

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))。

network parameter deployment method, an electronic device, and a computer-readable storage medium. In the present application, the second neural network code and the second weight parameter which are relatively important are stored in the AI chip, the security level of the second neural network code and the second weight parameter is improved, and the bandwidth requirement of the NPU to access the off-chip memory can be reduced.

(57) 摘要: 本申请实施例提供了一种AI计算装置, 涉及终端领域。AI计算装置包括AI集成芯片及片外存储器, AI集成芯片包括CPU、NPU、第一片内存储单元及第二片内存储单元, 第二片内存储单元设置仅有允许NPU读取的权限及仅允许CPU写入的权限。片外存储器存储与NPU关联的第一神经网络代码及第一权重参数, 第二片内存储单元存储与NPU关联的第二神经网络代码及第二权重参数。本申请实施例还提供一种AI集成芯片、神经网络参数部署方法、电子设备及计算机可读存储介质。本申请将相对重要的第二神经网络代码和第二权重参数存储在AI芯片内部, 提高第二神经网络代码和第二权重参数的安全等级, 且可降低NPU访问片外存储器的带宽需求。

神经网络参数部署方法、AI 集成芯片及其相关装置

本申请要求于 2021 年 08 月 11 日提交中国专利局，申请号为 202110919731.6、申请名称为“神经网络参数部署方法、AI 集成芯片及其相关装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及人工智能领域，尤其涉及一种神经网络参数部署方法、人工智能 (Artificial Intelligence, AI) 集成芯片、AI 计算装置、电子设备及计算机可读存储介质。

背景技术

随着信息技术与 AI 技术的演进，AI 计算逐渐从云端扩展到终端，具备 AI 计算能力的智能终端逐渐在智能手机、公共安全、汽车辅助驾驶、智慧家庭等领域普及。然而现有的 AI 计算所需要的高算力引起的高成本、性能瓶颈、功耗瓶颈、安全隐患等始终是智能终端设备面临的挑战。

如图 1 所示，现有的 AI 计算系统一般包括神经网络处理单元 (Neural-network Processing Unit, NPU)、中央处理单元 (Central Processing Unit, CPU)、片内静态随机访问存储器 (Static Random Access Memory, SRAM) 及片外动态随机访问存储器 (Dynamic Random Access Memory, DRAM)。DRAM 用于存储 NPU 计算所需要的软件代码，权重参数、待处理数据等。NPU 的软件代码和权重参数具有极高的商业价值，存储在片外 DRAM 很难进行保护，容易被破解。

发明内容

有鉴于此，有必要提供一种神经网络参数部署方法、AI 集成芯片、AI 计算装置、电子设备及计算机可读存储介质，其可提升 NPU 软件代码和权重参数的安全等级。

本申请实施例第一方面公开了一种 AI 计算装置，AI 计算装置包括 AI 集成芯片及片外存储器。AI 集成芯片包括 CPU、NPU、第一片内存储单元及第二片内存储单元。片外存储器用于存储与 NPU 关联的第一神经网络代码及第一权重参数。第二片内存储单元用于存储与 NPU 关联的第二神经网络代码及第二权重参数，NPU 用于基于第一神经网络代码、第二神经网络代码、第一权重参数及第二权重参数对指定数据进行处理。第二片内存储单元设置有权仅允许 NPU 读取的权限，及仅允许 CPU 写入的权限。

通过采用该技术方案，增设第二片内存储单元来存储与 NPU 关联的第二神经网络代码及第二权重参数，例如第二片内存储单元可以是嵌入式非易失存储器，将相对重要的第二神经网络代码与第二权重参数存储在第二片内存储单元，将相对不重要的第一神经网络代码与第一权重参数存储在片外存储器，可提升第二神经网络代码与第二权重参数的安全等级，NPU 可从第二片内存储单元读取第二神经网络代码与第二权重参数，可降低 NPU 访问片外存储器的带宽需求，同时将第二片内存储单元设置为仅允许 NPU 读取、CPU 写入，可以进一步提升第二神经网络代码与第二权重参数的安全等级。

在一些实施例中，AI 集成芯片还包括第三片内存单元，第三片内存单元用于缓存从片外存储器读取的第一神经网络代码及第一权重参数的至少部分。

通过采用该技术方案，增设第三片内存单元来缓存 NPU 从片外存储器读取的第一神经网络代码及第一权重参数，例如第三片内存单元可以是嵌入式非易失存储器，相比 NPU 从片外存储器读取第一神经网络代码及第一权重参数的速度，NPU 从第三片内存单元读取第一神经网络代码及第一权重参数的速度更快，可以提升 NPU 的数据处理效率。

