骤 S1644 可以由处理器 104 实现，也即是说，处理器 104 可以用于根据第一网络结构在搜索空间的每个操作层采样一个操作以得到第二网络结构，以及用于利用测试集的一批测试数据预测第二网络结构以确定第一网络结构的反馈量。

5 如此，可以根据搜索到的第二网络结构在测试集上进行预测从而得到第一网络结构的反馈量。

在本实施方式中，在搜索到第二网络结构后，可将搜索到的第二网络结构在测试集上预测，以得到反馈量来根据前述公式更新第一网络结构。请注意，并非直接用测试集训练 LSTM。

10 可选地，训练第一网络结构步骤 S164 循环预设次数，步骤 S164 包括：

步骤 S1646，利用每次循环确定的第一网络结构的反馈量更新第一网络结构，每次循环确定第一网络结构的反馈量的数量为预设数量。

对应地，处理器 104 用于将第一网络结构循环训练预设次数，以及用于利用每次循环确定的反馈量更新第一网络结构。其中，每次循环确定的反馈量的数量为预设数量。

15 如此，实现训练第一网络结构。具体地，在本实施方式中，步骤 S16 循环 50 次，也即是说，训练总图步骤 S162 和训练第一网络结构步骤 S164 迭代进行训练过程中，每次训练第一网络结构步骤 164 执行时，第一网络结构更新 50 次从而实现第一网络结构的优化。训练第一网络结构步骤 164 完成后，下迭代中训练总图步骤 162 以最后更新的第一网络结构训练总图。

20 可以理解，在其他的例子中，预设次数可为 10、20、30 或其他数值，在此不对预设次数的具体数值进行限定。

可选地，步骤 S1642 和步骤 S1644 循环进行以获得第一网络结构预设数量的反馈量。

可以理解，每次循环时，都可以采样到一个第二网络结构，这个网络结构在测试集的一批训练数据上进行测试后，可以得到一个关于第一网络结构的反馈量，循环进行足够次
25 数后可以获得第一网络结构预设数量的反馈量。

在一个例子中，预设数量为 20，即第一网络结构的反馈量为 20 个。可以理解，在其他的例子中，预设数量可为 10、15、25 或其他数值。在此不对预设数量的具体数值进行限定。

30 可选地，第一网络结构是根据长短期记忆网络模型来构建，步骤 S164 通过以下条件式实现：

$$\frac{1}{m} \sum_{k=1}^m \sum_{t=1}^T \nabla \theta_c \log P(a_t | a_{(t-1):1}; \theta_c) R_k ;$$

其中， R_k 为第 k 个反馈量， θ_c 为长短期记忆网络模型的参数， a_t 为在第 t 个操作层采

样到的所述操作, $P(a_t|a_{(t-1):1};\theta_c)$ 为采样到操作的概率, m 为反馈量的总数, T 为第一网络结构预测的超参数的数量。

如此, 实现根据多个反馈量的平均值更新第一网络结构。在本实施方式中, T 包括操作层和跳线。在其他的实施方式中, 还可能包含其他想要优化的超参数。在此不对 T 的具体内容进行限定。

在本实施方式中, 对总图的训练和对第一网络结构的更新是多次迭代进行的, 总图与第一网络结构交替训练的第二迭代总次数为 310 次。也即是说, 步骤 S14 和步骤 S16 迭代 310 次。如此, 可以最终获得效果较好的第二网络结构。

在每次交替训练时, 循环步骤 S16 的预设次数为 50 次。也即是说, 在步骤 S14 和步骤 S16 的每次迭代中, 第一网络结构更新 50 次, 根据上述条件式执行 50 次。可以理解, 将步骤 S16 的循环次数选取为 50, 可以减少采样带来的随机性优化。

每次循环步骤 S16 时, 采样的第二网络结构的数量为 20 个, 以得到 20 个反馈量, 将 20 个反馈量作为 R_k 代入上述条件式, 以实现第一网络结构的更新。也即是说, 上述条件式中, m 的值为 20。

请注意, 训练样本遍历一遍, 即为迭代一次。例如, 训练样本的数量为 10 个, 每次划分出 2 个样本作为训练集的一批来训练子图, 5 批后, 训练样本使用完毕, 则完成一次迭代。

可选的, 请参阅图 10, 步骤 S164 中, 确定反馈量的步骤包括:

步骤 S1646: 根据预设的惩罚模型调整第一网络结构的反馈量。

对应地, 步骤 S1646 可以由处理器 104 实现, 也即是说, 处理器 104 可以用于根据预设的惩罚模型调整第一网络结构的反馈量。

可以理解, 预训练过程可以不将总图训练到充分收敛, 此时, 在预训练过程中未能使得总图充分收敛的情况下, 根据预设的惩罚模型调整反馈量, 可以避免在总图未充分收敛时优化第一网络结构导致的偏见, 从而进一步提高第一网络结构的可信度, 保证搜索到的模型是全局最优。

可选地, 请参阅图 11, 步骤 S1646 包括:

步骤 S1648: 根据预设信息、第二网络结构、当前的迭代次数和第二迭代总次数确定惩罚项;

步骤 S1649: 根据惩罚项调整反馈量。

对应地, 步骤 S1648 和步骤 S1649 可以由处理器 104 实现, 也即是说, 处理器 104 用于根据预设信息、第二网络结构、当前的迭代次数和迭代总次数确定惩罚项; 以及用于根据惩罚项调整反馈量。

如此，实现根据惩罚模型调整反馈量。可以理解，为避免在搜索过程中，总图未充分收敛的情况下，优化第一网络结构引入偏见。我们给较大的卷积核加入惩罚项，对预测时的反馈量进行处理。

具体地，步骤 S1648 中的预设信息，可以是用户预先输入的信息，也可以是其他计算
5 模型将计算的结果输入。在此不对预设信息的来源进行限定。

请注意，在本实施方式中，步骤 S162 和步骤 S164 迭代进行。第二迭代总次数指步骤 S12 和步骤 S164 迭代的总次数。当前的迭代次数是指在步骤 S162 和步骤 S164 迭代过程中，当前的迭代是第几次迭代。例如，第二迭代总次数为 310，在第一次迭代时，当前的迭代次数为 1；在第二次迭代时，当前的迭代次数为 2；在第三次迭代时，当前的迭代次数为 3……
10 以此类推，步骤 S162 和步骤 S164 迭代完 310 次即停止迭代。

