

(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
2024 年 4 月 4 日 (04.04.2024)



(10) 国际公布号
WO 2024/067280 A1

(51) 国际专利分类号:

H04L 41/082 (2022.01)

SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ,
UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(21) 国际申请号:

PCT/CN2023/119938

(22) 国际申请日:

2023 年 9 月 20 日 (20.09.2023)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

202211177497.5 2022年9月26日 (26.09.2022) CN

(71) 申请人: 维沃移动通信有限公司 (**VIVO MOBILE COMMUNICATION CO., LTD.**) [CN/CN]; 中国广东省东莞市长安镇维沃路 1 号, Guangdong 523863 (CN)。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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))。

(72) 发明人: 杨昂 (**YANG, Ang**); 中国广东省东莞市长安镇维沃路 1 号, Guangdong 523863 (CN)。孙鹏 (**SUN, Peng**); 中国广东省东莞市长安镇维沃路 1 号, Guangdong 523863 (CN)。

(74) 代理人: 北京银龙知识产权代理有限公司 (**DRAGON INTELLECTUAL PROPERTY LAW FIRM**); 中国北京市海淀区西直门北大街 32 号院枫蓝国际中心 2 号楼 10 层, Beijing 100082 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,

(54) **Title:** METHOD AND APPARATUS FOR UPDATING AI MODEL PARAMETER, AND COMMUNICATION DEVICE

(54) 发明名称: 更新 AI 模型参数的方法、装置及通信设备

第一节点向第二节点发送与 AI 模型参数相关的第一信息;
其中, 所述第一信息包括以下至少一项: 所述 AI 模型参数的更新模式, 所述 AI 模型参数的指示方式

图 4

401

401 A first node sends, to a second node, first information related to an AI model parameter, wherein the first information comprises at least one of the following: an update mode for the AI model parameter, and an indication mode for the AI model parameter

(57) **Abstract:** Disclosed in the present application are a method and apparatus for updating an AI model parameter, and a communication device. The method comprises: a first node sending, to a second node, first information related to an AI model parameter, wherein the first information comprises at least one of the following: an update mode for the AI model parameter, and an indication mode for the AI model parameter.

(57) 摘要: 本申请公开了一种更新 AI 模型参数的方法、装置及通信设备, 该方法包括: 第一节点向第二节点发送与 AI 模型参数相关的第一信息; 其中, 所述第一信息包括以下至少一项: 所述 AI 模型参数的更新模式, 所述 AI 模型参数的指示方式。

WO 2024/067280 A1

更新 AI 模型参数的方法、装置及通信设备

相关申请的交叉引用

本申请主张在 2022 年 09 月 26 日在中国提交的中国专利申请 No. 202211177497.5 的优先权，其全部内容通过引用包含于此。

技术领域

本申请属于通信技术领域，具体涉及一种更新 AI 模型参数的方法、装置及通信设备。

背景技术

人工智能（Artificial Intelligence，AI）在各个领域获得了广泛的应用。AI 模型有多种实现方式，例如神经网络、决策树、支持向量机、贝叶斯分类器等。

相关技术中通过传输 AI 模型进行模型更新，然而这种更新方式效率较低以及网络信令开销较大。

发明内容

本申请实施例提供一种更新 AI 模型参数的方法、装置及通信设备，解决相关技术中 AI 模型的更新方式效率较低以及网络信令开销较大的问题。

第一方面，提供一种更新 AI 模型参数的方法，包括：

第一节点向第二节点发送与 AI 模型参数相关的第一信息；

其中，所述第一信息包括以下至少一项：所述 AI 模型参数的更新模式，所述 AI 模型参数的更新参数的指示方式。

第二方面，提供一种更新 AI 模型参数的方法，包括：

第二节点接收第一节点发送的与 AI 模型参数相关的第一信息；

所述第二节点根据所述第一信息更新所述 AI 模型参数；

其中，所述第一信息包括以下至少一项：所述 AI 模型参数的更新模式，所述 AI 模型参数的更新参数的指示方式。

第三方面，提供一种更新 AI 模型参数的装置，包括：

第一发送模块，用于向第二节点发送与 AI 模型参数相关的第一信息；

其中，所述第一信息包括以下至少一项：所述 AI 模型参数的更新模式，所述 AI 模型参数的更新参数的指示方式。

第四方面，提供一种更新 AI 模型参数的装置，包括：

第一接收模块，用于接收第一节点发送的与 AI 模型参数相关的第一信息；

更新模块，用于根据所述第一信息更新所述 AI 模型参数；

其中，所述第一信息包括以下至少一项：所述 AI 模型参数的更新模式，所述 AI 模型参数的更新参数的指示方式。

第五方面，提供了一种通信设备，包括：处理器，存储器及存储在所述存储器上并可在所述处理器上运行的程序或指令，所述程序或指令被所述处理器执行时实现如第一方面或第二方面所述的方法的步骤。

第六方面，提供了一种可读存储介质，所述可读存储介质上存储程序或指令，所述程序或指令被处理器执行时实现如第一方面或第二方面所述的方法的步骤。

第七方面，提供了一种芯片，所述芯片包括处理器和通信接口，所述通信接口和所述处理器耦合，所述处理器用于运行程序或指令，实现如第一方面或第二方面所述的方法的步骤。

第八方面，提供了一种计算机程序/程序产品，所述计算机程序/程序产品被存储在非瞬态的存储介质中，所述程序/程序产品被至少一个处理器执行以实现如第一方面或第二方面所述的方法的步骤。

第九方面，提供一种通信系统，所述通信系统包括终端与网络侧设备，所述终端用于执行如第一方面所述的方法的步骤，所述网络侧设备用于执行如第二方面所述的方法的步骤。

在本申请实施例中，第一节点可以向第二节点发送所述 AI 模型参数的更新模式和/或 AI 模型参数的更新参数的指示方式，第二节点不需要对 AI 模型进行编译或重编译，第二节点根据接收到的 AI 模型参数的更新模式和/或 AI 模型参数的更新参数的指示方式更新已有 AI 模型参数，能够有效提升无线通信系统中传输 AI 模型的效率，以及减少网络信令开销。

附图说明

图 1 为神经网络的示意图；

图 2 为神经元的示意图；

图 3 为本申请实施例的无线通信系统的架构示意图；

图 4 为本申请实施例的更新 AI 模型参数的方法的流程图之一；

图 5 为本申请实施例的更新 AI 模型参数的方法的流程图之二；

图 6 为本申请实施例的 AI 模型参数的指示方式的示意图之一；

图 7 为本申请实施例的 AI 模型参数的指示方式的示意图之二；

图 8 为本申请实施例的 AI 模型参数的指示方式的示意图之三；

图 9 为本申请实施例的 AI 模型参数的指示方式的示意图之四；

图 10 为本申请实施例的更新 AI 模型参数的装置的结构图之一；

图 11 为本申请实施例的更新 AI 模型参数的装置的结构图之二；

图 12 为本申请实施例的终端的示意图；

图 13 为本申请实施例的网络侧设备的示意图；

图 14 为本申请实施例的通信设备的示意图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员所获得的所有其他实施例，都属于本申请保护的范围。

本申请的说明书和权利要求书中的术语“第一”、“第二”等是用于区别类似的对象，而不适用于描述特定的顺序或先后次序。应该理解这样使用的术语在适当情况下可以互换，以便本申请的实施例能够以除了在这里图示或描述的那些以外的顺序实施，且“第一”、“第二”所区别的对象通常为一类，并不限定对象的个数，例如第一对象可以是一个，也可以是多个。此外，说明书以及权利要求中“和/或”表示所连接对象的至少其中之一，字符“/”一般表示前后关联对象是一种“或”的关系。

值得指出的是，本申请实施例所描述的技术不限于长期演进型(Long Term Evolution, LTE)/LTE 的演进(LTE-Advanced, LTE-A)系统，还可用于其他无线通信系统，诸如码分多址(Code Division Multiple Access, CDMA)、时分多址(Time Division Multiple Access, TDMA)、频分多址(Frequency Division Multiple Access, FDMA)、正交频分多址(Orthogonal Frequency Division Multiple Access, OFDMA)、单载波频分多址(Single-carrier Frequency Division Multiple Access, SC-FDMA)和其他系统。本申请实施例中的术语“系统”和“网络”常被可互换地使用，所描述的技术既可用于以上提及的系统 and 无线电技术，也可用于其他系统和无线电技术。以下描述出于示例目的描述了新空口(New Radio, NR)系统，并且在以下大部分描述中使用 NR 术语，但是这些技术也可应用于 NR 系统应用以外的应用，如第 6 代(6th Generation, 6G)通信系统。

为了便于理解本申请的实施方式，下面先介绍以下技术点。

1、关于神经网络的介绍

本申请以神经网络为例进行说明，但是并不限定 AI 模块的具体类型，神经网络的结构如图 1 所示。

其中，神经网络由神经元组成，神经元的示意图如图 2 所示。其中 $a_1, a_2, \dots, a_k$ 为输入， w 为权值(乘性系数)， b 为偏置(加性系数)， $\sigma(\cdot)$ 为激活函数， $z = a_1 * w_1 + \dots + a_k * w_k + \dots + a_k * w_k + b$ 。常见的激活函数包括 Sigmoid 函数、tanh 函数、修正线性单元(Rectified Linear Unit, ReLU) 等等。

神经网络的参数可以通过优化算法进行优化。优化算法就是一种能够最小化或者最大化目标函数(有时候也叫损失函数)的一类算法。而目标函数往往是模型参数和数据的数学组合。例如给定数据 X 和其对应的标签 Y ，构建一个神经网络模型 $f(\cdot)$ ，有了模型后，

根据输入 x 就可以得到预测输出 $f(x)$ ，并且可以计算出预测值和真实值之间的差距 $(f(x)-Y)$ ，这个就是损失函数。如果找到合适的 W, b 使上述的损失函数的值达到最小，损失值越小，则说明模型越接近于真实情况。

目前常见的优化算法，基本都是基于误差反向传播（error Back Propagation, BP）算法。BP 算法的基本思想是，学习过程由信号的正向传播与误差的反向传播两个过程组成。正向传播时，输入样本从输入层传入，经各隐藏层逐层处理后，传向输出层。若输出层的实际输出与期望的输出不符，则转入误差的反向传播阶段。误差反传是将输出误差以某种形式通过隐藏层向输入层逐层反传，并将误差分摊给各层的所有单元，从而获得各层单元的误差信号，此误差信号即作为修正各单元权值的依据。这种信号正向传播与误差反向传播的各层权值调整过程，是周而复始地进行的。权值不断调整的过程，也就是网络的学习训练过程。此过程一直进行到网络输出的误差减少到可接受的程度，或进行到预先设定的学习次数为止。

一般而言，根据解决类型不同，选取的 AI 算法和采用的模型也有所差别。根据相关技术，借助 AI 提升第五代移动通信技术（5th Generation, 5G）网络性能的主要方法是通过基于神经网络的算法和模型增强或者替代目前已有的算法或处理模块。在特定场景下，基于神经网络的算法和模型可以取得比基于确定性算法更好的性能。比较常用的神经网络包括深度神经网络、卷积神经网络和循环神经网络等。借助已有 AI 工具，可以实现神经网络的搭建、训练与验证工作。

常见的优化算法有梯度下降（Gradient Descent）、随机梯度下降（Stochastic Gradient Descent, SGD）、小批量梯度下降（mini-batch gradient descent）、动量法（Momentum）、内斯特洛夫（Nesterov）加速算法、自适应梯度下降（ADaptive GRADient descent, Adagrad）、Adadelta、均方根误差降速（Root Mean Square prop, RMSprop）、自适应动量估计（Adaptive Moment Estimation, Adam）等。