在一些实施例中，CPU 还用于接收密文数据，并将密文数据写入至第二片内存单元，密文数据为对第二神经网络代码及第二权重参数进行加密得到的。

通过采用该技术方案，AI 厂商将神经网络代码和权重参数使用加密密钥加密后发送给设备厂商，设备厂商将密文的神经网络代码以及权重参数写入芯片的内置 eNVM（上述第二片内存单元）中，而该芯片内的一次性可编程 OTP 器件中内置有解密密钥，该解密密钥用于对密文的神经网络代码以及权重参数进行解密，该解密密钥是由 AI 厂商和芯片厂商协商好后并由芯片厂商植入芯片内部的，AI 厂商无需将解密密钥发送给设备厂商，而是由芯片厂商内置在芯片中，设备厂商的设备在使用了上述芯片后可以正常使用该解密密钥解密密文的神经网络代码和权重参数，设备厂商或其他任何第三方厂商无法获取明文的神经网络代码，有效保护了 AI 厂商的知识产权。

在一些实施例中，AI 集成芯片还包括密钥存储单元及解密单元，解密密钥在密文数据写入至第二片内存单元之前存储至密钥存储单元，密钥存储单元设置仅有允许解密单元读取的权限，NPU 还用于从第二片内存单元读取密文数据，并调用解密单元使用解密密钥对密文数据进行解密，得到第二神经网络代码及第二权重参数。

通过采用该技术方案，由芯片厂商将解密密钥先写入至芯片内的密钥存储单元如 OTP 器件中，避免解密密钥泄露给设备厂商，同时可以仅为密钥存储单元设置一个读端口，该读端口提供给解密单元，实现仅允许解密单元读取解密密钥，进一步避免解密密钥泄露，再者可以在 NPU 从第二片内存单元读取密文数据时再对密文数据进行解密处理，避免设备厂商获取明文的第二神经网络代码及第二权重参数。

在一些实施例中，AI 集成芯片还包括密钥存储单元及解密单元，密钥存储单元用于存储解密密钥，密钥存储单元设置仅有允许解密单元读取的权限，CPU 用于接收密文数据，并调用解密单元使用解密密钥对密文数据进行解密，密文数据为对第二神经网络代码及第二权重参数进行加密得到的，CPU 还用于将解密得到的第二神经网络代码及第二权重参数写入至第二片内存单元。

通过采用该技术方案，由芯片厂商将解密密钥先写入至芯片内的密钥存储单元如 OTP 器件中，示例性的，该 OTP 器件仅设置有一个读端口，该读端口连接至芯片内的解密单元，该解密密钥是由 AI 厂商和芯片厂商协商好后并由芯片厂商植入芯片内部的。设备厂商从 AI 厂商购买密文的神经网络代码和权重参数，以及从芯片厂商购买植入了解密密钥的芯片，该芯片内还设置有 eNVM，设备厂商购买了该密文的神经网络代码和权重参数后，将该密文的神经网络代码和权重参数写入该芯片内的 eNVM 中，芯片内的解密单元通过调用内置在 OTP 器件中的解密密钥对密文的神经网络代码和权重参数进行解密后使用。这样，AI 厂商无需将解密密钥发送给设备厂商，而是由芯片厂商内置在芯片中，设备厂商的设备在使用了上述芯片后可以正常使用该解密密钥解密密文的神经网络代码和权重参数，设备厂商或其他任何第三方厂商无法获取明文的神经网络代码，有效保护了 AI 厂商的知识产权。

在一些实施例中，密钥存储单元包括一次性可编程 OTP 器件。

通过采用该技术方案，使用 OTP 器件存储解密密钥，由于 OTP 器件的一次性可编程特性，可以避免解密密钥被篡改，且由于 OTP 器件具有很强的抗攻击、抗反向工程能力，可提升解密密钥的存储安全性。

在一些实施例中，片外存储器还用于存储指定数据及指定数据的处理结果，第一片内存存储单元用于缓存从片外存储器读取的指定数据。

通过采用该技术方案，通过第一片内存存储单元来缓存从片外存储器读取的指定数据，例如第一片内存存储单元可以是片内 SRAM，相比 NPU 从片外存储器读取指定数据的速度，NPU 从第一片内存存储单元读取指定数据的速度更快，可以加快 NPU 读取指定数据的速度，提升 NPU 的数据处理效率。

在一些实施例中，第一片内存存储单元包括 SRAM，片外存储器包括 DRAM，第二片内存存储单元包括 eFlash、eNVM、MRAM 或 RRAM 中的任意一种，第三片内存存储单元包括 eFlash、eNVM、MRAM 或 RRAM 中的任意一种。