因此，第二迭代总次数和每次迭代中当前的迭代次数也是可以确定的。其中，第二迭代总次数可以由用户进行设定。当前的迭代次数随着迭代的进行是可以累计得到的。

请一并参阅图 12，由于随着迭代的进行，总图逐渐收敛。可以理解，惩罚项的值随着总图的收敛过程而增大，其中，总图在训练的后期收敛了，应该予以正常的反馈量，不再
15 对反馈量进行惩罚，因此，在总图收敛时，惩罚项的值为 1。请注意，随着步骤 S162 和步骤 S164 的迭代的进行，总图逐渐收敛，在迭代完第二迭代总次数时，总图呈收敛状态。

从图 12 中可以看出，惩罚项的数值可以从 0.76 逐步过渡到 1。对应地，调整后的反馈量，逐步从 $ACC \times 0.76$ 到 ACC 。

在一些实施例中，惩罚项的大小选取可以与预训练时总图的收敛程度相关，此时，预设信息可以包括预训练的第一迭代总次数，也即是说，步骤 S1648 可以根据预训练的第一
20 迭代总次数、第二网络结构、当前的迭代次数和第二迭代总次数确定惩罚项。

本申请实施方式还提供一种计算机存储介质，其上存储有计算机程序，计算机程序被计算机执行时使得，计算机执行上述任一实施方式的方法。

本申请实施方式还提供一种包含指令的计算机程序产品，指令被计算机执行时使得计
25 算机执行上述任一实施方式的方法。

本申请实施方式的计算机存储介质和计算机程序产品，将网络结构和第一网络结构交替进行优化，并根据预设的惩罚模型调整反馈量，可以避免在总图充分收敛时优化第一网络结构导致的偏见，从而提高第一网络结构的可信度，保证搜索到的模型是全局最优。

在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其他任意组合来实现。
30 当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机程序指令时，全部或部

分地产生按照本申请实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一个计算机可读存储介质传输，例如，所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线（例如同轴电缆、光纤、
5 数字用户线（digital subscriber line, DSL））或无线（例如红外、无线、微波等）方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质（例如，软盘、硬盘、磁带）、光介质（例如数字视频光盘（digital video disc, DVD））、或者半导体介质（例如固态硬盘（solid state disk, SSD））等。
10

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、或者计算机软件和电子硬件的结合来实现。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
15 申请的范围。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统、装置和方法，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。另一点，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，可以是电性，机械或其它的形式。
20

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，也可以两个或两个以上单元集成在一个单元中。
25

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，可轻易想到变化或替换，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所述权利要求的保护范围为准。

权利要求书

1. 一种网络结构搜索的方法，其特征在于，包括：

定义搜索空间步骤：确定待进行网络结构搜索的神经网络模型的搜索空间，所述搜索空间定义了所述神经网络模型中每两个节点之间的操作层上的多种操作；

预训练步骤：根据第一网络结构以所述第一网络结构的预设参数训练所述搜索空间的总图以得到具有预训练参数的所述总图，所述总图由所述操作构成；

训练步骤：根据所述第一网络结构训练具有所述预训练参数的所述总图和根据所述第一网络结构的反馈量更新所述第一网络结构。

2. 根据权利要求 1 所述的网络结构搜索的方法，其特征在于，所述预训练步骤，包括：

根据所述第一网络结构以所述第一网络结构的预设参数在所述搜索空间的每个所述操作层采样一个所述操作以得到所述总图的一个子图；

利用训练集的一批训练数据训练所述总图的子图以得到具有所述预训练参数的所述总图。

3. 根据权利要求 1 所述的网络结构搜索的方法，其特征在于，所述训练步骤包括：

训练总图步骤：根据所述第一网络结构训练具有所述预训练参数的所述总图；

训练第一网络结构步骤：确定所述反馈量并根据所述反馈量更新所述第一网络结构。

4. 根据权利要求 3 所述的网络结构搜索的方法，其特征在于，所述训练总图步骤和所述训练第一网络结构步骤迭代进行。

5. 根据权利要求 3 所述的网络结构搜索的方法，其特征在于，所述训练总图步骤包括：

根据所述第一网络结构在所述搜索空间的每个所述操作层采样一个所述操作以得到具有所述预训练参数的所述总图的一个子图；

利用训练集的一批训练数据训练所述子图。

6. 根据权利要求 3 所述的网络结构搜索的方法，其特征在于，所述训练第一网络结构步骤循环预设次数，根据所述反馈量更新所述第一网络结构，包括：

利用每次循环确定的所述反馈量更新所述第一网络结构，每次循环确定的所述反馈量的数量为预设数量。

7. 根据权利要求 6 所述的网络结构搜索的方法, 其特征在于, 确定所述反馈量, 包括:
根据所述第一网络结构在所述搜索空间的每个所述操作层采样一个所述操作以得到第二网络结构;

5 利用测试集的一批测试数据预测所述第二网络结构以确定所述反馈量。

8. 根据权利要求 6 所述的网络结构搜索方法, 其特征在于, 所述第一网络结构是根据长短期记忆网络模型来构建, 根据所述反馈量更新所述第一网络结构, 通过以下条件式实现:

$$10 \quad \frac{1}{m} \sum_{k=1}^m \sum_{t=1}^T \nabla \theta_c \log P(a_t | a_{(t-1):1}; \theta_c) R_k;$$

其中, R_k 为第 k 个所述反馈量, θ_c 为所述长短期记忆网络模型的参数, a_t 为在第 t 个所述操作层采样到的所述操作, $P(a_t | a_{(t-1):1}; \theta_c)$ 为采样到所述操作的概率, m 为所述反馈量的总数, T 为所述第一网络结构预测的超参数的数量。

15 9. 一种网络结构搜索的装置, 其特征在于, 包括处理器和存储器, 所述存储器存储有一个或多个程序, 在所述程序被处理器执行的情况下, 使得所述处理器用于执行:

定义搜索空间步骤: 确定待进行网络结构搜索的神经网络模型的搜索空间, 所述搜索空间定义了所述神经网络模型中每两个节点之间的操作层上的多种操作;

预训练步骤: 根据第一网络结构以所述第一网络结构的预设参数训练所述搜索空间的
20 总图以得到具有预训练参数的所述总图, 所述总图由所述操作构成;

训练步骤: 根据所述第一网络结构训练具有所述预训练参数的所述总图和根据所述第一网络结构的反馈量更新所述第一网络结构。

10. 根据权利要求 9 所述的网络结构搜索的装置, 其特征在于, 所述处理器用于根据
25 所述第一网络结构以所述第一网络结构的预设参数在所述搜索空间的每个所述操作层采样一个所述操作以得到所述总图的一个子图; 以及用于利用训练集的一批训练数据训练所述总图的子图以得到具有所述预训练参数的所述总图。