这些优化算法在误差反向传播时，都是根据损失函数得到的误差/损失，对当前神经元求导数/偏导，加上学习速率、之前的梯度/导数/偏导等影响，得到梯度，将梯度传给上一层。图 3 示出本申请实施例可应用的一种无线通信系统的框图。无线通信系统包括终端 31 和网络侧设备 32。

其中，终端 31 可以是手机、平板电脑（Tablet Personal Computer）、膝上型电脑（Laptop Computer）或称为笔记本电脑、个人数字助理（Personal Digital Assistant, PDA）、掌上电脑、上网本、超级移动个人计算机（ultra-mobile personal computer, UMPC）、移动上网装置（Mobile Internet Device, MID）、增强现实（augmented reality, AR）/虚拟现实（virtual reality, VR）设备、机器人、可穿戴式设备（Wearable Device）、车载设备（Vehicle User Equipment, VUE）、行人终端（Pedestrian User Equipment, PUE）、智能家居（具有无线通信功能的家居设备，如冰箱、电视、洗衣机或者家具等）、游戏机、个人计算机（personal computer, PC）、柜员机或者自助机等终端侧设备，可穿戴式设备包括：智能手表、智能手环、智能

耳机、智能眼镜、智能首饰（智能手镯、智能手链、智能戒指、智能项链、智能脚镯、智能脚链等）、智能腕带、智能服装等。除了上述终端设备，本申请涉及的终端也可以是终端内的芯片，例如调制解调器（Modem）芯片，系统级芯片（System on Chip, SoC）。需要说明的是，在本申请实施例并不限定终端 31 的具体类型。

网络侧设备 32 可以包括接入网设备或核心网设备，其中，接入网设备也可以称为无线接入网设备、无线接入网（Radio Access Network, RAN）、无线接入网功能或无线接入网单元。接入网设备可以包括基站、无线局域网（Wireless Local Area Network, WLAN）接入点或 WiFi 节点等，基站可被称为节点 B、演进节点 B（eNB）、接入点、基收发机站（Base Transceiver Station, BTS）、无线电基站、无线电收发机、基本服务集（Basic Service Set, BSS）、扩展服务集（Extended Service Set, ESS）、家用 B 节点、家用演进型 B 节点、发送接收点（Transmitting Receiving Point, TRP）或所述领域中其他某个合适的术语，只要达到相同的技术效果，所述基站不限于特定技术词汇，需要说明的是，在本申请实施例中仅以 NR 系统中的基站为例进行介绍，并不限定基站的具体类型。

核心网设备可以包含但不限于如下至少一项：核心网节点、核心网功能、移动管理实体（Mobility Management Entity, MME）、接入和移动管理功能（Access and Mobility Management Function, AMF）、会话管理功能（Session Management Function, SMF）、用户平面功能（User Plane Function, UPF）、策略控制功能（Policy Control Function, PCF）、策略与计费规则功能单元（Policy and Charging Rules Function, PCRF）、边缘应用服务发现功能（Edge Application Server Discovery Function, EASDF）、统一数据管理（Unified Data Management, UDM）、统一数据仓储（Unified Data Repository, UDR）、归属用户服务器（Home Subscriber Server, HSS）、集中式网络配置（Centralized network configuration, CNC）、网络存储功能（Network Repository Function, NRF）、网络开放功能（Network Exposure Function, NEF）、本地 NEF（Local NEF, 或 L-NEF）、绑定支持功能（Binding Support Function, BSF）、应用功能（Application Function, AF）、非 3GPP 互通功能（Non-3GPP InterWorking Function, N3IWF）等。需要说明的是，在本申请实施例中仅以 NR 系统中的核心网设备为例进行介绍，并不限定核心网设备的具体类型。

下面结合附图，通过一些实施例及其应用场景对本申请实施例提供的更新 AI 模型参数的方法、装置、通信设备及可读存储介质进行详细地说明。

参见图 4，本申请实施例提供一种更新 AI 模型参数的方法，应用于第一节点，该第一节点可以是网络侧设备或终端，具体步骤包括：步骤 401。

步骤 401：第一节点向第二节点发送与 AI 模型参数相关的第一信息；

其中，所述第一信息包括以下至少一项：所述 AI 模型参数的更新模式，所述 AI 模型参数的指示方式。

在本实施例中，所述第一信息用于第二节点更新 AI 模型参数，第一信息中的 AI 模型参数的更新模式用于表示采用哪种方式更新 AI 模型参数，第一信息中的 AI 模型参数的指

示方式用于指示更新参数的位置或者用于指示 AI 模型参数的发送顺序，比如第一节点可以向第二节点指示所有 AI 模型参数的发送顺序，第二节点根据情况对接收到的所有 AI 模型参数中的部分参数进行 AI 模型更新，或者第一节点可以向第二节点指示部分 AI 模型参数的发送顺序，第二节点根据接收到的部分 AI 模型参数进行 AI 模型更新，其中的部分 AI 模型参数也可以称为 AI 模型更新参数。

在本实施例中，第一节点可以向第二节点发送所述 AI 模型参数的更新模式和/或 AI 模型参数的指示方式，第二节点不需要对 AI 模型进行编译或重编译，第二节点根据接收到的 AI 模型参数的更新模式和/或 AI 模型参数的指示方式更新已有 AI 模型参数，能够有效提升无线通信系统中传输 AI 模型的效率，以及减少网络信令开销。

在本申请的一种实施例中，所述第一节点为第一网络侧设备或第一终端，所述第二节点为第二网络侧设备或第二终端，比如，第一节点为第一网络侧设备，第二节点为第二终端，或者，第一节点为第一终端，第二节点为第二网络侧设备，或者，第一节点为第一网络侧设备，第二节点为第二网络侧设备，或者第一节点为第一终端，第二节点为第二终端。

在本申请的一种实施例中，所述更新模式包括以下之一：

(1) 第一模式；

可选的，所述第一模式包括：更新第一参数子集中的所有参数，所述第一参数子集是所述 AI 模型参数的子集，所述第一参数子集是预配置或者协议约定的或者网络侧指示的；

(2) 第二模式；

可选的，所述第二模式包括：更新第二参数子集中的所有参数，所述第二参数子集是所述第一参数子集的子集；

(3) 第三模式；

可选的，所述第三模式包括：更新所述 AI 模型的所有参数；

(4) 第四模式；

可选的，所述第四模式包括：更新第三参数子集，所述第三参数子集是所述 AI 模型参数的子集。

在本申请的一种实施例中，所述指示方式包括以下之一：

(1) 第一信息，所述第一信息用于指示更新参数在所述 AI 模型中的位置；

(2) 第二信息，所述第二信息用于指示所述 AI 模型中更新参数的发送顺序。

示例性的，本申请实施例适用的一种神经网络的结构如图 1 所示，一层对应一列，即输入层为第一列，输出层为最后一列，隐藏层为中间列，当然本申请实施例也可以适用其他结构的神经网络。

在一种实施方式中，可以基于图 1 所示的神经网络的结构“指示更新参数在所述 AI 模型中的位置”，以及“指示所述 AI 模型中更新参数的发送顺序”。

在另一种实施方式中，神经网络的结构可以是根据图 1 旋转得到的神经网络的另一种结构，这样基于该神经网络的另一种结构“指示更新参数在所述 AI 模型中的位置”，以及

“指示所述 AI 模型中更新参数的发送顺序”，指示方式应用在图 1 所示的神经网络的结构，与应用在神经网络的其他结构相似，在此不再赘述。

比如，神经网络的另一种结构可以是相对于图 1 的示意向右旋转 90 度，即一层对应一行，输入层为第一行，输出层为最后一行，隐藏层为中间行；又比如，神经网络的另一种结构可以是相对于图 1 的示意的神经网络结构向左旋转 90 度，同样一层对应一行，输入层为最后一行，输出层为第一行，隐藏层为中间行，可以理解的是神经网络的结构并不限于以上三种形式。

在本申请的一种实施例中，所述 AI 模型为神经网络，所述神经网络包括输入层、一个或多个隐藏层和输出层，所述输入层、一个或多个隐藏层、输出层，以及各层的神经元组成多行多列的结构，所述神经网络中神经元的参数的发送顺序是根据所述神经元在所述神经网络中的位置确定的。

在本申请的一种实施例中，所述神经网络中神经元的参数的发送顺序，包括以下之一：

(1) 第一列所有神经元的参数，第二列所有神经元的参数，以此类推，直至最后一列所有神经元的参数；

在一种实施方式中，所述第一列对应所述输入层，所述最后一列对应所述输出层；或者，所述第一列对应所述输出层，所述最后一列对应所述输入层。

(2) 第一行所有神经元的参数，第二行所有神经元的参数，以此类推，直至最后一行所有神经元的参数。

在一种实施方式中，所述第一行对应所述输入层，所述最后一行对应所述输出层；或者，所述第一行对应所述输出层，所述最后一行对应所述输入层；

在一种实施方式中，所述第一行是所述神经网络中最上方一行，所述最后一行是所述神经网络中最下方一行；或者，所述第一行是所述神经网络中最下方一行，所述最后一行是所述神经网络中最上方一行。

在本申请的一种实施例中，在所述神经网络中神经元的参数的发送顺序包括：第一列所有神经元的参数，第二列所有神经元的参数，以此类推，直至最后一列所有神经元的参数的情况下，所述神经网络中神经元的参数的发送顺序还包括以下之一：

(1) 每一列中神经元的参数按照从上往下的神经元的顺序发送；

(2) 每一列中神经元的参数按照从下往上的神经元的顺序发送；

(3) 每一列中神经元的参数的发送顺序相同，如图 6 所示；

(4) 相邻列的神经元的参数的发送顺序相反，如图 7 所示。

在本申请的一种实施例中，在所述神经网络中神经元的参数的发送顺序包括第一行所有神经元的参数，第二行所有神经元的参数，以此类推，直至最后一行所有神经元的参数的情况下，所述神经网络中神经元的参数的发送顺序还包括以下之一：

(1) 每一行中神经元的参数按照从左往右的神经元的顺序发送；

(2) 每一行中神经元的参数按照从右往左的神经元的顺序发送；

(3) 每一行中神经元的参数的发送顺序相同，如图 8 所示；

(4) 相邻行的神经元的参数的发送顺序相反，如图 9 所示。

在本申请的一种实施例中，当每层的神经元数目不是完全相同时，通过以下至少一项发送神经元的参数：

(1) 每一行都按最多神经元数量的行发送所述神经元的参数，缺少神经元的层补预设值（比如，0）。

比如，第一层 50 个神经元，第二层 40 个神经元，第三层 60 个神经元，第四层 30 个神经元，如第 50 行，第二层、第四层没有神经元，则发送时第 50 行的第二层、第四层补 0。

(2) 每一行都按实际神经元数量发送所述神经元的参数。

比如，第一层 50 个神经元，第二层 40 个神经元，第三层 60 个神经元，第四层 30 个神经元，如第 50 行，第二层、第四层没有神经元，则发送时跳过第二层和第四层，第 50 行发送第一层和第三层的神经元的参数。

在本申请的一种实施例中，所述 AI 模型参数的发送顺序是预配置的或协议约定的或者网络侧指示的。

在本申请的一种实施例中，所述 AI 模型参数包括乘性系数和加性系数，所述乘性系数的发送优先级高于加性系数的发送优先级；或者，所述加性系数的发送优先级高于乘性系数的发送优先级。