本申请实施例第二方面公开了一种 AI 集成芯片，AI 集成芯片包括 CPU、NPU、第一片内存存储单元及第二片内存存储单元。第二片内存存储单元用于存储与 NPU 关联的神经网络代码及权重参数，NPU 用于基于神经网络代码及权重参数对指定数据进行处理。第二片内存存储单元包括 eFlash、eNVM、MRAM 或 RRAM 中的任意一种。

通过采用该技术方案，增设第二片内存存储单元来存储与 NPU 关联的神经网络代码及权重参数，相比现有将神经网络代码及权重参数存储在片外存储器，可提升神经网络代码与权重参数的安全等级，NPU 从第二片内存存储单元读取神经网络代码与权重参数，可降低 NPU 访问片外存储器的带宽需求，提升 NPU 性能，同时将神经网络代码及权重参数写入至片内 eFlash、eNVM、MRAM 或 RRAM 进行储存，设备厂商无法窃取 AI 厂商所提供的神经网络代码及权重参数。

在一些实施例中，第二片内存存储单元仅设置一个读端口与一个写端口，读端口电连接于 NPU，写端口电连接于所述 CPU。

通过采用该技术方案，将第二片内存存储单元设置为仅允许 NPU 读取的权限及仅允许 CPU 写入的权限，可以进一步提高第二片内存存储单元存储的神经网络代码与权重参数的安全等级。

在一些实施例中，CPU 还用于接收密文数据，并将密文数据写入至第二片内存存储单元，密文数据为对神经网络代码及权重参数进行加密得到的。

通过采用该技术方案，AI 厂商将神经网络代码和权重参数使用加密密钥加密后发送给设备厂商，设备厂商将密文的神经网络代码以及权重参数写入芯片的内置 eNVM（上述第二片内存存储单元）中，而该芯片内的一次性可编程 OTP 器件中内置有解密密钥，该解密密钥用于对密文的神经网络代码以及权重参数进行解密，该解密密钥是由 AI 厂商和芯片厂商协商好后并由芯片厂商植入芯片内部的，AI 厂商无需将解密密钥发送给设备厂商，而是由芯片厂商内置在芯片中，设备厂商的设备在使用了上述芯片后可以正常使用该解密密钥解密密文的神经网络代码和权重参数，设备厂商或其他任何第三方厂商无法获取明文的神经网络代码，有效保护了 AI 厂商的知识产权。

在一些实施例中，AI 集成芯片还包括密钥存储单元及解密单元，解密密钥在密文数据写入至第二片内存存储单元之前存储至密钥存储单元，密钥存储单元设置有仅允许解密单元读

取的权限，NPU 还用于从第二片内存储单元读取密文数据，并调用解密单元使用解密密钥对密文数据进行解密，得到神经网络代码及权重参数。

通过采用该技术方案，由芯片厂商将解密密钥先写入至芯片内的密钥存储单元如 OTP 器件中，避免解密密钥泄露给设备厂商，同时可以仅为密钥存储单元设置一个读端口，该读端口提供给解密单元，实现仅允许解密单元读取解密密钥，进一步避免解密密钥泄露，再者可以在 NPU 从第二片内存储单元读取密文数据时再对密文数据进行解密处理，避免设备厂商获取明文的第二神经网络代码及第二权重参数。

在一些实施例中，AI 集成芯片还包括密钥存储单元及解密单元，密钥存储单元用于存储解密密钥，密钥存储单元设置有仅允许解密单元读取的权限，CPU 用于接收密文数据，并调用解密单元使用解密密钥对密文数据进行解密，密文数据为对神经网络代码及权重参数进行加密得到的，CPU 还用于将解密得到的神经网络代码及权重参数写入至第二片内存储单元。

通过采用该技术方案，由芯片厂商将解密密钥先写入至芯片内的密钥存储单元如 OTP 器件中，示例性的，该 OTP 器件仅设置有一个读端口，该读端口连接至芯片内的解密单元，该解密密钥是由 AI 厂商和芯片厂商协商好后并由芯片厂商植入芯片内部的。设备厂商从 AI 厂商购买密文的神经网络代码和权重参数，以及从芯片厂商购买植了解密密钥的芯片，该芯片内还设置有 eNVM，设备厂商购买了该密文的神经网络代码和权重参数后，将该密文的神经网络代码和权重参数写入该芯片内的 eNVM 中，芯片内的解密单元通过调用内置在 OTP 器件中的解密密钥对密文的神经网络代码和权重参数进行解密后使用。这样，AI 厂商无需将解密密钥发送给设备厂商，而是由芯片厂商内置在芯片中，设备厂商的设备在使用了上述芯片后可以正常使用该解密密钥解密密文的神经网络代码和权重参数，设备厂商或其他任何第三方厂商无法获取明文的神经网络代码，有效保护了 AI 厂商的知识产权。