11. 根据权利要求 9 所述的网络结构搜索的装置, 其特征在于, 所述处理器用于训练
30 总图: 根据所述第一网络结构训练具有所述预训练参数的所述总图; 以及用于训练第一网络结构: 确定所述反馈量并根据所述反馈量更新所述第一网络结构。

12. 根据权利要求 11 所述的网络结构搜索的装置，其特征在于，所述处理器用于迭代进行所述训练总图和所述训练第一网络结构。

13. 根据权利要求 11 所述的网络结构搜索的装置，其特征在于，所述处理器用于根据
5 所述第一网络结构在所述搜索空间的每个所述操作层采样一个所述操作以得到具有所述预训练参数的所述总图的一个子图，以及用于利用训练集的一批训练数据训练所述子图。

14. 根据权利要求 11 所述的网络结构搜索的装置，其特征在于，所述处理器用于将所
10 述训练第一网络结构步骤循环预设次数，以及用于利用每次循环确定的所述反馈量更新所
述第一网络结构，每次循环确定的所述反馈量的数量为预设数量。

15 15. 根据权利要求 14 所述的网络结构搜索的装置，其特征在于，所述处理器用于根据
第一网络结构在所述搜索空间的每个所述操作层采样一个所述操作以得到第二网络结构；
以及用于利用测试集的一批测试数据预测所述第二网络结构以确定所述反馈量。

16. 根据权利要求 14 所述的网络结构搜索的装置，其特征在于，所述第一网络结构是
根据长短期记忆网络模型来构建，所述处理器用于利用所述反馈量更新所述第一网络结构，
通过以下条件式实现：

$$\frac{1}{m} \sum_{k=1}^m \sum_{t=1}^T \nabla_{\theta_c} \log P(a_t | a_{(t-1):1}; \theta_c) R_k ;$$

20 其中， R_k 为第 k 个所述反馈量， θ_c 为所述长短期记忆网络模型的参数， a_t 为在第 t 个
所述操作层采样到的所述操作， $P(a_t | a_{(t-1):1}; \theta_c)$ 为采样到所述操作的概率， m 为所述反馈量
的总数， T 为所述第一网络结构预测的超参数的数量。