在本申请的一种实施例中，所述 AI 模型的量化等级和/或压缩方式是预配置的或者协议约定的或者网络侧指示的。

在本申请的一种实施例中，更新后的 AI 模型的模型标识(Identity, ID)与更新前的 AI 模型的模型 ID 相同，不保存所述更新前的 AI 模型；

或者，

更新后的 AI 模型的模型 ID 与更新前的 AI 模型的模型 ID 不同，保存所述更新前的 AI 模型，且所述更新前的 AI 模型的模型 ID 不变。

在本申请的一种实施方式中，所述 AI 模型包括第一功能模块，所述第一功能模块用于以下至少一项：

(1) 信号处理，包括但不限于以下至少一项：信号检测、滤波、均衡等，其中信号包括但不限于以下至少一项：解调参考信号（Demodulation Reference Signal, DMRS）、探测参考信号（Sounding Reference Signal, SRS）、同步信号块（Synchronization Signal and PBCH block, SSB）、跟踪参考信号（Tracking Reference Signal, TRS）、相位跟踪参考信号（Phase-tracking reference signals, PTRS）、信道状态信息参考信号（Channel-State-Information Reference Signal, CSI-RS）等；

(2) 信道信号传输、信道信号接收、信道解调或信道信号发送，其中信道包括但不限于以下至少一项：物理下行控制信道（Physical Downlink Control Channel, PDCCH）、

物理下行共享信道 (Physical Downlink Shared Channel, PDSCH)、物理上行控制信道 (Physical Uplink Control Channel, PUCCH)、物理上行共享信道 (Physical Uplink Shared Channel, PUSCH)、物理随机接入信道 (Physical Random Access Channel, PRACH)、物理广播信道 (Physical Broadcast Channel, PBCH);

(3) 获取处理信道状态信息;

比如, 信道状态信息反馈, 包括但不限于以下至少一项: 信道相关信息、信道矩阵相关信息、信道特征信息、信道矩阵特征信息、预编码矩阵指示 (Precoding matrix indicator, PMI)、秩指示 (Rank indicator, RI)、CSI-RS 资源指示 (CSI-RS Resource Indicator, CRI)、信道质量指示 (Channel quality indicator, CQI)、层指示 (Layer Indicator, LI) 等。

又比如, 频分双工 (Frequency Division Duplex, FDD) 上下行部分互易性。对于 FDD 系统, 根据部分互易性, 基站根据上行信道获取角度和时延信息, 可以通过 CSI-RS 预编码或者直接指示的方法, 将角度信息和时延信息通知终端, 终端根据基站的指示上报或者在基站的指示范围内选择并上报, 从而减少终端的计算量和 CSI 上报的开销。

(4) 波束管理, 包括但不限于以下至少一项: 波束测量、波束上报、波束预测、波束失败检测、波束失败恢复、波束失败恢复中的新波束指示;

(5) 信道预测, 包括但不限于以下至少一项: 信道状态信息的预测、波束预测;

(6) 干扰抑制, 包括但不限于以下至少一项: 小区内干扰、小区间干扰、带外干扰、交调干扰等;

(7) 定位, 比如通过参考信号 (例如 SRS), 估计出终端的具体位置 (包括水平位置和/或垂直位置) 或未来可能的轨迹, 或估计出终端的辅助位置估计或轨迹估计的信息;

(8) 预测或管理高层业务和/或高层参数, 包括但不限于以下至少一项: 吞吐量、所需数据包大小、业务需求、移动速度、噪声信息等;

(9) 解析控制信令, 包括但不限于以下至少一项: 功率控制的相关信令, 波束管理的相关信令。

在本申请实施例中, 能够有效提升无线通信系统中传输 AI 模型的效率, 以及减少网络信令开销。

参见图 5, 本申请实施例提供一种更新 AI 模型参数的方法, 应用于第二节点, 该第二节点为网络侧设备或终端, 该方法包括: 步骤 501 和步骤 502。

步骤 501: 第二节点接收第一节点发送的与 AI 模型参数相关的第一信息;

步骤 502: 第二节点根据所述第一信息更新所述 AI 模型参数;

其中, 所述第一信息包括以下至少一项: 所述 AI 模型参数的更新模式, 所述 AI 模型参数的指示方式。

在本申请的一种实施例中, 所述第一节点为第一网络侧设备或第一终端, 所述第二节点为第二网络侧设备或第二终端, 比如, 第一节点为第一网络侧设备, 第二节点为第二终

端,或者,第一节点为第一终端,第二节点为第二网络侧设备,或者,第一节点为第一网络侧设备,第二节点为第二网络侧设备,或者第一节点为第一终端,第二节点为第二终端。

在本申请的一种实施例中,所述更新模式包括以下之一:

(1) 第一模式;

可选的,所述第一模式包括:更新第一参数子集中的所有参数,所述第一参数子集是所述 AI 模型参数的子集,所述第一参数子集是预配置或者协议约定的或者网络侧指示的;

(2) 第二模式;

可选的,所述第二模式包括:更新第二参数子集中的所有参数,所述第二参数子集是所述第一参数子集的子集;

(3) 第三模式;

可选的,所述第三模式包括:更新所述 AI 模型的所有参数;

(4) 第四模式;

可选的,所述第四模式包括:更新第三参数子集,所述第三参数子集是所述 AI 模型参数的子集。

在本申请的一种实施例中,所述指示方式包括以下之一:

(1) 第一信息,所述第一信息用于指示所述 AI 模型参数在 AI 模型中的位置;

(2) 第二信息,所述第二信息用于指示所述 AI 模型中更新参数的发送顺序。

在本申请的一种实施例中,所述 AI 模型为神经网络,所述神经网络包括输入层、一个或多个隐藏层和输出层,所述输入层、一个或多个隐藏层、输出层,以及各层的神经元组成多行多列的结构,所述神经网络中各个神经元的参数的发送顺序是根据所述神经元在所述神经网络中的位置确定的。

在本申请的一种实施例中,所述神经网络中各个神经元的参数的发送顺序,包括:

第一列所有神经元的参数,第二列所有神经元的参数,以此类推,直至最后一列所有神经元的参数;

或者,

第一行所有神经元的参数,第二行所有神经元的参数,以此类推,直至最后一行所有神经元的参数。

在本申请的一种实施例中,所述第一列对应所述输入层,所述最后一列对应所述输出层;或者,所述第一列对应所述输出层,所述最后一列对应所述输入层;

或者,

所述第一行对应所述输入层,所述最后一行对应所述输出层;或者,所述第一行对应所述输出层,所述最后一行对应所述输入层。

在本申请的一种实施例中,每一列中神经元的参数按照从上往下的神经元的顺序发送,或者每一列中神经元的参数按照从下往上的神经元的顺序发送,或者每一列中神经元的参数的发送顺序相同,或者相邻列的神经元的参数的发送顺序相反。

在本申请的一种实施例中，所述第一行是所述神经网络中最上方一行，所述最后一行是所述神经网络中最下方一行；或者，所述第一行是所述神经网络中最下方一行，所述最后一行是所述神经网络中最上方一行。

在本申请的一种实施例中，每一行中神经元的参数按照从左往右的神经元的顺序发送，或者每一行中神经元的参数按照从右往左的神经元的顺序发送，或者每一行中神经元的参数的发送顺序相同，或者相邻行的神经元的参数的发送顺序相反。

在本申请的一种实施例中，在每层的神经元数目不是完全相同的情况下，每一行都按最多神经元数量的行发送所述神经元的参数，缺少神经元的层补预设值，或者，每一行都按实际神经元数量发送所述神经元的参数。

在本申请的一种实施例中，所述 AI 模型参数的发送顺序是预配置的或协议约定的或者网络侧指示的。

在本申请的一种实施例中，所述 AI 模型参数包括乘性系数和加性系数，所述 AI 模型中乘性系数的发送优先级高于加性系数的发送优先级；或者，所述 AI 模型中加性系数的发送优先级高于乘性系数的发送优先级。

在本申请的一种实施例中，所述 AI 模型的量化等级和/或压缩方式是预配置的或者协议约定的或者网络侧指示的。

在本申请的一种实施例中，更新后的 AI 模型的模型 ID 与更新前的 AI 模型的模型 ID 相同，不保存所述更新前的 AI 模型；

或者，

更新后的 AI 模型的模型 ID 与更新前的 AI 模型的模型 ID 不同，保存所述更新前的 AI 模型，且所述更新前的 AI 模型的模型 ID 不变。

在本申请的一种实施方式中，所述 AI 模型包括第一功能模块，所述第一功能模块用于以下至少一项：

(1) 信号处理，包括但不限于以下至少一项：信号检测、滤波、均衡等，其中信号包括但不限于以下至少一项：DMRS、SRS、SSB、TRS、PTRS、CSI-RS 等；

(2) 信道信号传输、信道信号接收、信道解调或信道信号发送，其中信道包括但不限于以下至少一项：PDCCH、PDSCH、PUCCH、PUSCH、PRACH、PBCH；

(3) 获取处理信道状态信息；

比如，信道状态信息反馈，包括但不限于以下至少一项：信道相关信息、信道矩阵相关信息、信道特征信息、信道矩阵特征信息、PMI、RI、CRI、CQI、LI 等。

又比如，FDD 上下行部分互易性。对于 FDD 系统，根据部分互异性，基站根据上行信道获取角度和时延信息，可以通过 CSI-RS 预编码或者直接指示的方法，将角度信息和时延信息通知终端，终端根据基站的指示上报或者在基站的指示范围内选择并上报，从而减少终端的计算量和 CSI 上报的开销。

(4) 波束管理，包括但不限于以下至少一项：波束测量、波束上报、波束预测、波

束失败检测、波束失败恢复、波束失败恢复中的新波束指示；

(5) 信道预测，包括但不限于以下至少一项：信道状态信息的预测、波束预测；

(6) 干扰抑制，包括但不限于以下至少一项：小区内干扰、小区间干扰、带外干扰、交调干扰等；

(7) 定位，比如通过参考信号（例如 SRS），估计出终端的具体位置（包括水平位置和/或垂直位置）或未来可能的轨迹，或估计出终端的辅助位置估计或轨迹估计的信息；

(8) 预测或管理高层业务和/或高层参数，包括但不限于以下至少一项：吞吐量、所需数据包大小、业务需求、移动速度、噪声信息等；

(9) 解析控制信令，包括但不限于以下至少一项：功率控制的相关信令，波束管理的相关信令。

在本申请实施例中，能够有效提升无线通信系统中传输 AI 模型的效率，以及减少网络信令开销。

参见图 10，本申请实施例提供一种更新 AI 模型参数的装置，应用于第一节点，该装置 1000 包括：

第一发送模块 1001，用于向第二节点发送与 AI 模型参数相关的第一信息；

其中，所述第一信息包括以下至少一项：所述 AI 模型参数的更新模式，所述 AI 模型参数的指示方式。

在本申请的一种实施方式中，所述更新模式包括以下之一：第一模式、第二模式、第三模式、第四模式；

其中，所述第一模式包括：更新第一参数子集中的所有参数，所述第一参数子集是所述 AI 模型参数的子集，所述第一参数子集是预配置或者协议约定的或者网络侧指示的；

所述第二模式包括：更新第二参数子集中的所有参数，所述第二参数子集是所述第一参数子集的子集；

所述第三模式包括：更新所述 AI 模型的所有参数；

所述第四模式包括：更新第三参数子集，所述第三参数子集是所述 AI 模型参数的子集。

在本申请的一种实施方式中，所述指示方式包括以下之一：

第一信息，所述第一信息用于指示所述 AI 模型参数在所述 AI 模型中的位置；

第二信息，所述第二信息用于指示所述 AI 模型参数的发送顺序。

在本申请的一种实施方式中，所述 AI 模型为神经网络，所述神经网络包括输入层、一个或多个隐藏层和输出层，所述输入层、一个或多个隐藏层、输出层，以及各层的神经元组成多行多列的结构，所述神经网络中各个神经元的参数的发送顺序是根据所述神经元在所述神经网络中的位置确定的。