在一些实施例中，密钥存储单元包括一次性可编程 OTP 器件。

通过采用该技术方案，使用 OTP 器件存储解密密钥，由于 OTP 器件的一次性可编程特性，可以避免解密密钥被篡改，且由于 OTP 器件具有很强的抗攻击、抗反向工程能力，可提升解密密钥的存储安全性。

第三方面，本申请实施例提供一种神经网络参数部署方法，应用于 AI 计算装置，AI 计算装置包括 AI 集成芯片及片外存储器。AI 集成芯片包括 CPU、NPU、第一片内存储单元及第二片内存储单元。神经网络参数部署包括：将与 NPU 关联的第一神经网络代码、第一权重参数存储至片外存储器；将与 NPU 关联的第二神经网络代码及第二权重参数存储至第二片内存储单元；其中，NPU 用于基于第一神经网络代码、第二神经网络代码、第一权重参数及第二权重参数对指定数据进行处理，第二片内存储单元设置有仅允许 NPU 读取的权限，及仅允许 CPU 写入的权限。

通过采用该技术方案，增设第二片内存储单元来存储与 NPU 关联的第二神经网络代码及第二权重参数，例如第二片内存储单元可以是嵌入式非易失存储器，将相对重要的第二神经网络代码与第二权重参数存储在第二片内存储单元，将相对不重要的第一神经网络代码与第一权重参数存储在片外存储器，可提升第二神经网络代码与第二权重参数的安全等级，NPU 可从第二片内存储单元读取神经网络代码与权重参数，可降低 NPU 访问片外存储器的带宽需求，同时将第二片内存储单元设置为仅允许 NPU 读取、CPU 写入，可以进一步提升第二神经网络代码与第二权重参数的安全等级。

在一些实施例中，AI 集成芯片还包括第三片内存储单元，神经网络参数部署方法还

包括：从片外存储器读取第一神经网络代码及第一权重参数后缓存在第三片内存储单元中。

通过采用该技术方案，增设第三片内存储单元来缓存 NPU 从片外存储器读取的第一神经网络代码及第一权重参数，例如第三片内存储器可以是嵌入式非易失存储器，相比 NPU 从片外存储器读取第一神经网络代码及第一权重参数的速度，NPU 从第三片内存储单元读取第一神经网络代码及第一权重参数的速度更快，可以提升 NPU 的数据处理效率。

在一些实施例中，神经网络参数部署方法还包括：接收加密装置发送的密文数据，并将密文数据写入至第二片内存储单元；其中，密文数据为对第二神经网络代码及第二权重参数进行加密得到的。

通过采用该技术方案，AI 厂商将神经网络代码和权重参数使用加密密钥加密后发送给设备厂商，设备厂商将密文的神经网络代码以及权重参数写入芯片的内置 eNWM（上述第二片内存储单元）中，而该芯片内的一次性可编程 OTP 器件中内置有解密密钥，该解密密钥用于对密文的神经网络代码以及权重参数进行解密，该解密密钥是由 AI 厂商和芯片厂商协商好后并由芯片厂商植入芯片内部的，AI 厂商无需将解密密钥发送给设备厂商，而是由芯片厂商内置在芯片中，设备厂商的设备在使用了上述芯片后可以正常使用该解密密钥解密密文的神经网络代码和权重参数，设备厂商或其他任何第三方厂商无法获取明文的神经网络代码，有效保护了 AI 厂商的知识产权。

在一些实施例中，神经网络参数部署方法还包括：当 NPU 从第二片内存储单元读取密文数据时，调用预先存储的解密密钥对密文数据进行解密，得到第二神经网络代码及第二权重参数。

通过采用该技术方案，由芯片厂商将解密密钥先写入至芯片内的密钥存储单元如 OTP 器件中，避免解密密钥泄露给设备厂商，同时可以仅为密钥存储单元设置一个读端口，该读端口提供给解密单元，实现仅允许解密单元读取解密密钥，进一步避免解密密钥泄露，再者可以在 NPU 从第二片内存储单元读取密文数据时再对密文数据进行解密处理，避免设备厂商获取明文的第二神经网络代码及第二权重参数。

在一些实施例中，神经网络参数部署方法还包括：接收加密装置发送的密文数据，并调用预先存储的解密密钥对密文数据进行解密；将解密得到的第二神经网络代码及第二权重参数写入至第二片内存储单元；其中，密文数据为对第二神经网络代码及第二权重参数进行加密得到的。