25 17. 一种计算机存储介质，其特征在于，其上存储有计算机程序，所述计算机程序被
计算机执行时使得，所述计算机执行权利要求 1 至 8 中任一项所述的方法。

18. 一种包含指令的计算机程序产品，其特征在于，所述指令被计算机执行时使得所
述计算机执行权利要求 1 至 8 中任一项所述的方法。

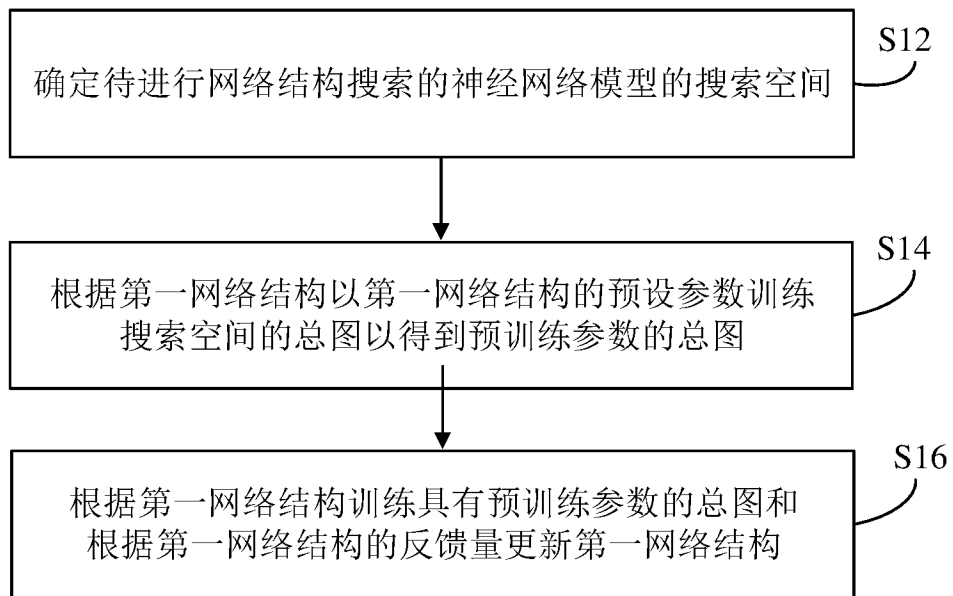


图 1

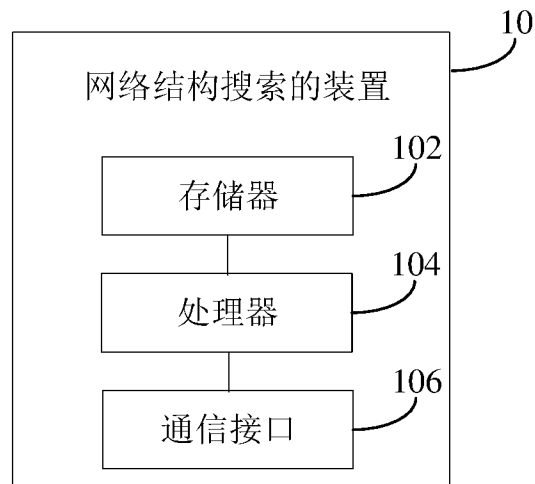