在本申请的一种实施方式中，所述神经网络中各个神经元的参数的发送顺序，包括：第一列所有神经元的参数，第二列所有神经元的参数，以此类推，直至最后一列所有

神经元的参数；或，

第一行所有神经元的参数，第二行所有神经元的参数，以此类推，直至最后一行所有神经元的参数。

在本申请的一种实施方式中，所述第一列对应所述输入层，所述最后一列对应所述输出层，或者，所述第一列对应所述输出层，所述最后一列对应所述输入层；

或者，

所述第一行对应所述输入层，所述最后一行对应所述输出层，或者，所述第一行对应所述输出层，所述最后一行对应所述输入层。

在本申请的一种实施方式中，在所述神经网络中神经元的参数的发送顺序包括第一列所有神经元的参数，第二列所有神经元的参数，以此类推，直至最后一列所有神经元的参数的情况下，所述神经网络中神经元的参数的发送顺序还包括以下之一：

每一列中神经元的参数按照从上往下的神经元的顺序发送；

每一列中神经元的参数按照从下往上的神经元的顺序发送；

每一列中神经元的参数的发送顺序相同；

相邻列的神经元的参数的发送顺序相反。

在本申请的一种实施方式中，所述第一行是所述神经网络中最上方一行，所述最后一行是所述神经网络中最下方一行；或者，所述第一行是所述神经网络中最下方一行，所述最后一行是所述神经网络中最上方一行。

在本申请的一种实施方式中，在所述神经网络中神经元的参数的发送顺序包括第一行所有神经元的参数，第二行所有神经元的参数，以此类推，直至最后一行所有神经元的参数的情况下，所述神经网络中神经元的参数的发送顺序还包括以下之一：

每一行中神经元的参数按照从左往右的神经元的顺序发送；

每一行中神经元的参数按照从右往左的神经元的顺序发送；

每一行中神经元的参数的发送顺序相同；

相邻行的神经元的参数的发送顺序相反。

在本申请的一种实施方式中，所述 AI 模型参数的发送顺序是预配置的或协议约定的或者网络侧指示的。

在本申请的一种实施方式中，所述 AI 模型参数包括乘性系数和加性系数，所述乘性系数的发送优先级高于加性系数的发送优先级；或者，所述加性系数的发送优先级高于乘性系数的发送优先级。

在本申请的一种实施方式中，所述 AI 模型的量化等级和/或压缩方式是预配置的或者协议约定的或者网络侧指示的。

在本申请的一种实施方式中，更新后的 AI 模型的模型 ID 与更新前的 AI 模型的模型 ID 相同，不保存所述更新前的 AI 模型；

或者，

更新后的 AI 模型的模型 ID 与更新前的 AI 模型的模型 ID 不同，保存所述更新前的 AI 模型，且所述更新前的 AI 模型的模型 ID 不变。

在本申请的一种实施方式中，所述 AI 模型包括第一功能模块，所述第一功能模块用于以下一项或多项：

信号处理；
信道传输、信道接收、信道解调或信道发送；
获取信道状态信息；
波束管理；
信道预测；
干扰抑制；
定位；
预测或管理高层业务和/或高层参数的预测；
解析控制信令。

在本申请的一种实施方式中，所述第一节点为第一网络侧设备或第一终端，所述第二节点为第二网络侧设备或第二终端。

本申请实施例提供的装置能够实现图 4 的方法实施例实现的各个过程，并达到相同的技术效果，为避免重复，这里不再赘述。

参见图 11，本申请实施例提供一种更新 AI 模型参数的装置，应用于第二节点，该装置 1100 包括：

第一接收模块 1101，用于接收第一节点发送的与 AI 模型参数相关的第一信息；
更新模块 1102，用于根据所述第一信息更新所述 AI 模型参数；

其中，所述第一信息包括以下至少一项：所述 AI 模型参数的更新模式，所述 AI 模型参数的指示方式。

在本申请的一种实施方式中，所述更新模式包括以下之一：第一模式、第二模式、第三模式、第四模式；

其中，所述第一模式包括：更新第一参数子集中的所有参数，所述第一参数子集是所述 AI 模型参数的子集，所述第一参数子集是预配置或者协议约定的或者网络侧指示的；

所述第二模式包括：更新第二参数子集中的所有参数，所述第二参数子集是所述第一参数子集的子集，所述第一参数子集是所述 AI 模型参数的子集，所述第一参数子集是预先指示或协议约定的；

所述第三模式包括：更新所述 AI 模型的所有参数；

所述第四模式包括：更新第三参数子集，所述第三参数子集是所述 AI 模型参数的子集。

在本申请的一种实施方式中，所述指示方式包括以下之一：

第一信息，所述第一信息用于指示所述 AI 模型参数在所述 AI 模型中的位置；

第二信息，所述第二信息用于指示所述 AI 模型参数的发送顺序。

在本申请的一种实施方式中，所述 AI 模型为神经网络，所述神经网络包括输入层、一个或多个隐藏层和输出层，所述输入层、一个或多个隐藏层、输出层，以及各层的神经元组成多行多列的结构，所述神经网络中各个神经元的参数的发送顺序是根据所述神经元在所述神经网络中的位置确定的。

在本申请的一种实施方式中，所述神经网络中各个神经元的参数的发送顺序，包括：

第一列所有神经元的参数，第二列所有神经元的参数，以此类推，直至最后一列所有神经元的参数；

或者，

第一行所有神经元的参数，第二行所有神经元的参数，以此类推，直至最后一行所有神经元的参数。

在本申请的一种实施方式中，所述第一列对应所述输入层，所述最后一列对应所述输出层，或者，所述第一列对应所述输出层，所述最后一列对应所述输入层；

或者，

所述第一行对应所述输入层，所述最后一行对应所述输出层，或者，所述第一行对应所述输出层，所述最后一行对应所述输入层。

在本申请的一种实施方式中，在所述神经网络中神经元的参数的发送顺序包括第一列所有神经元的参数，第二列所有神经元的参数，以此类推，直至最后一列所有神经元的参数的情况下，所述神经网络中神经元的参数的发送顺序还包括以下之一：

每一列中神经元的参数按照从上往下的神经元的顺序发送；

每一列中神经元的参数按照从下往上的神经元的顺序发送；

每一列中神经元的参数的发送顺序相同；

相邻列的神经元的参数的发送顺序相反。

在本申请的一种实施方式中，所述第一行是所述神经网络中最上方一行，所述最后一行是所述神经网络中最下方一行；或者，所述第一行是所述神经网络中最下方一行，所述最后一行是所述神经网络中最上方一行。

在本申请的一种实施方式中，在所述神经网络中神经元的参数的发送顺序包括第一行所有神经元的参数，第二行所有神经元的参数，以此类推，直至最后一行所有神经元的参数的情况下，所述神经网络中神经元的参数的发送顺序还包括以下之一：

每一行中神经元的参数按照从左往右的神经元的顺序发送；

每一行中神经元的参数按照从右往左的神经元的顺序发送；

每一行中神经元的参数的发送顺序相同；

相邻行的神经元的参数的发送顺序相反。

在本申请的一种实施方式中，在每层的神经元数目不是完全相同的情况下，每一行都按最多神经元数量的行发送所述神经元的参数，缺少神经元的层补预设值，或者，每一行

都按实际神经元数量发送所述神经元的参数。

在本申请的一种实施方式中，所述 AI 模型参数的发送顺序是预配置的或协议约定的或者网络侧指示的。

在本申请的一种实施方式中，所述 AI 模型参数包括乘性系数和加性系数，乘性系数的发送优先级高于加性系数的发送优先级；或者，加性系数的发送优先级高于乘性系数的发送优先级。

在本申请的一种实施方式中，所述 AI 模型的量化等级和/或压缩方式是预配置的或者协议约定的或者网络侧指示的。

在本申请的一种实施方式中，更新后的 AI 模型的模型 ID 与更新前的 AI 模型的模型 ID 相同，不保存所述更新前的 AI 模型；

或者，

更新后的 AI 模型的模型 ID 与更新前的 AI 模型的模型 ID 不同，保存所述更新前的 AI 模型，且所述更新前的 AI 模型的模型 ID 不变。

在本申请的一种实施方式中，所述 AI 模型包括第一功能模块，所述第一功能模块用于以下一项或多项：

信号处理；

信道传输、信道接收、信道解调或信道发送；

获取信道状态信息；

波束管理；

信道预测；

干扰抑制；

定位；

预测或管理高层业务和/或高层参数的预测；

解析控制信令。

在本申请的一种实施方式中，所述第二节点为第二网络侧设备或第二终端，所述第一节点为第一网络侧设备或第一终端。

本申请实施例提供的装置能够实现图 5 的方法实施例实现的各个过程，并达到相同的技术效果，为避免重复，这里不再赘述。

图 12 为实现本申请实施例的一种终端的硬件结构示意图。该终端 1200 包括但不限于：射频单元 1201、网络模块 1202、音频输出单元 1203、输入单元 1204、传感器 1205、显示单元 1206、用户输入单元 1207、接口单元 1208、存储器 1209 以及处理器 1210 等中的至少部分部件。

本领域技术人员可以理解，终端 1200 还可以包括给各个部件供电的电源（比如电池），电源可以通过电源管理系统与处理器 1210 逻辑相连，从而通过电源管理系统实现管理充电、放电、以及功耗管理等功能。图 12 中示出的终端结构并不构成对终端的限定，终端

可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件布置，在此不再赘述。

应理解的是，本申请实施例中，输入单元 1204 可以包括图形处理单元（Graphics Processing Unit, GPU）12041 和麦克风 12042，图形处理器 12041 对在视频捕获模式或图像捕获模式中由图像捕获装置（如摄像头）获得的静态图片或视频的图像数据进行处理。显示单元 1206 可包括显示面板 12061，可以采用液晶显示器、有机发光二极管等形式来配置显示面板 12061。用户输入单元 1207 包括触控面板 12071 以及其他输入设备 12072 中的至少一种。触控面板 12071，也称为触摸屏。触控面板 12071 可包括触摸检测装置和触摸控制器两个部分。其他输入设备 12072 可以包括但不限于物理键盘、功能键（比如音量控制按键、开关按键等）、轨迹球、鼠标、操作杆，在此不再赘述。

本申请实施例中，射频单元 1201 接收来自网络侧设备的下行数据后，可以传输给处理器 1210 进行处理；另外，射频单元 1201 可以向网络侧设备发送上行数据。通常，射频单元 1201 包括但不限于天线、放大器、收发信机、耦合器、低噪声放大器、双工器等。