通过采用该技术方案，由芯片厂商将解密密钥先写入至芯片内的密钥存储单元如 OTP 器件中，示例性的，该 OTP 器件仅设置有一个读端口，该读端口连接至芯片内的解密单元，该解密密钥是由 AI 厂商和芯片厂商协商好后并由芯片厂商植入芯片内部的。设备厂商从 AI 厂商购买密文的神经网络代码和权重参数，以及从芯片厂商购买植了解密密钥的芯片，该芯片内还设置有 eNVM，设备厂商购买了该密文的神经网络代码和权重参数后，将该密文的神经网络代码和权重参数写入该芯片内的 eNVM 中，芯片内的解密单元通过调用内置在 OTP 器件中的解密密钥对密文的神经网络代码和权重参数进行解密后使用。这样，AI 厂商无需将解密密钥发送给设备厂商，而是由芯片厂商内置在芯片中，设备厂商的设备在使用了上述芯片后可以正常使用该解密密钥解密密文的神经网络代码和权重参数，设备厂商或其他任何第三方厂商无法获取明文的神经网络代码，有效保护了 AI 厂商的知识产权。

第四方面，本申请实施例提供一种神经网络参数部署方法，包括：第一处理端将解

密密钥写入至 AI 集成芯片的 OTP 器件；第二处理端对神经网络代码及权重参数进行加密，得到密文数据；第三处理端将密文数据写入至 AI 集成芯片的片内存储单元；其中，解密密钥用于对密文数据进行解密，片内存储单元包括 eFalsh、eNVM、MRAM 或 RRAM 中的任意一种。

通过采用该技术方案，相比现有方案由设备厂商将解密密钥写入至 AI 集成芯片，改由芯片厂商将解密密钥先写入至芯片内的密钥存储单元如 OTP 器件中，示例性的，该 OTP 器件仅设置有一个读端口，该读端口连接至芯片内的解密单元，示例性的，该解密密钥是由 AI 厂商和芯片厂商协商好后并由芯片厂商植入芯片内部的，设备厂商从 AI 厂商购买密文的神经网络代码和权重参数，以及从芯片厂商购买植入了解密密钥的芯片，该芯片内还设置有 eFalsh、eNVM、MRAM 或 RRAM，设备厂商购买了该密文的神经网络代码和权重参数后，将该密文的神经网络代码和权重参数写入该芯片内的 eFalsh、eNVM、MRAM 或 RRAM 中，芯片内的解密单元通过调用内置在 OTP 器件中的解密密钥对密文的神经网络代码和权重参数进行解密后使用。这样，AI 厂商无需将解密密钥发送给设备厂商，而是由芯片厂商内置在芯片中，设备厂商的设备在使用了上述芯片后可以正常使用该解密密钥解密密文的神经网络代码和权重参数，设备厂商或其他任何第三方厂商无法获取明文的神经网络代码，有效保护了 AI 厂商的知识产权，再者将密文数据存储于片内 eFalsh、eNVM、MRAM 或 RRAM 中，相比现有方案将神经网络代码与权重参数存储在片外存储器，可以进一步提升神经网络代码与权重参数的安全等级。

第五方面，本申请实施例提供一种神经网络参数部署方法，包括：获取密文数据，密文数据为对神经网络代码及权重参数进行加密得到的；将密文数据写入至 AI 集成芯片的片内存储单元；其中，AI 集成芯片的 OTP 器件中预先写入有对密文数据进行解密的解密密钥，片内存储单元包括 eFalsh、eNVM、MRAM 或 RRAM 中的任意一种。

通过采用该技术方案，对于设备厂商而言，解密密钥是由芯片厂商写入至 AI 集成芯片的 OTP 器件，例如该解密密钥是由 AI 厂商和芯片厂商协商好后，由芯片厂商植入芯片内部，设备厂商从 AI 厂商购买密文的神经网络代码和权重参数，以及从芯片厂商购买植入了解密密钥的芯片，该芯片内还设置有 eFalsh、eNVM、MRAM 或 RRAM，设备厂商购买了该密文的神经网络代码和权重参数后，将该密文的神经网络代码和权重参数写入该芯片内的 eFalsh、eNVM、MRAM 或 RRAM 中，芯片内的解密单元通过调用内置在 OTP 器件中的解密密钥对密文的神经网络代码和权重参数进行解密后使用，AI 厂商无需将解密密钥发送给设备厂商，而是由芯片厂商内置在芯片中，设备厂商的设备在使用了上述芯片后可以正常使用该解密密钥解密密文的神经网络代码和权重参数，设备厂商或其他任何第三方厂商无法获取明文的神经网络代码，有效保护了 AI 厂商的知识产权，再者将密文数据存储于片内 eFalsh、eNVM、MRAM 或 RRAM 中，相比现有方案将神经网络代码与权重参数存储在片外存储器，可以进一步提升神经网络代码与权重参数的安全等级。