图 2

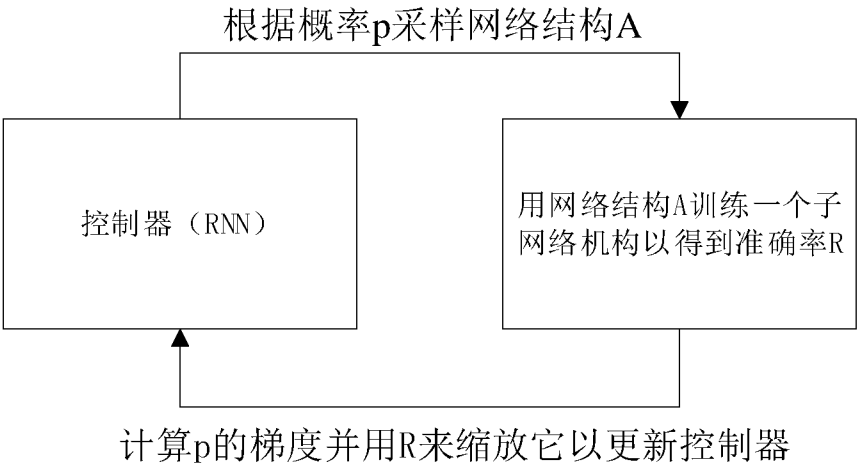


图 3

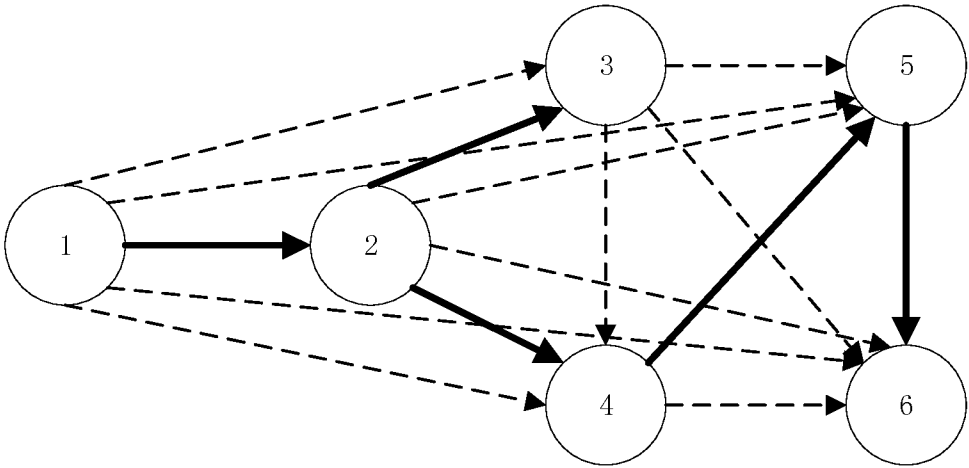


图 4

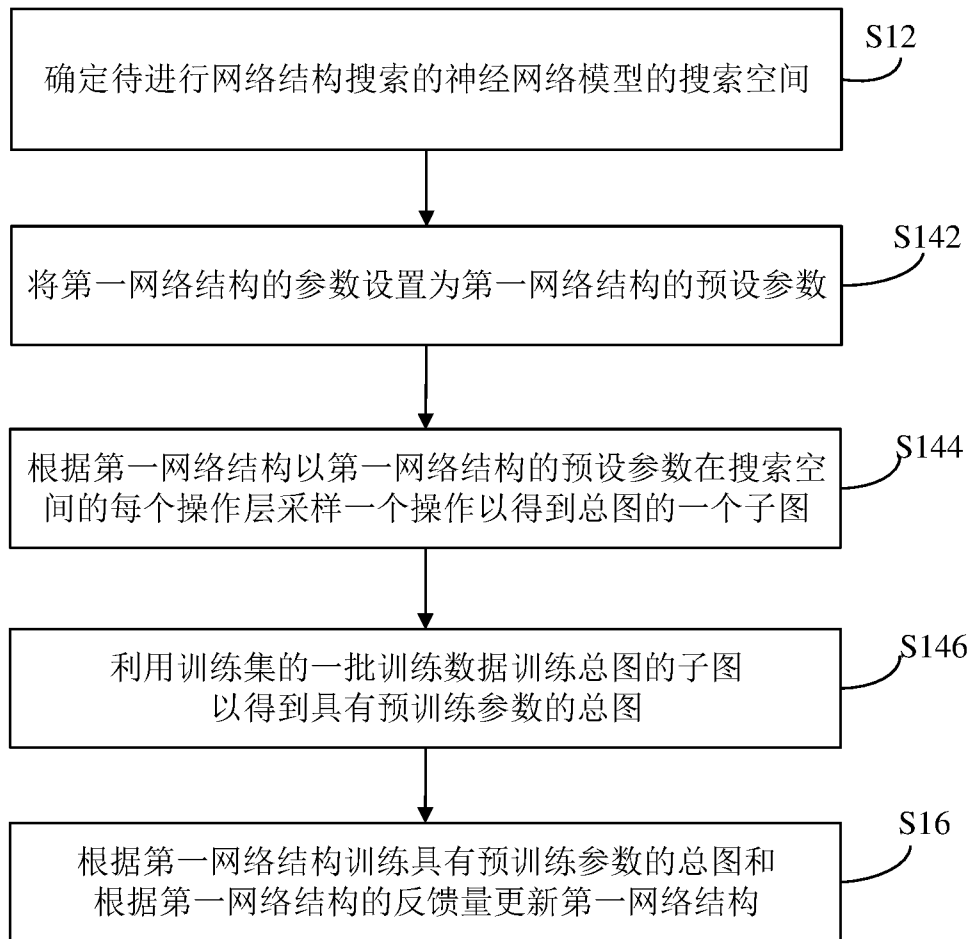


图 5

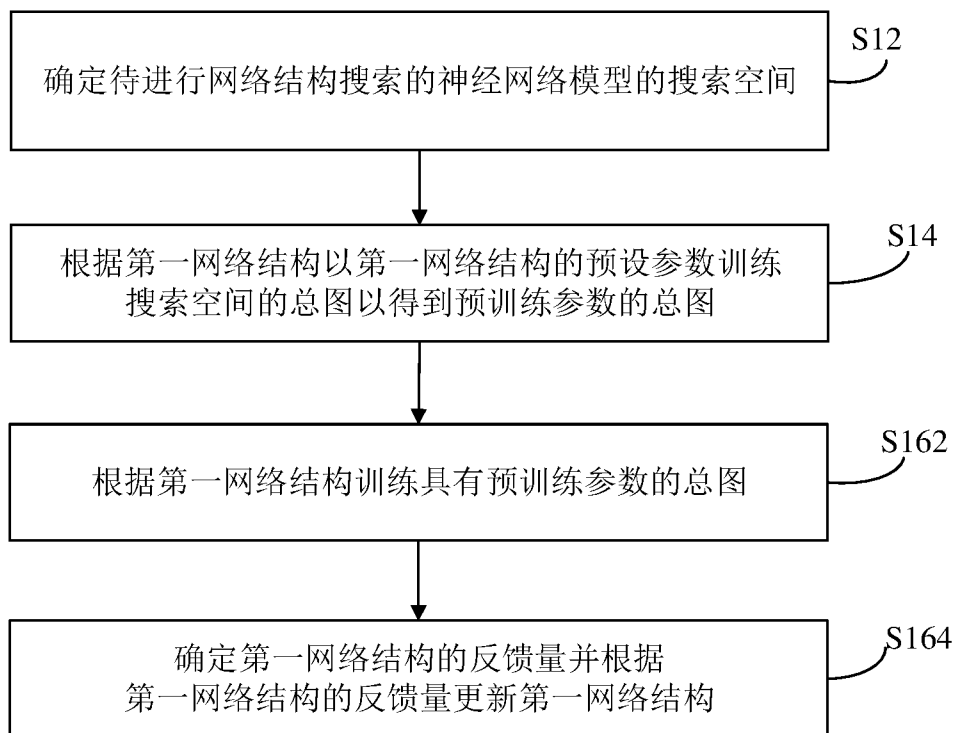


图 6

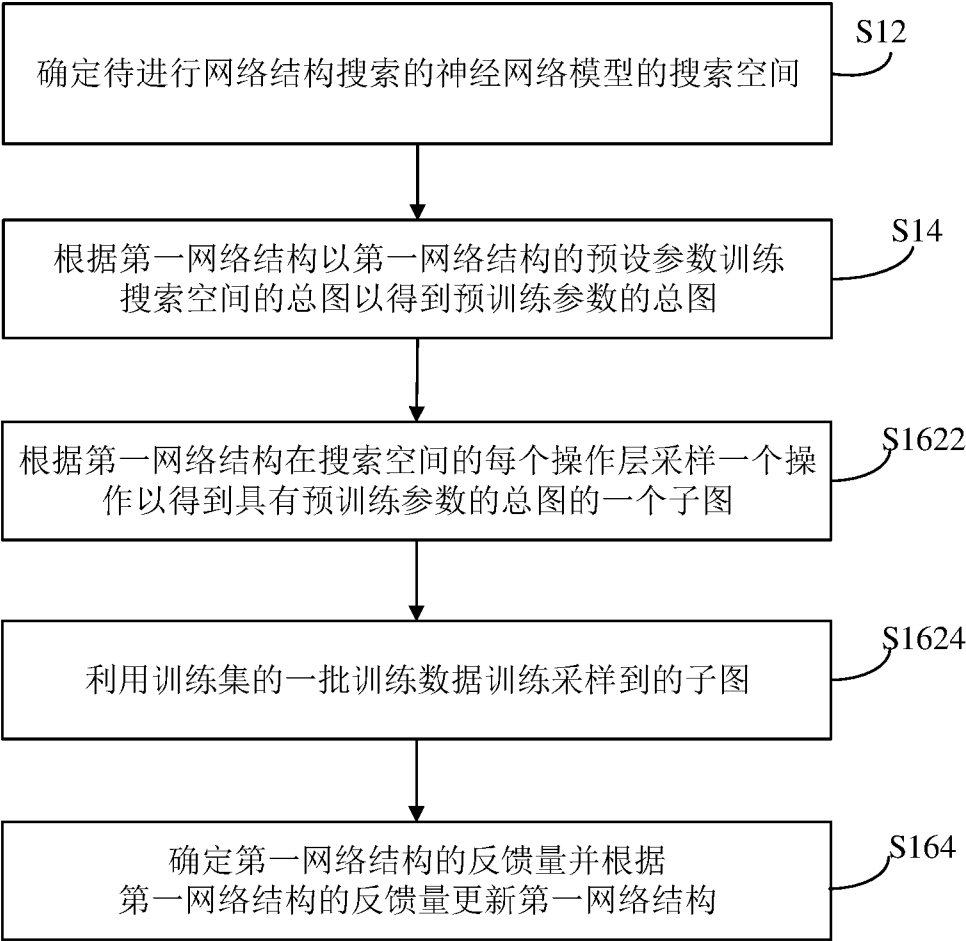