存储器 1209 可用于存储软件程序或指令以及各种数据。存储器 1209 可主要包括存储程序或指令的第一存储区和存储数据的第二存储区，其中，第一存储区可存储操作系统、至少一个功能所需的应用程序或指令（比如声音播放功能、图像播放功能等）等。此外，存储器 1209 可以包括易失性存储器或非易失性存储器，或者，存储器 1209 可以包括易失性和非易失性存储器两者。其中，非易失性存储器可以是只读存储器(Read-Only Memory, ROM)、可编程只读存储器(Programmable ROM, PROM)、可擦除可编程只读存储器(Erasable PROM, EPROM)、电可擦除可编程只读存储器(Electrically EPROM, EEPROM)或闪存。易失性存储器可以是随机存取存储器(Random Access Memory, RAM)，静态随机存取存储器(Static RAM, SRAM)、动态随机存取存储器(Dynamic RAM, DRAM)、同步动态随机存取存储器(Synchronous DRAM, SDRAM)、双倍数据速率同步动态随机存取存储器(Double Data Rate SDRAM, DDRSDRAM)、增强型同步动态随机存取存储器(Enhanced SDRAM, ESDRAM)、同步连接动态随机存取存储器(Synch link DRAM, SLDRAM)和直接内存总线随机存取存储器(Direct Rambus RAM, DRRAM)。本申请实施例中的存储器 1209 包括但不限于这些和任意其它适合类型的存储器。

处理器 1210 可包括一个或多个处理单元；可选的，处理器 1210 集成应用处理器和调制解调处理器，其中，应用处理器主要处理涉及操作系统、用户界面和应用程序等的操作，调制解调处理器主要处理无线通信信号，如基带处理器。可以理解的是，上述调制解调处理器也可以不集成到处理器 1210 中。

本申请实施例提供的终端能够实现图 4 的方法实施例实现的各个过程，并达到相同的技术效果，为避免重复，这里不再赘述。

请参阅图 13，图 13 是本申请实施例应用的网络侧设备的结构图，如图 13 所示，网络侧设备 1300 包括：处理器 1301、收发机 1302、存储器 1303 和总线接口，其中，处理

器 1301 可以负责管理总线架构和通常的处理。存储器 1303 可以存储处理器 1301 在执行操作时所使用的数据。

在本申请的一个实施例中，网络侧设备 1300 还包括：存储在存储器 1303 并可在处理器 1301 上运行的程序，程序被处理器 1301 执行时实现以上图 4 所示方法中的步骤。

在图 13 中，总线架构可以包括任意数量的互联的总线和桥，具体由处理器 1301 代表的一个或多个处理器和存储器 1303 代表的存储器的各种电路链接在一起。总线架构还可以将诸如外围设备、稳压器和功率管理电路等之类的各种其他电路链接在一起，这些都是本领域所公知的，因此，本文不再对其进行进一步描述。总线接口提供接口。收发机 1302 可以是多个元件，即包括发送机和接收机，提供用于在传输介质上与各种其他装置通信的单元。

可选的，如图 14 所示，本申请实施例还提供一种网络侧设备 1300，包括处理器 1401 和存储器 1402，存储器 1402 上存储有可在所述处理器 1401 上运行的程序或指令，例如，该通信设备 1400 为第一节点时，该程序或指令被处理器 1401 执行时实现上述图 4 方法实施例的各个步骤，该通信设备 1400 为第二节点时，该程序或指令被处理器 1401 执行时实现上述图 5 方法实施例的各个步骤且能达到相同的技术效果，为避免重复，这里不再赘述。

本申请实施例还提供一种可读存储介质，所述可读存储介质上存储有程序或指令，该程序或指令被处理器执行时实现图 4 或图 5 方法及上述各个实施例的各个过程，且能达到相同的技术效果，为避免重复，这里不再赘述。

其中，所述处理器为上述实施例中所述的终端中的处理器。所述可读存储介质，包括计算机可读存储介质，如计算机只读存储器 ROM、随机存取存储器 RAM、磁碟或者光盘等。

本申请实施例另提供了一种芯片，所述芯片包括处理器和通信接口，所述通信接口和所述处理器耦合，所述处理器用于运行程序或指令，实现图 4 或图 5 所示及上述各个方法实施例的各个过程，且能达到相同的技术效果，为避免重复，这里不再赘述。

应理解，本申请实施例提到的芯片还可以称为系统级芯片，系统芯片，芯片系统或片上系统芯片等。

本申请实施例另提供了一种计算机程序/程序产品，所述计算机程序/程序产品被存储在存储介质中，所述计算机程序/程序产品被至少一个处理器执行以实现图 4 或图 5 所示及上述各个方法实施例的各个过程，且能达到相同的技术效果，为避免重复，这里不再赘述。

本申请实施例另提供一种通信系统，所述通信系统包括终端与网络侧设备，所述终端用于执行如图 4 及上述各个方法实施例的各个过程，所述网络侧设备用于执行如图 5 及上述各个方法实施例的各个过程，且能达到相同的技术效果，为避免重复，这里不再赘述。

需要说明的是，在本文中，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素，

而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者装置所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括该要素的过程、方法、物品或者装置中还存在另外的相同要素。此外，需要指出的是，本申请实施方式中的方法和装置的范围不限按示出或讨论的顺序来执行功能，还可包括根据所涉及的功能按基本同时的方式或按相反的顺序来执行功能，例如，可以按不同于所描述的次序来执行所描述的方法，并且还可以添加、省去、或组合各种步骤。另外，参照某些示例所描述的特征可在其他示例中被组合。

通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以计算机软件产品的形式体现出来，该计算机软件产品存储在一个存储介质（如 ROM/RAM、磁碟、光盘）中，包括若干指令用以使得一台终端（可以是手机，计算机，服务器，空调器，或者网络设备等等）执行本申请各个实施例所述的方法。

上面结合附图对本申请的实施例进行了描述，但是本申请并不局限于上述的具体实施方式，上述的具体实施方式仅仅是示意性的，而不是限制性的，本领域的普通技术人员在本申请的启示下，在不脱离本申请宗旨和权利要求所保护的范围情况下，还可做出很多形式，均属于本申请的保护之内。

权利要求书

1. 一种更新 AI 模型参数的方法，包括：

第一节点向第二节点发送与 AI 模型参数相关的第一信息；

其中，所述第一信息包括以下至少一项：所述 AI 模型参数的更新模式，所述 AI 模型参数的指示方式。

2. 根据权利要求 1 所述的方法，其中，所述更新模式包括以下之一：第一模式、第二模式、第三模式、第四模式；

其中，所述第一模式包括：更新第一参数子集中的所有参数，所述第一参数子集是所述 AI 模型参数的子集，所述第一参数子集是预配置或者协议约定的或者网络侧指示的；

所述第二模式包括：更新第二参数子集中的所有参数，所述第二参数子集是所述第一参数子集的子集；

所述第三模式包括：更新所述 AI 模型的所有参数；

所述第四模式包括：更新第三参数子集，所述第三参数子集是所述 AI 模型参数的子集。

3. 根据权利要求 1 所述的方法，其中，所述指示方式包括以下之一：

第一信息，所述第一信息用于指示所述 AI 模型参数在 AI 模型中的位置；

第二信息，所述第二信息用于指示所述 AI 模型参数的发送顺序。

4. 根据权利要求 3 所述的方法，其中，所述 AI 模型为神经网络，所述神经网络包括输入层、一个或多个隐藏层和输出层，所述输入层、一个或多个隐藏层、输出层，以及各层的神经元组成多行多列的结构，所述神经网络中神经元的参数的发送顺序是根据所述神经元在所述神经网络中的位置确定的。

5. 根据权利要求 4 所述的方法，其中，所述神经网络中神经元的参数的发送顺序，包括：

第一列所有神经元的参数，第二列所有神经元的参数，以此类推，直至最后一列所有神经元的参数；或，

第一行所有神经元的参数，第二行所有神经元的参数，以此类推，直至最后一行所有神经元的参数。

6. 根据权利要求 5 所述的方法，其中，

所述第一列对应所述输入层，所述最后一列对应所述输出层，或者，所述第一列对应所述输出层，所述最后一列对应所述输入层；

或者，

所述第一行对应所述输入层，所述最后一行对应所述输出层；或者，所述第一行对应所述输出层，所述最后一行对应所述输入层。

7. 根据权利要求 5 所述的方法, 其中, 在所述神经网络中神经元的参数的发送顺序包括第一列所有神经元的参数, 第二列所有神经元的参数, 以此类推, 直至最后一列所有神经元的参数的情况下, 所述神经网络中神经元的参数的发送顺序还包括以下之一:

每一列中神经元的参数按照从上往下的神经元的顺序发送;

每一列中神经元的参数按照从下往上的神经元的顺序发送;

每一列中神经元的参数的发送顺序相同;

相邻列的神经元的参数的发送顺序相反。

8. 根据权利要求 5 所述的方法, 其中, 所述第一行是所述神经网络中最上方一行, 所述最后一行是所述神经网络中最下方一行; 或者, 所述第一行是所述神经网络中最下方一行, 所述最后一行是所述神经网络中最上方一行。

9. 根据权利要求 5 所述的方法, 其中, 在所述神经网络中神经元的参数的发送顺序包括第一行所有神经元的参数, 第二行所有神经元的参数, 以此类推, 直至最后一行所有神经元的参数的情况下, 所述神经网络中神经元的参数的发送顺序还包括以下之一:

每一行中神经元的参数按照从左往右的神经元的顺序发送;

每一行中神经元的参数按照从右往左的神经元的顺序发送;

每一行中神经元的参数的发送顺序相同;

相邻行的神经元的参数的发送顺序相反。

10. 根据权利要求 9 所述的方法, 其中, 在每层的神经元数目不是完全相同的情况下, 每一行都按最多神经元数量的行发送所述神经元的参数, 缺少神经元的层补预设值, 或者, 每一行都按实际神经元数量发送所述神经元的参数。

11. 根据权利要求 3 所述的方法, 其中, 所述 AI 模型参数的发送顺序是预配置的或协议约定的或者网络侧指示的。

12. 根据权利要求 3 所述的方法, 其中, 所述 AI 模型参数包括乘性系数和加性系数, 所述乘性系数的发送优先级高于所述加性系数的发送优先级;

或者,

所述加性系数的发送优先级高于所述乘性系数的发送优先级。

13. 根据权利要求 1 所述的方法, 其中, 所述 AI 模型的量化等级和/或压缩方式是预配置的或者协议约定的或者网络侧指示的。

14. 根据权利要求 1 所述的方法, 其中, 更新后的 AI 模型的模型 ID 与更新前的 AI 模型的模型 ID 相同, 不保存所述更新前的 AI 模型;

或者,

更新后的 AI 模型的模型 ID 与更新前的 AI 模型的模型 ID 不同, 保存所述更新前的 AI 模型, 且所述更新前的 AI 模型的模型 ID 不变。

15. 根据权利要求 1 所述的方法, 其中, 所述 AI 模型包括第一功能模块, 所述第一功能模块用于以下一项或多项:

信号处理；
信道传输、信道接收、信道解调或信道发送；
获取信道状态信息；
波束管理；
信道预测；
干扰抑制；
定位；
预测或管理高层业务和/或高层参数的预测；
解析控制信令。

16. 根据权利要求 1 所述的方法，其中，所述第一节点为第一网络侧设备或第一终端，所述第二节点为第二网络侧设备或第二终端。

17. 一种更新 AI 模型参数的方法，包括：

第二节点接收第一节点发送的与 AI 模型参数相关的第一信息；

所述第二节点根据所述第一信息更新所述 AI 模型参数；

其中，所述第一信息包括以下至少一项：所述 AI 模型参数的更新模式，所述 AI 模型参数的指示方式。

18. 根据权利要求 17 所述的方法，其中，所述更新模式包括以下之一：第一模式、第二模式、第三模式、第四模式；

其中，所述第一模式包括：更新第一参数子集中的所有参数，所述第一参数子集是所述 AI 模型参数的子集，所述第一参数子集是预配置或者协议约定的或者网络侧指示的；

所述第二模式包括：更新第二参数子集中的所有参数，所述第二参数子集是所述第一参数子集的子集，所述第一参数子集是所述 AI 模型参数的子集，所述第一参数子集是预先指示或协议约定的；