在一些实施例中，AI 集成芯片包括元 CPU、NPU、第一片内存储单元及第二片内存储单元，将密文数据写入至 AI 集成芯片的片内存储单元，包括：将密文数据写入至第二片内存储单元，其中第二片内存储单元设置仅有允许 NPU 读取的权限，及仅允许 CPU 写入的权限。

通过采用该技术方案，NPU 可从第二片内存储单元读取神经网络代码与权重参数，可降低 NPU 访问片外存储器的带宽需求，同时将第二片内存储单元设置为仅允许 NPU 读取、CPU 写入，可以进一步提升第二神经网络代码与第二权重参数的安全等级。

在一些实施例中，神经网络参数部署方法还包括：当 NPU 从第二片内存储单元读取密文数据时，调用 OTP 器件存储的解密密钥对密文数据进行解密，得到神经网络代码及权重参数。

通过采用该技术方案，由芯片厂商将解密密钥先写入至芯片内的密钥存储单元如 OTP 器件中，避免解密密钥泄露给设备厂商，同时可以仅为密钥存储单元设置一个读端口，该读端口提供给解密单元，实现仅允许解密单元读取解密密钥，进一步避免解密密钥泄露，再者可以在 NPU 从第二片内存储单元读取密文数据时再对密文数据进行解密处理，避免设备厂商获取明文的第二神经网络代码及第二权重参数。

在一些实施例中，将密文数据写入至 AI 集成芯片的片内存储单元，包括：调用解密密钥对密文数据进行解密；将解密得到的神经网络代码及权重参数写入至第二片内存储单元。

通过采用该技术方案，由芯片厂商将解密密钥先写入至芯片内的密钥存储单元如 OTP 器件中，避免解密密钥泄露给设备厂商，同时可以仅为密钥存储单元设置一个读端口，该读端口提供给解密单元，实现仅允许解密单元读取解密密钥，进一步避免解密密钥泄露，再者可以在将密文数据写入至 AI 集成芯片的片内存储单元时再对密文数据进行解密处理，避免设备厂商获取明文的第二神经网络代码及第二权重参数。

第六方面，本申请实施例提供一种计算机可读存储介质，包括计算机指令，当计算机指令在电子设备上运行时，使得电子设备执行如第三方面所述的神经网络参数部署方法。

第七方面，本申请实施例提供一种电子设备，包括如第一方面所述的 AI 计算装置或者包括如第二方面所述的 AI 集成芯片。

第八方面，本申请实施例提供一种计算机程序产品，当计算机程序产品在计算机上运行时，使得计算机执行如第三方面所述的神经网络参数部署方法。

可以理解地，上述提供的第四方面所述的计算机可读存储介质，第五方面所述的电子设备，第六方面所述的计算机程序产品，第七方面所述的装置均与上述第二方面或第三方面的方法对应，因此，其所能达到的有益效果可参考上文所提供的对应的方法中的有益效果，此处不再赘述。

附图说明

- 图 1 为现有技术中的 AI 计算系统的架构示意图；
- 图 2 为本申请一实施例提供的 AI 计算装置的架构示意图；
- 图 3a~3b 为本申请另一实施例提供的 AI 计算装置的架构示意图；
- 图 4 为本申请一实施例提供的解密单元对密文数据进行解密的示意图；
- 图 5 为本申请一实施例提供的神经网络参数部署方法的流程示意图；
- 图 6 为本申请另一实施例提供的神经网络参数部署方法的流程示意图；
- 图 7 为本申请又一实施例提供的神经网络参数部署方法的流程示意图；
- 图 8 为本申请实施例提供的一种可能的电子设备。

具体实施方式

需要说明的是，本申请中“至少一个”是指一个或者多个，“多个”是指两个或多于两个。“和/或”，描述关联对象的关联关系，表示可以存在三种关系，例如，A 和/或 B 可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 的情况，其中 A，B 可以是单数或者

复数。本申请的说明书和权利要求书及附图中的术语“第一”、“第二”、“第三”、“第四”等（如果存在）是用于区别类似的对象，而不是用于描述特定的顺序或先后次序。

下面结合图 2 示例性的介绍本发明一实施例提供的人工智能(Artificial Intelligence, AI)计算装置的架构示意图。

AI 计算装置 100 可以应用于智能终端，智能终端可以是智能手机、平板电脑、智能汽车、智能家居设备等。AI 计算装置 100 可以包括 AI 集成芯片 10 及片外存储器 20。AI 集成芯片 10 可以包括中央处理单元（Central Processing Unit, CPU）11、神经网络处理单元（Neural-network Processing Unit, NPU）12、第一片内存单元 13 及第二片内存单元 14。