图 7

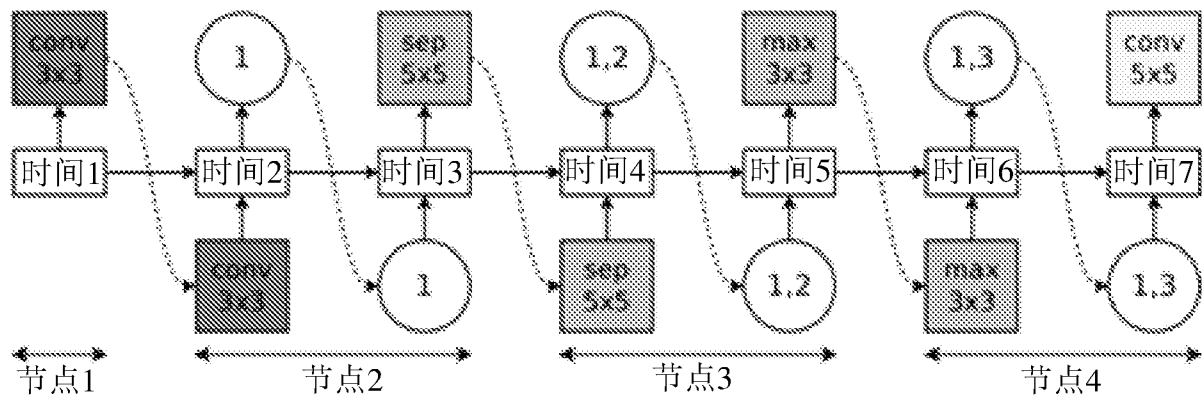


图 8

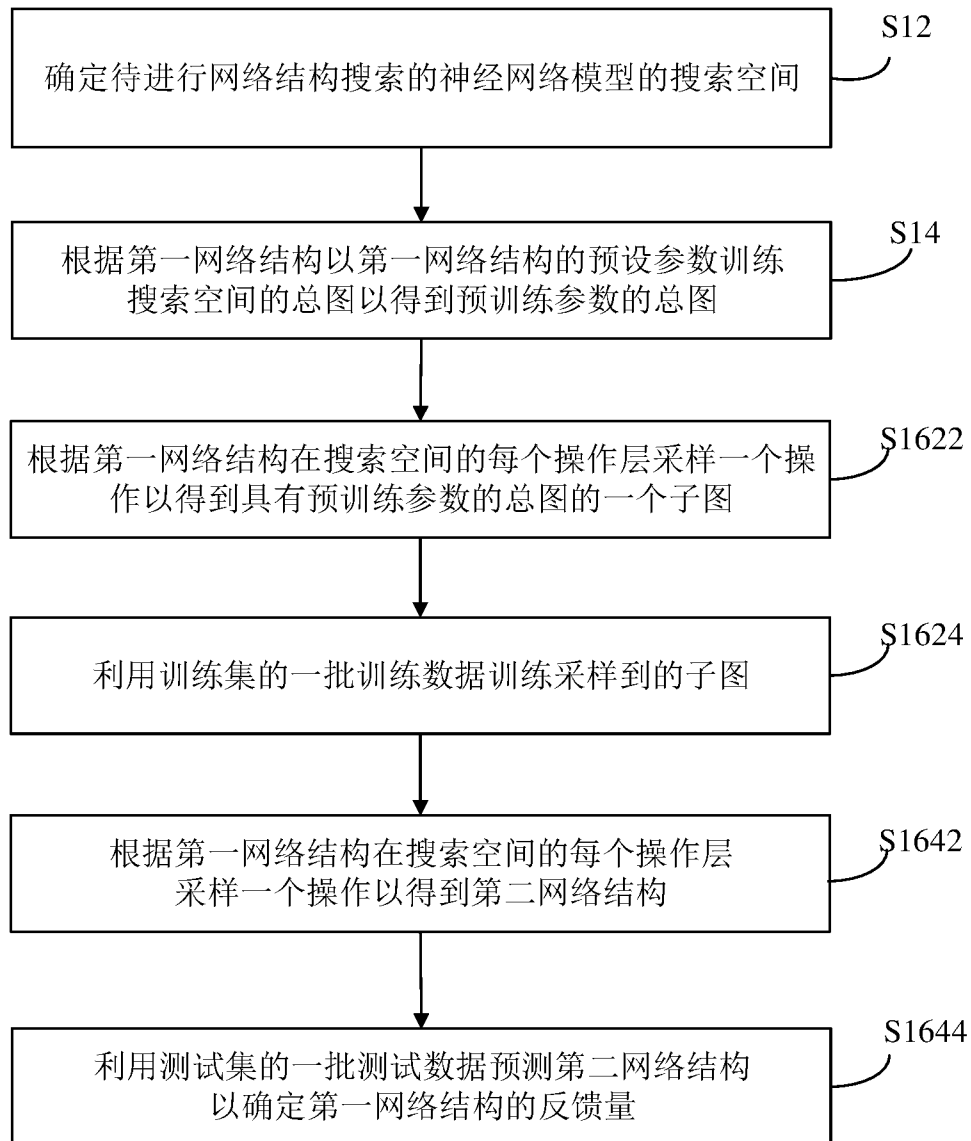


图 9

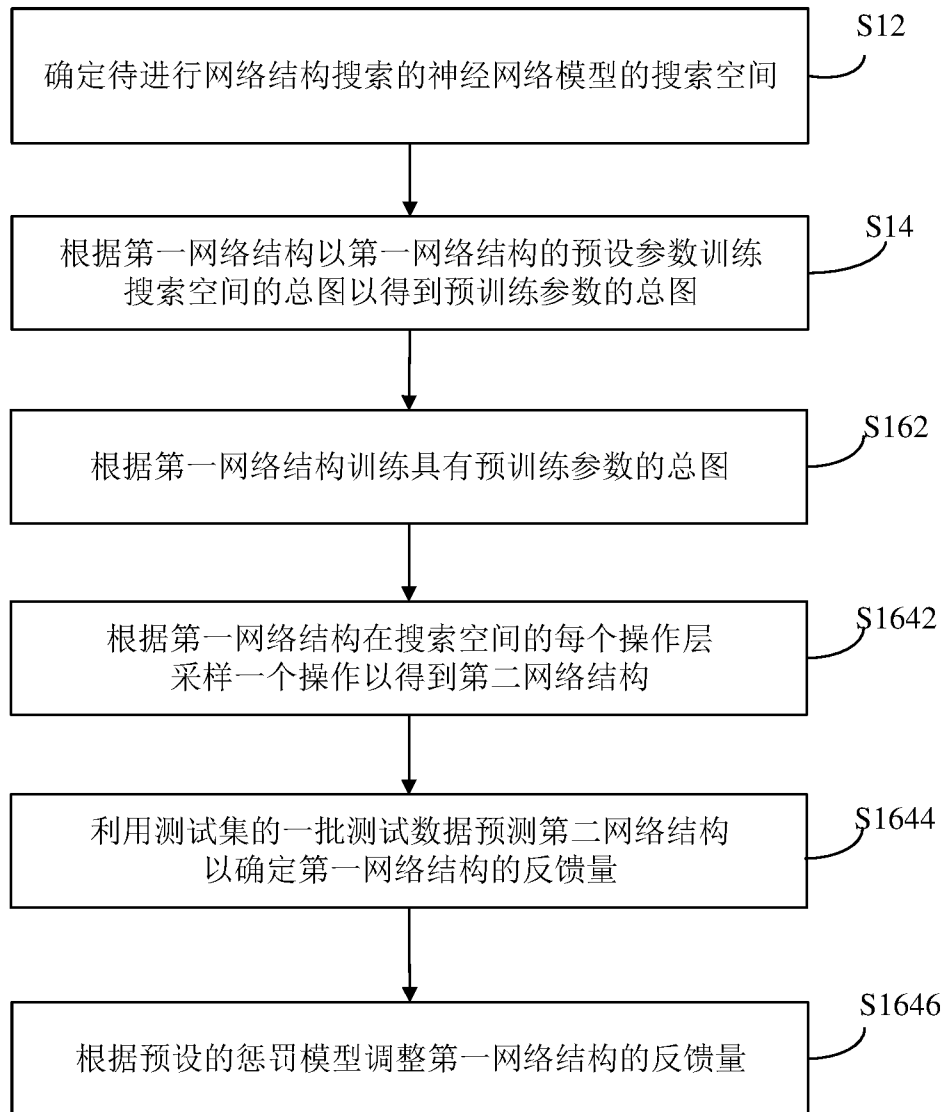


图 10

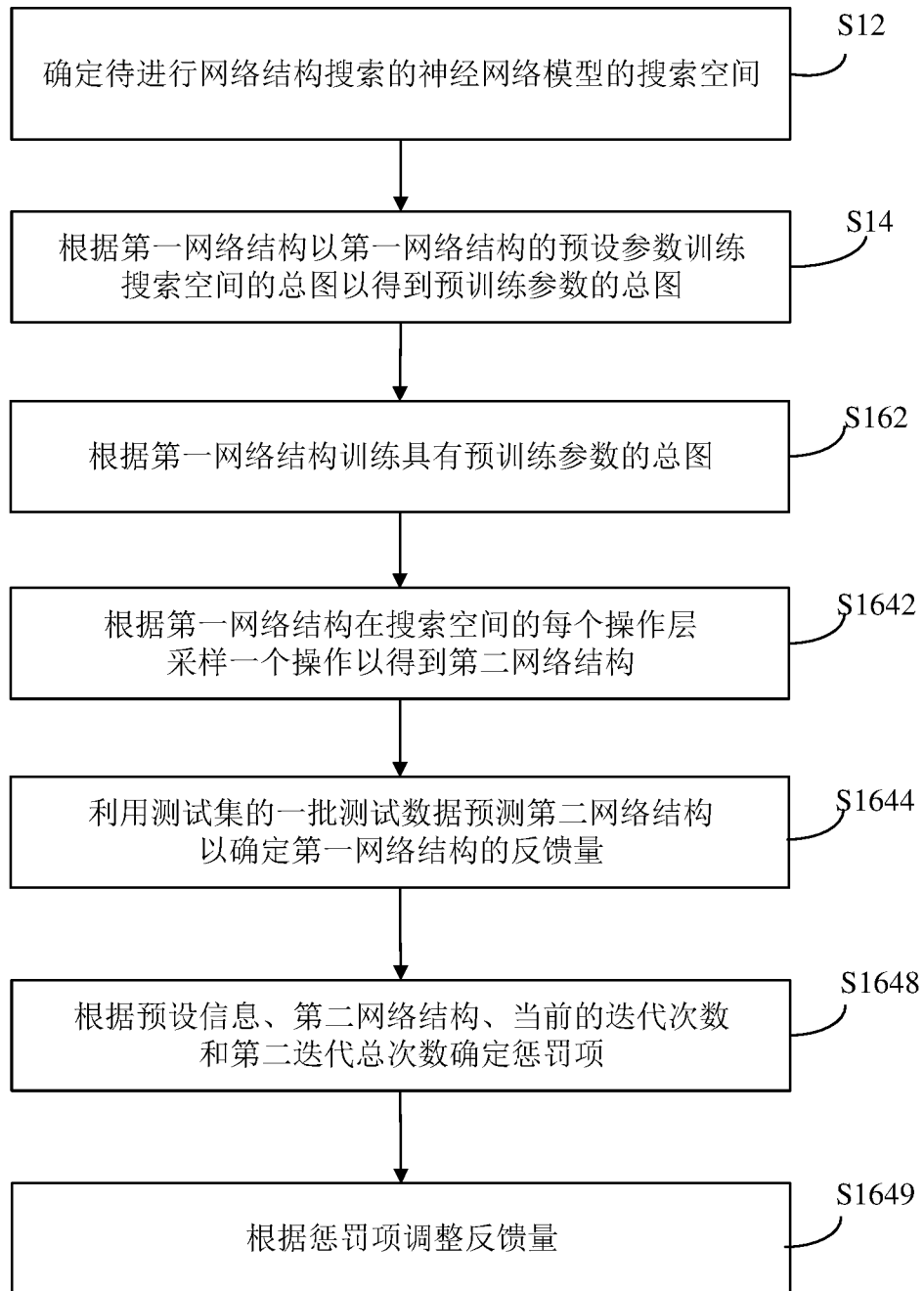


图 11

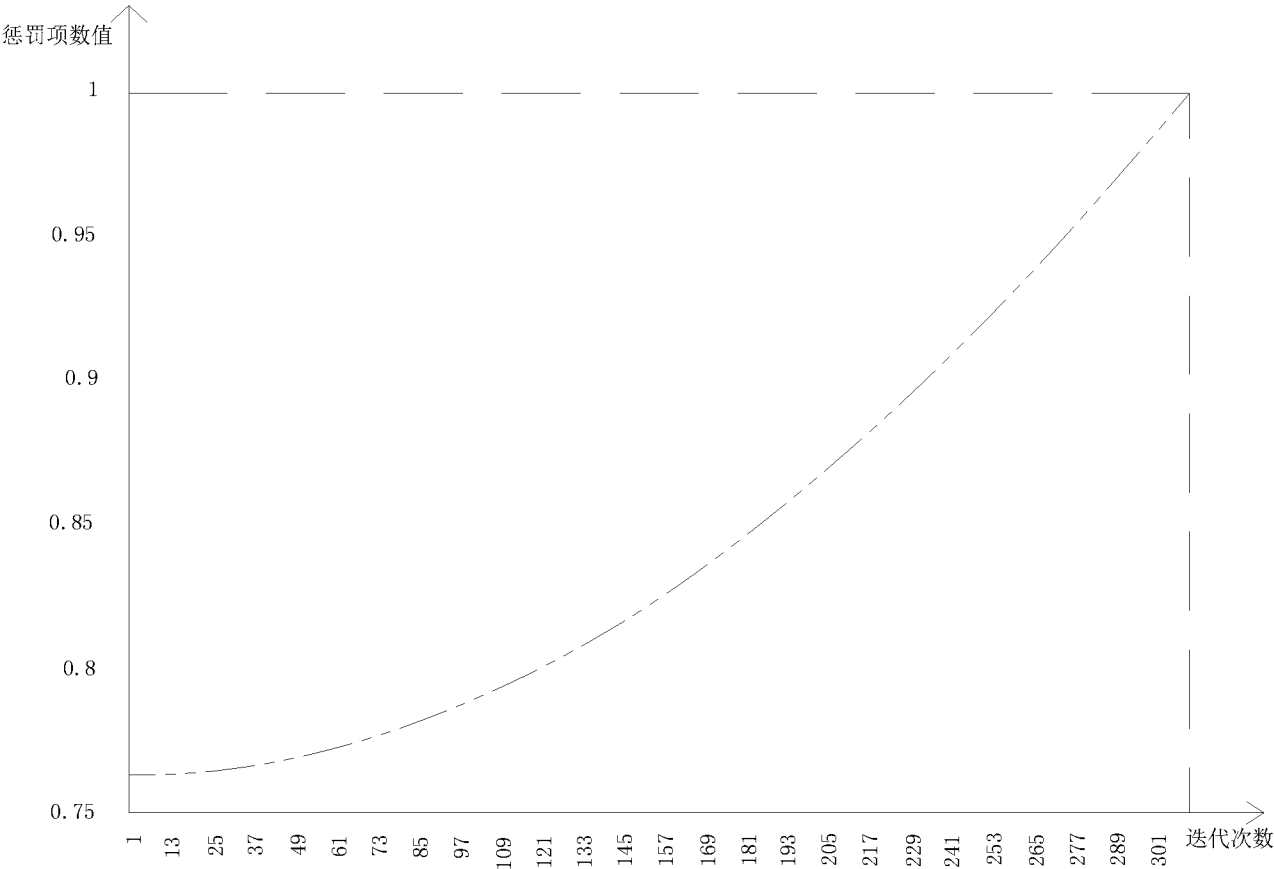


图 12

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/089697

A. CLASSIFICATION OF SUBJECT MATTER

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)

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)