所述第三模式包括：更新所述 AI 模型的所有参数；

所述第四模式包括：更新第三参数子集，所述第三参数子集是所述 AI 模型参数的子集。

19. 根据权利要求 17 所述的方法，其中，所述指示方式包括以下之一：

第一信息，所述第一信息用于指示所述 AI 模型参数在所述 AI 模型中的位置；

第二信息，所述第二信息用于指示所述 AI 模型参数的发送顺序。

20. 根据权利要求 19 所述的方法，其中，所述 AI 模型为神经网络，所述神经网络包括输入层、一个或多个隐藏层和输出层，所述输入层、一个或多个隐藏层、输出层，以及各层的神经元组成多行多列的结构，所述神经网络中神经元的参数的发送顺序是根据所述神经元在所述神经网络中的位置确定的。

21. 根据权利要求 20 所述的方法，其中，所述神经网络中神经元的参数的发送顺序，包括：

第一列所有神经元的参数，第二列所有神经元的参数，以此类推，直至最后一列所有神经元的参数；

第一行所有神经元的参数，第二行所有神经元的参数，以此类推，直至最后一行所有神经元的参数。

22. 根据权利要求 21 所述的方法，其中，

所述第一列对应所述输入层，所述最后一列对应所述输出层，或者，所述第一列对应所述输出层，所述最后一列对应所述输入层；

或者，

所述第一行对应所述输入层，所述最后一行对应所述输出层；或者，所述第一行对应所述输出层，所述最后一行对应所述输入层。

23. 根据权利要求 21 所述的方法，其中，在所述神经网络中神经元的参数的发送顺序包括第一列所有神经元的参数，第二列所有神经元的参数，以此类推，直至最后一列所有神经元的参数的情况下，所述神经网络中神经元的参数的发送顺序还包括以下之一：

每一列中神经元的参数按照从上往下的神经元的顺序发送；

每一列中神经元的参数按照从下往上的神经元的顺序发送；

每一列中神经元的参数的发送顺序相同；

相邻列的神经元的参数的发送顺序相反。

24. 根据权利要求 21 所述的方法，其中，所述第一行是所述神经网络中最上方一行，所述最后一行是所述神经网络中最下方一行；或者，所述第一行是所述神经网络中最下方一行，所述最后一行是所述神经网络中最上方一行。

25. 根据权利要求 21 所述的方法，其中，在所述神经网络中神经元的参数的发送顺序包括第一行所有神经元的参数，第二行所有神经元的参数，以此类推，直至最后一行所有神经元的参数的情况下，所述神经网络中神经元的参数的发送顺序还包括以下之一：

每一行中神经元的参数按照从左往右的神经元的顺序发送；

每一行中神经元的参数按照从右往左的神经元的顺序发送；

每一行中神经元的参数的发送顺序相同；

相邻行的神经元的参数的发送顺序相反。

26. 根据权利要求 25 所述的方法，其中，在每层的神经元数目不是完全相同的情况下，每一行都按最多神经元数量的行发送所述神经元的参数，缺少神经元的层补预设值，或者，每一行都按实际神经元数量发送所述神经元的参数。

27. 根据权利要求 19 所述的方法，其中，所述 AI 模型参数的发送顺序是预配置的或协议约定的或者网络侧指示的。

28. 根据权利要求 19 所述的方法，其中，所述 AI 模型中乘性系数的发送优先级高于加性系数的发送优先级；

或者，

所述 AI 模型中加性系数的发送优先级高于乘性系数的发送优先级。

29. 根据权利要求 17 所述的方法，其中，所述 AI 模型的量化等级和/或压缩方式是预配置的或者协议约定的或者网络侧指示的。

30. 根据权利要求 17 所述的方法，其中，更新后的 AI 模型的模型 ID 与更新前的 AI 模型的模型 ID 相同，不保存所述更新前的 AI 模型；

或者，

更新后的 AI 模型的模型 ID 与更新前的 AI 模型的模型 ID 不同，保存所述更新前的 AI 模型，且所述更新前的 AI 模型的模型 ID 不变。

31. 根据权利要求 17 所述的方法，其中，所述 AI 模型包括第一功能模块，所述第一功能模块用于以下一项或多项：

信号处理；

信道传输、信道接收、信道解调或信道发送；

获取信道状态信息；

波束管理；

信道预测；

干扰抑制；

定位；

预测或管理高层业务和/或高层参数的预测；

解析控制信令。

32. 根据权利要求 17 所述的方法，其中，所述第二节点为第二网络侧设备或第二终端，所述第一节点为第一网络侧设备或第一终端。

33. 一种更新 AI 模型参数的装置，包括：

第一发送模块，用于向第二节点发送与 AI 模型参数相关的第一信息；

其中，所述第一信息包括以下至少一项：所述 AI 模型参数的更新模式，所述 AI 模型参数的指示方式。

34. 一种更新 AI 模型参数的装置，包括：

第一接收模块，用于接收第一节点发送的与 AI 模型参数相关的第一信息；

更新模块，用于根据所述第一信息更新所述 AI 模型参数；

其中，所述第一信息包括以下至少一项：所述 AI 模型参数的更新模式，所述 AI 模型参数的指示方式。

35. 一种通信设备，包括处理器，存储器及存储在所述存储器上并可在所述处理器上运行的程序或指令，所述程序或指令被所述处理器执行时实现如权利要求 1 至 32 中任一项所述的方法的步骤。