CPU 11 可以包括 NPU 驱动，CPU 11 可以通过 NPU 驱动配置 NPU 12。例如，CPU 11 可以通过 NPU 驱动配置 NPU 12 对指定数据进行处理，为处理指定数据对 NPU 12 内的寄存器进行分配。指定数据可以是指由 CPU 11 所指定的数据。可以由 CPU 11 分配任务，NPU 1 再执行相应任务。例如，以智能汽车为例，摄像头拍摄的每帧图像可以自动存储在某个存储器中，每次有图像存储进来，CPU 11 可以向 NPU 12 下发执行命令，指示 NPU 12 从该存储器中调用该图像进行 AI 模型推理。

NPU 12 为神经网络(neural-network, NN)计算处理器，通过借鉴生物神经网络结构，例如借鉴人脑神经元之间传递模式，对输入信息快速处理，还可以不断的自学习。通过 NPU 12 可以实现智能终端的智能认知等应用，例如：图像识别，人脸识别，语音识别，文本理解等。NPU 12 在数据处理过程中所需的网络代码可以被划分为第一神经网络代码与第二神经网络代码，在数据处理过程中所需的权重数据可以被划分为第一权重参数与第二权重参数。第二神经网络代码可以是指该网络代码中具有高价值的神经网络代码/神经模型代码，第二神经网络代码可以由开发者进行指定。第一神经网络代码可以是指该网络代码中除了第二神经网络代码之外的其余代码。第二权重参数可以是指该权重数据中具有高价值的权重参数，第二权重参数可以由开发者进行指定。第一权重参数可以是指该权重数据中除了第二权重参数之外的其余权重参数。NPU 12 可基于第一神经网络代码、第二神经网络代码、第一权重参数及第二权重参数对指定数据进行处理。

基于硬件成本考量，AI 计算装置 100 所采用的第二片内存单元 14 容量有限，可能无法存储 NPU 12 所需的所有网络代码和权重数据，可以利用片外存储器 20 存储第一神经网络代码、第一权重参数、指定数据及指定数据的处理结果，利用第二片内存单元 14 存储第二神经网络代码及第二权重参数，使得 NPU 12 的核心神经网络代码和核心权重参数存储在 AI 集成芯片 10，可提高核心神经网络代码和核心权重参数的安全等级。

例如，第二片内存单元 14 可以是嵌入式非易失存储器，AI 集成芯片 10 通过增设第二片内存单元 14 来存储相对重要的第二神经网络代码与第二权重参数，将相对不重要的第一神经网络代码与第一权重参数存储在片外存储器 20，可提升第二神经网络代码与第二权重参数的安全等级，由于 NPU 12 可从第二片内存单元读取第二神经网络代码与第二权重参数，可降低 NPU 12 访问片外存储器 20 的带宽需求，提升 NPU 12 性能。若 AI 集成芯片 10 已包含有嵌入式非易失存储器，AI 集成芯片 10 亦可复用该嵌入式非易失存储器来存储相对重要的第二神经网络代码与第二权重参数，无需额外增设一嵌入式非易失存储器作为第二片内存单元 14。

第一片内存单元 13 可用于缓存从片外存储器 20 读取的指定数据，例如当 NPU 12 对

存储在片外存储器 20 中的第一数据进行处理时，该第一数据可以从片外存储器 20 读到第一片内存单元 13，通过第一片内存单元 13 缓存该第一数据，可以加快 NPU 处理数据的速度、提升数据处理效率。

在一些实施例中，第一片内存单元 13 可以包括静态随机访问存储器（Static Random Access Memory, SRAM），第二片内存单元 14 可以包括嵌入式闪存（Embedded Flash, eFlash）、嵌入式非易失性存储器（Embedded Non-volatile Memory, eNVM）、磁性随机访问存储器（Magnetic Random Access Memory, MRAM）或阻性随机访问存储器（Resistive Random Access Memory, RRAM）中的任意一种。第二片内存单元 14 也可以为其他类型的嵌入式存储器件。片外存储器 20 可以包括动态随机访问存储器（Dynamic Random Access Memory, DRAM）。

在一些实施例中，为了提高第二片内存单元 14 所存储的数据私密性，可以将第二片内存单元 14 设计为仅允许 NPU 12 读取的权限，及仅允许 CPU 11 写入的权限。例如，第二片内存单元 14 可以仅设置一个读端口与一个写端口。第二片内存单元 14 的写端口仅电连接于 CPU 11，实现仅允许 CPU 11 写入的权限。第二片内存单元 14 的读端口仅电连接于 NPU 12，实现仅允许 NPU 12 读取的权限。