CNPAT, CNKI, WPI, EPODOC, IEEE: 网络结构搜索, 神经网络, 搜索, NAS, ENAS, 预训练, 预设, 预置, 固定, 参数, 训练, 反馈, 偏见, 卷积, 池化, 图, neural architecture search, neural network, search, preset, train+, fixed, parameter, feedback, bias, graph, convolut+, pool+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 109284820 A (BEIJING TUSIMPLE FUTURE TECHNOLOGY CO., LTD.) 29 January 2019 (2019-01-29) claims 7 and 8, and description, paragraphs [0003], [0004], [0040]-[0068], [0109] and [0110]	1-18
A	CN 108229647 A (BEIJING SENSETIME SCIENCE TECHNOLOGY LTD.) 29 June 2018 (2018-06-29) entire document	1-18
A	CN 109299142 A (SUN YAT-SEN UNIVERSITY) 01 February 2019 (2019-02-01) entire document	1-18
A	CN 108921287 A (OPEN AI LAB (SHANGHAI) CO., LTD.) 30 November 2018 (2018-11-30) entire document	1-18
A	CN 108780519 A (MAGIC LEAP INC.) 09 November 2018 (2018-11-09) entire document	1-18
A	WO 2019018375 A1 (GOOGLE LLC.) 24 January 2019 (2019-01-24) entire document	1-18



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

“&” document member of the same patent family

Date of the actual completion of the international search

26 February 2020

Date of mailing of the international search report

03 March 2020

Name and mailing address of the ISA/CN

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

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/089697

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109284820	A	29 January 2019	None			
CN	108229647	A	29 June 2018	US	2019095780	A1	28 March 2019
				WO	2019034129	A1	21 February 2019
CN	109299142	A	01 February 2019	None			
CN	108921287	A	30 November 2018	None			
CN	108780519	A	09 November 2018	AU	2017230184	A1	06 September 2018
				US	10255529	B2	09 April 2019
				US	2019286951	A1	19 September 2019
				US	2017262737	A1	14 September 2017
				KR	20180117704	A	29 October 2018
				IL	261245	D0	31 October 2018
				WO	2017156547	A1	14 September 2017
				CA	3015658	A1	14 September 2017
				JP	2019511777	A	25 April 2019
				EP	3427192	A1	16 January 2019
WO	2019018375	A1	24 January 2019	US	2019026639	A1	24 January 2019
				US	10521729	B2	31 December 2019
				EP	3583553	A1	25 December 2019
				CN	110476172	A	19 November 2019

国际检索报告

国际申请号

PCT/CN2019/089697

A. 主题的分类 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) 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)) CNPAT, CNKI, WPI, EPDOC, IEEE: 网络结构搜索, 神经网络, 搜索, NAS, ENAS, 预训练, 预设, 预置, 固定, 参数, 训练, 反馈, 偏见, 卷积, 池化, 图, neural architecture search, neural network, search, preset, train+, fixed, parameter, feedback, bias, graph, convolut+, pool+																							
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 109284820 A (北京图森未来科技有限公司) 2019年 1月 29日 (2019 - 01 - 29) 权利要求7-8, 说明书第0003-0004、0040-0068、0109-0110段</td> <td>1-18</td> </tr> <tr> <td>A</td> <td>CN 108229647 A (北京市商汤科技开发有限公司) 2018年 6月 29日 (2018 - 06 - 29) 全文</td> <td>1-18</td> </tr> <tr> <td>A</td> <td>CN 109299142 A (中山大学) 2019年 2月 1日 (2019 - 02 - 01) 全文</td> <td>1-18</td> </tr> <tr> <td>A</td> <td>CN 108921287 A (开放智能机器上海有限公司) 2018年 11月 30日 (2018 - 11 - 30) 全文</td> <td>1-18</td> </tr> <tr> <td>A</td> <td>CN 108780519 A (奇跃公司) 2018年 11月 9日 (2018 - 11 - 09) 全文</td> <td>1-18</td> </tr> <tr> <td>A</td> <td>WO 2019018375 A1 (GOOGLE LLC.) 2019年 1月 24日 (2019 - 01 - 24) 全文</td> <td>1-18</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 109284820 A (北京图森未来科技有限公司) 2019年 1月 29日 (2019 - 01 - 29) 权利要求7-8, 说明书第0003-0004、0040-0068、0109-0110段	1-18	A	CN 108229647 A (北京市商汤科技开发有限公司) 2018年 6月 29日 (2018 - 06 - 29) 全文	1-18	A	CN 109299142 A (中山大学) 2019年 2月 1日 (2019 - 02 - 01) 全文	1-18	A	CN 108921287 A (开放智能机器上海有限公司) 2018年 11月 30日 (2018 - 11 - 30) 全文	1-18	A	CN 108780519 A (奇跃公司) 2018年 11月 9日 (2018 - 11 - 09) 全文	1-18	A	WO 2019018375 A1 (GOOGLE LLC.) 2019年 1月 24日 (2019 - 01 - 24) 全文	1-18
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
X	CN 109284820 A (北京图森未来科技有限公司) 2019年 1月 29日 (2019 - 01 - 29) 权利要求7-8, 说明书第0003-0004、0040-0068、0109-0110段	1-18																					
A	CN 108229647 A (北京市商汤科技开发有限公司) 2018年 6月 29日 (2018 - 06 - 29) 全文	1-18																					
A	CN 109299142 A (中山大学) 2019年 2月 1日 (2019 - 02 - 01) 全文	1-18																					
A	CN 108921287 A (开放智能机器上海有限公司) 2018年 11月 30日 (2018 - 11 - 30) 全文	1-18																					
A	CN 108780519 A (奇跃公司) 2018年 11月 9日 (2018 - 11 - 09) 全文	1-18																					
A	WO 2019018375 A1 (GOOGLE LLC.) 2019年 1月 24日 (2019 - 01 - 24) 全文	1-18																					
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																							
<table border="0"> <tr> <td> * 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td> “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																			
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																						
国际检索实际完成的日期 2020年 2月 26日		国际检索报告邮寄日期 2020年 3月 3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 李晶 电话号码 86-10-53961705																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/089697

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	109284820	A	2019年 1月 29日	无			
CN	108229647	A	2018年 6月 29日	US	2019095780	A1	2019年 3月 28日
				WO	2019034129	A1	2019年 2月 21日
CN	109299142	A	2019年 2月 1日	无			
CN	108921287	A	2018年 11月 30日	无			
CN	108780519	A	2018年 11月 9日	AU	2017230184	A1	2018年 9月 6日
				US	10255529	B2	2019年 4月 9日
				US	2019286951	A1	2019年 9月 19日
				US	2017262737	A1	2017年 9月 14日
				KR	20180117704	A	2018年 10月 29日
				IL	261245	D0	2018年 10月 31日
				WO	2017156547	A1	2017年 9月 14日
				CA	3015658	A1	2017年 9月 14日
				JP	2019511777	A	2019年 4月 25日
				EP	3427192	A1	2019年 1月 16日
WO	2019018375	A1	2019年 1月 24日	US	2019026639	A1	2019年 1月 24日
				US	10521729	B2	2019年 12月 31日
				EP	3583553	A1	2019年 12月 25日
				CN	110476172	A	2019年 11月 19日