36. 一种可读存储介质，所述可读存储介质上存储程序或指令，所述程序或指令被处理器执行时实现如权利要求 1 至 32 中任一项所述的方法的步骤。

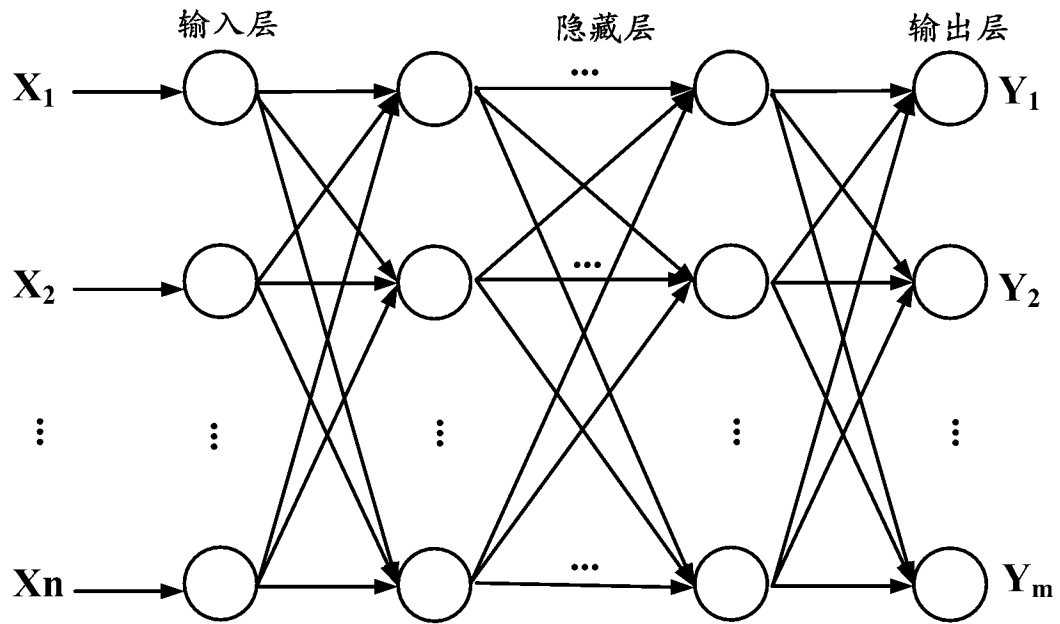


图 1

$$z = a_1 w_1 + \cdots + a_k w_k + \cdots + a_K w_K + b$$

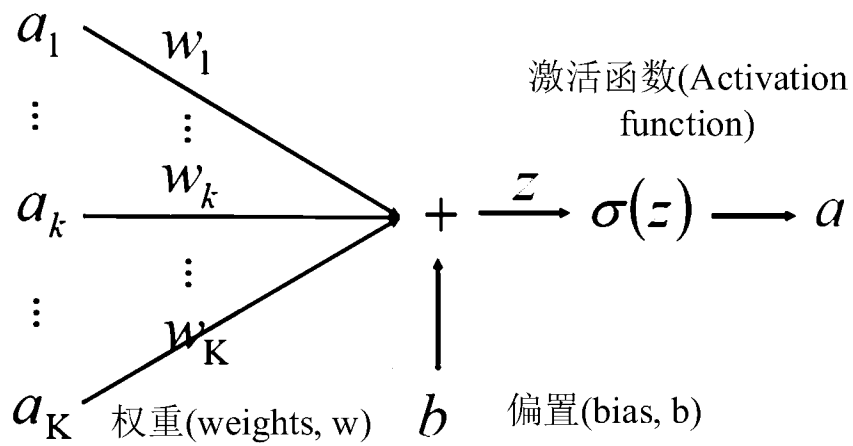


图 2

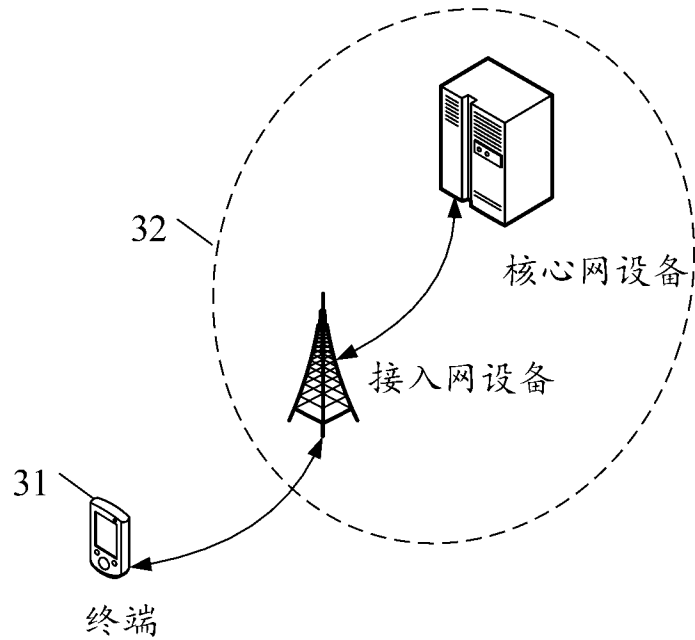


图 3

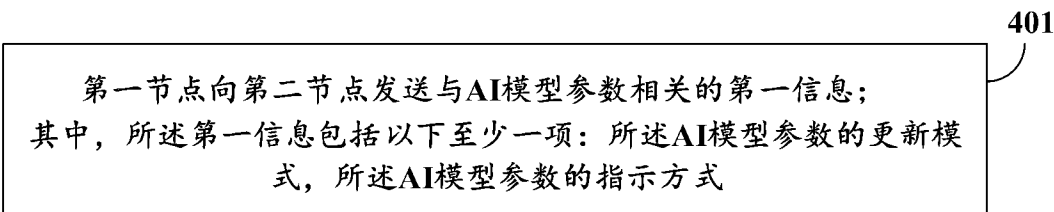


图 4

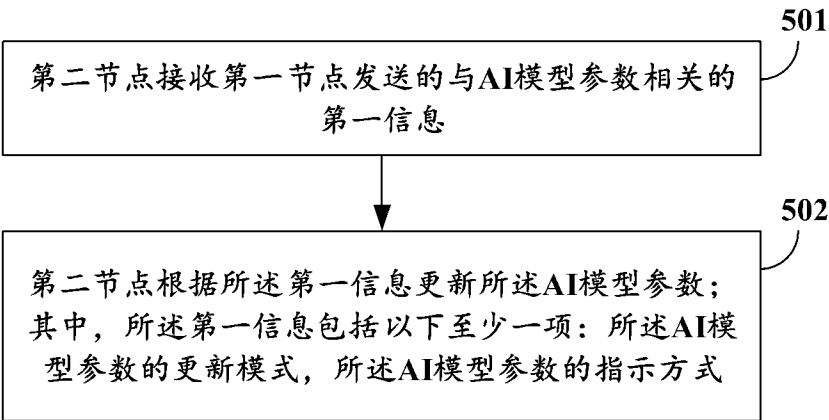


图 5

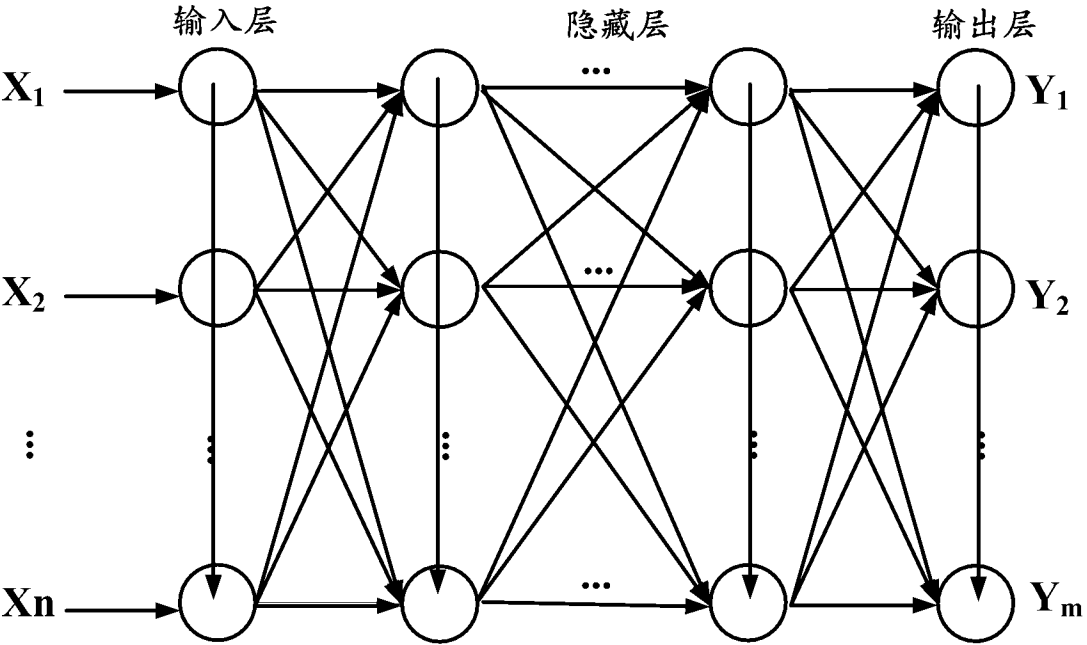


图 6

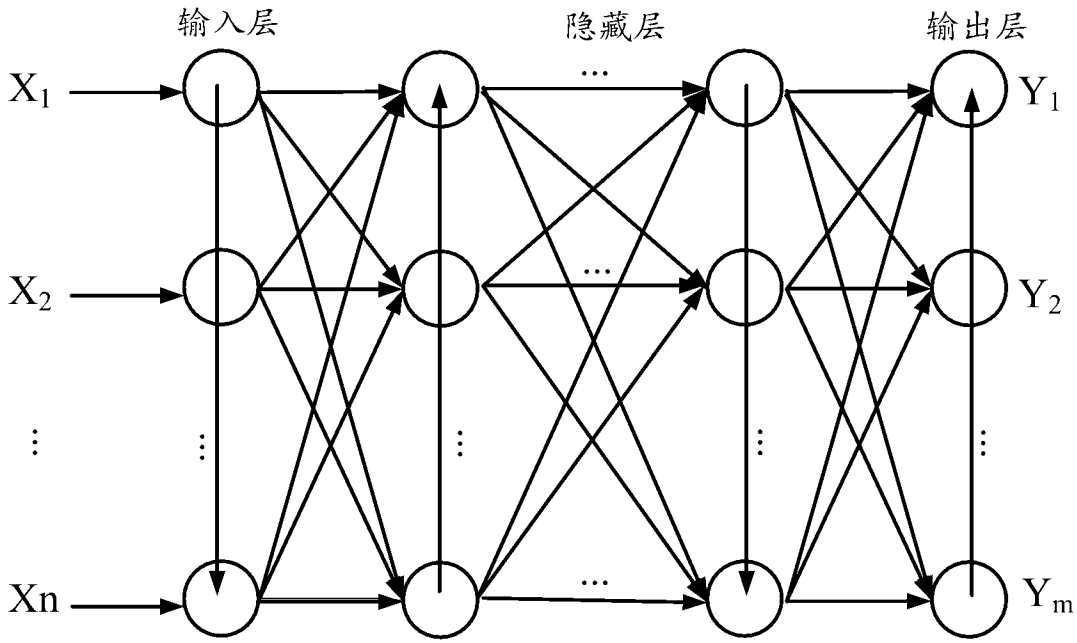


图 7

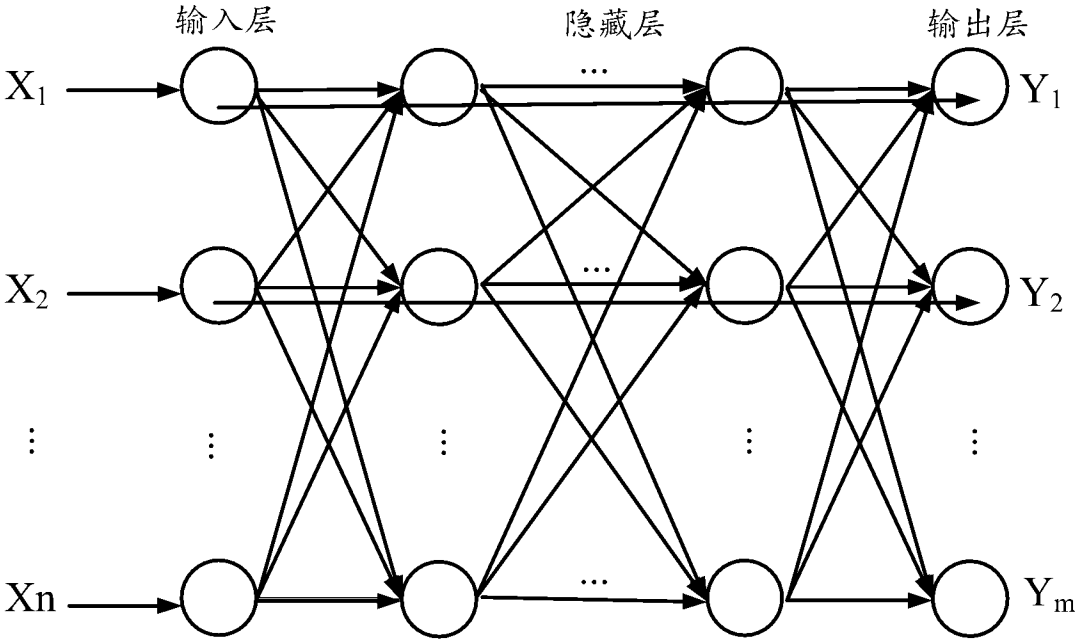


图 8

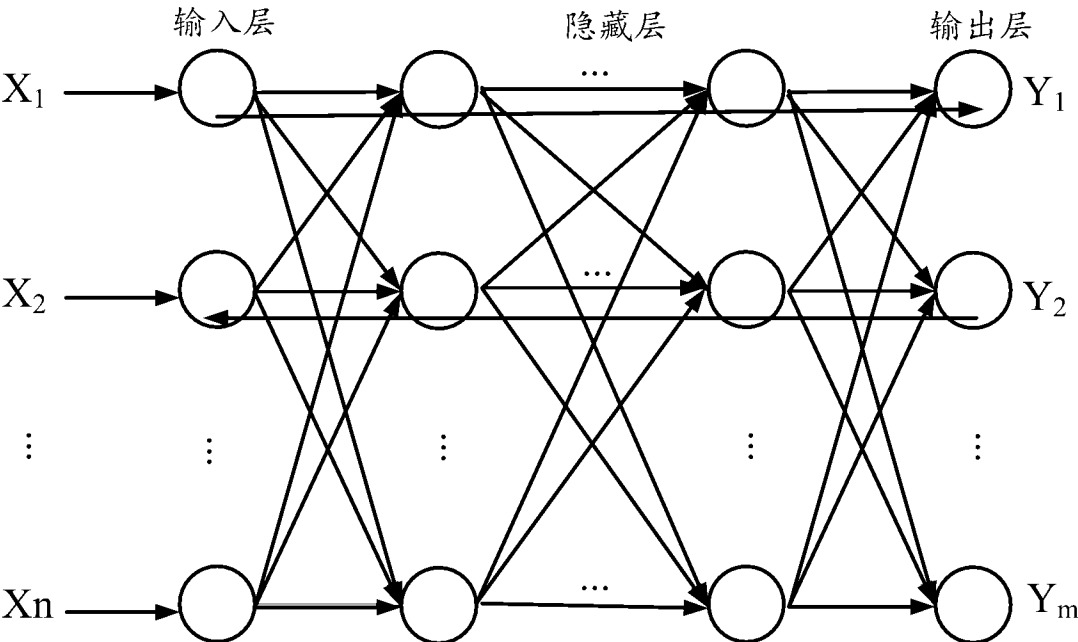


图 9

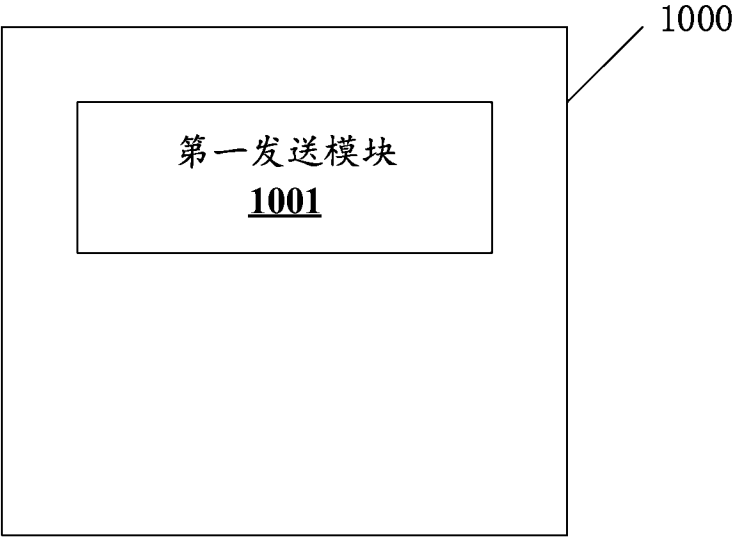


图 10

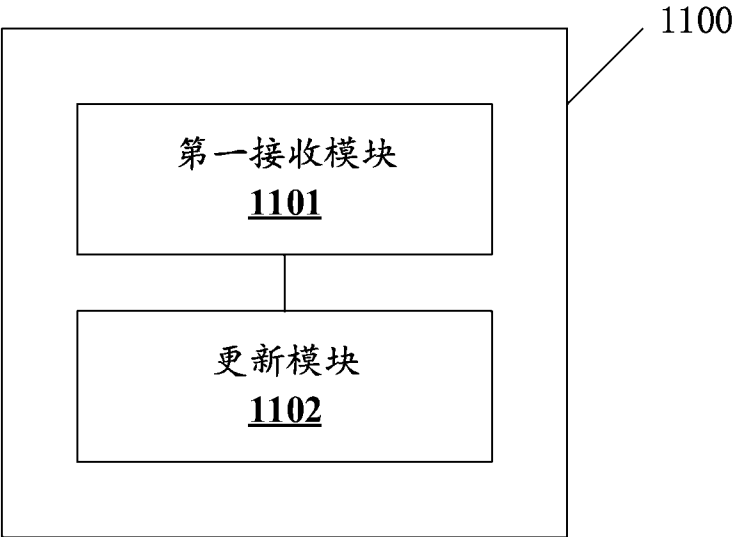


图 11

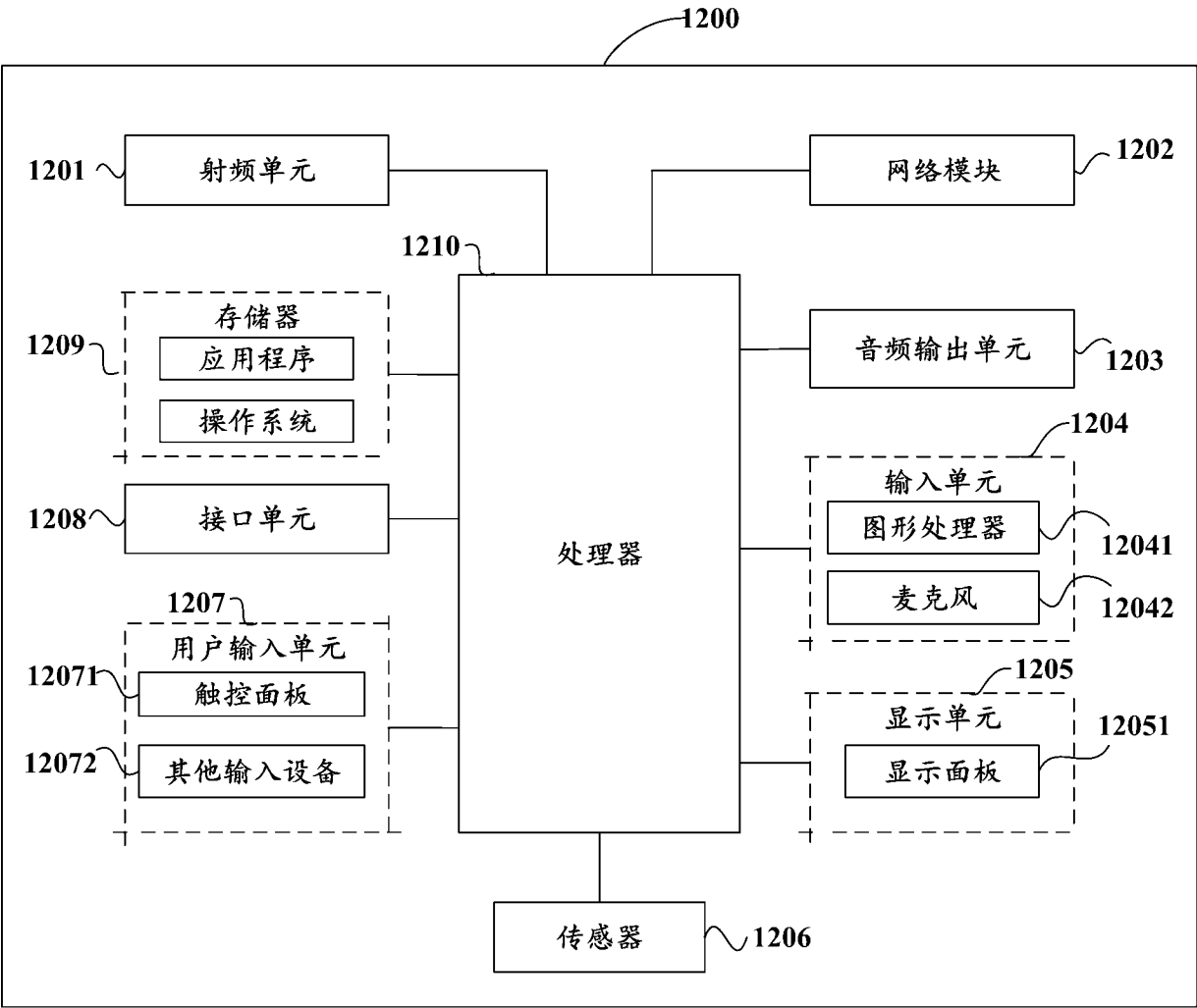


图 12

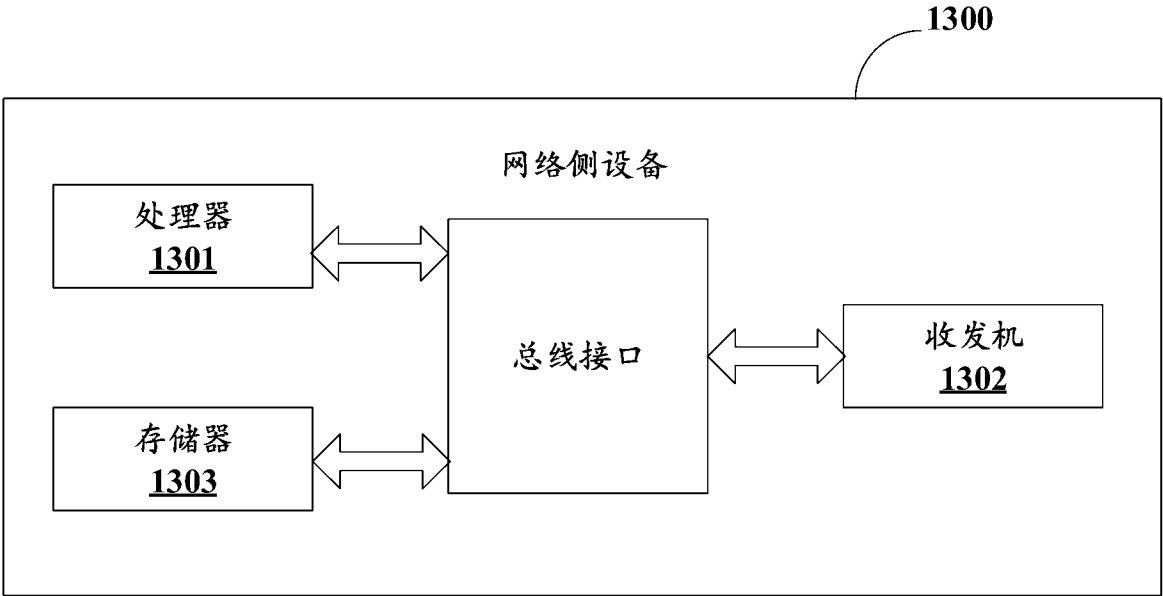


图 13

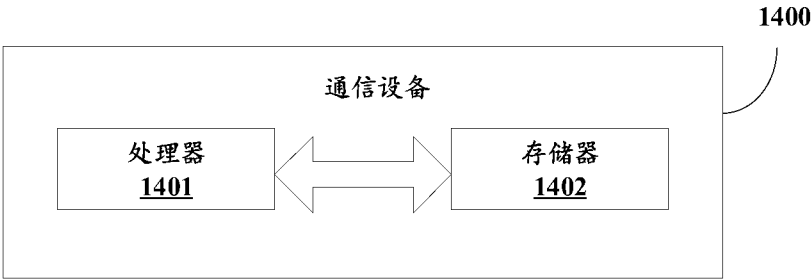


图 14

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/119938

A. CLASSIFICATION OF SUBJECT MATTER

H04L 41/082(2022.01)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)

IPC: H04L H04W 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)

CNXTX, ENTXTX, ENTXT, VEN: AI, 模型, 神经元, 神经网络, 参数, 更新, 子集, 层, 列, 行, 位置, 上, 下, 左, 右, 顺序, 指示, 人工智能, artificial, intelligence, AI, model, neuron, neural network, parameter, update, subset, layer, column, row, location, up, down, left, right, order, sequence, indicate

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 114363921 A (VIVO COMMUNICATION TECHNOLOGY CO., LTD.) 15 April 2022 (2022-04-15) description, paragraphs 22-102, 164 and 180	1-36
A	CN 114091679 A (HUAWEI TECHNOLOGIES CO., LTD.) 25 February 2022 (2022-02-25) entire document	1-36
A	CN 108924910 A (OPPO GUANGDONG MOBILE TELECOMMUNICATIONS CORP., LTD.) 30 November 2018 (2018-11-30) entire document	1-36
A	CN 114519435 A (VIVO COMMUNICATION TECHNOLOGY CO., LTD.) 20 May 2022 (2022-05-20) entire document	1-36
A	US 2020151558 A1 (GYRFALCON TECHNOLOGY INC.) 14 May 2020 (2020-05-14) entire document	1-36



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
 “D” document cited by the applicant in the international application
 “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

23 November 2023

Date of mailing of the international search report

29 November 2023

Name and mailing address of the ISA/CN

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

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/119938

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	114363921	A	15 April 2022	WO	2022078276	A1	21 April 2022
CN	114091679	A	25 February 2022	WO	2022041947	A1	03 March 2022
CN	108924910	A	30 November 2018	None			
CN	114519435	A	20 May 2022	None			
US	2020151558	A1	14 May 2020	US	2020151584	A1	14 May 2020
				US	2020151551	A1	14 May 2020

A. 主题的分类 H04L 41/082(2022.01)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) IPC: H04L H04W 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)) CNTXT,ENTXTC,ENTXT,VEN: AI, 模型, 神经元, 神经网络, 参数, 更新, 子集, 层, 列, 行, 位置, 上, 下, 左, 右, 顺序, 指示, 人工智能, artificial, intelligence, AI, model, neuron, neural network, parameter, update, subset, layer, column, row, location, up, down, left, right, order, sequence, indicate		
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
X	CN 114363921 A (维沃移动通信有限公司) 2022年4月15日 (2022 - 04 - 15) 说明书第22-102, 164, 180段	1-36
A	CN 114091679 A (华为技术有限公司) 2022年2月25日 (2022 - 02 - 25) 全文	1-36
A	CN 108924910 A (OPPO广东移动通信有限公司) 2018年11月30日 (2018 - 11 - 30) 全文	1-36
A	CN 114519435 A (维沃移动通信有限公司) 2022年5月20日 (2022 - 05 - 20) 全文	1-36
A	US 2020151558 A1 (GYRFALCON TECHNOLOGY INC.) 2020年5月14日 (2020 - 05 - 14) 全文	1-36
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
★ 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “D” 申请人在国际申请中引证的文件 “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 2023年11月23日		国际检索报告邮寄日期 2023年11月29日
ISA/CN的名称和邮寄地址 中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088		授权官员 王健 电话号码 (+86) 010-53961733

国际申请号
PCT/CN2023/119938

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	114363921	A	2022年4月15日	WO	2022078276	A1	2022年4月21日
CN	114091679	A	2022年2月25日	WO	2022041947	A1	2022年3月3日
CN	108924910	A	2018年11月30日	无			
CN	114519435	A	2022年5月20日	无			
US	2020151558	A1	2020年5月14日	US	2020151584	A1	2020年5月14日
				US	2020151551	A1	2020年5月14日