在一些实施例中，第二片内存单元 14 存储的第二神经网络代码与第二权重参数可以选择性进行密码学加密处理或数学运算加扰处理，以提高第二神经网络代码与第二权重参数的安全性。加密处理或加扰处理的算法可以根据实际需求进行选定，本申请对此不做限定，例如加密算法可以是对称加密算法（AES 算法、DES 算法、SM4 算法等）或非对称加密算法。加扰算法可以是开发者自定义的算法，例如开发者自定义的数学查表、倒序处理等可逆的加扰算法。加密处理或加扰处理可以在 AI 集成芯片 10 外部离线处理，例如可以由 AI 厂商（进行神经网络代码与权重参数开发的厂商）采用加密装置（独立于 AI 集成芯片 10 之外的元件或者装置，如电脑、服务器等设备）对第二神经网络代码与第二权重参数进行加密或加扰处理，得到密文数据，加密装置可以将密文数据发送给 AI 集成芯片 10。密文数据的解密或解扰处理可以在 AI 集成芯片 10 内部进行。

在一些实施例中，如图 3a 与 3b 所示，AI 集成芯片 10 还可以包括密钥存储单元 15、解密单元 16 及第三片内存单元 17。密钥存储单元 15 用于存储解密密钥或解扰密钥。解密单元 16 用于根据密钥存储单元 15 存储的解密密钥或解扰密钥对密文数据进行解密或解扰处理。密钥存储单元 15 可以是设置 AI 集成芯片 10 内部的一次性可编程（One Time Programmable, OTP）器件，解密单元 16 可以是设置在 AI 集成芯片 10 内部的用于对密文数据进行解密或解扰处理的硬件电路。

第三片内存单元 17 用于缓存从片外存储器 20 读取的第一神经网络代码及第一权重参数的至少部分。例如，NPU 12 在需要使用第一神经网络代码及第一权重参数对指定数据进行处理时，存储在片外存储器 20 的部分或者全部的第一神经网络代码及第一权重参数可以从片外存储器 20 读到第三片内存单元 17，通过第三片内存单元 17 缓存第一神经网络代码与第一权重参数，可以加快 NPU 12 读取第一神经网络代码与第一权重参数的速度，提升数据处理效率。第三片内存单元 17 可以是 eFlash、eNVM、MRAM、或 RRAM 中的任意一种。第三片内存单元 17 也可以为其他类型的嵌入式存储器件。

在一些实施例中，若第二片内存单元 14 能够存储 NPU 12 所需的所有网络代码和权重数据，即将第一神经网络代码、第二神经网络代码、第一权重参数及第二权重参数存储

在第二片内存储单元 14，此时片外存储器 20 可用于存储指定数据及指定数据的处理结果，第三片内存储单元 17 亦可被省略（即无需设置第三片内存储单元 17 缓存从片外存储器 20 读取的神经网络代码与权重参数）。

如图 3a 所示，AI 集成芯片 10 可以在密文数据写入第二片内存储单元 14 时对密文数据进行解密或解扰处理，实现将第二神经网络代码与第二权重参数以明文形式存储在第二片内存储单元 14，例如，CPU 11 在接收到密文数据时，可以调用解密单元 16 对密文数据进行解密或解扰处理，再将解密或解扰处理得到的第二神经网络代码与第二权重参数写入至第二片内存储单元 14 进行存储。

如图 3b 所示，AI 集成芯片 10 也可以在密文数据文件写入第二片内存储单元 14 时不进行解密或解扰处理，将第二神经网络代码与第二权重参数以密文形式存储在第二片内存储单元 14，NPU 12 从第二片内存储单元 14 读取密文数据时再对密文数据进行解密或解扰处理。例如，NPU 从第二片内存储单元 14 读取密文数据时，可以调用解密单元 16 对密文数据进行解密或解扰处理，得到第二神经网络代码及第二权重参数。

在一些实施例中，在 AI 集成芯片 10 上电后，解密单元 16 可以自动对存储在密钥存储单元 15 内的解密密钥或解扰密钥进行读取。例如，OTP 器件包括 OTP 存储器与 OTP 控制器，当 AI 集成芯片 10 上电后，OTP 控制器自动控制 OTP 存储器将解密密钥或解扰密钥传送给解密单元 16，使得解密单元 16 可以在 AI 集成芯片 10 上电后自动完成解密密钥或解扰密钥的读取。密钥存储单元 15 可以设置仅允许解密单元 16 读取的权限，避免解密密钥或解扰密钥的泄露。例如密钥存储单元 15 仅设置一个读取端口，该读取端口仅电连接于解密单元 16。

举例而言，第二神经网络代码与第二权重参数以密文形式存储在第二片内存储单元 14，当 NPU 12 从第二片内存储单元 14 读取密文数据时，NPU 12 可以发送调用指令至解密单元 16，解密单元 16 接收到 NPU 12 的调用指令时才使用解密密钥对密文数据进行解密处理，得到第二神经网络代码及第二权重参数。

上述 AI 计算装置，增设第二片内存储单元来存储与 NPU 关联的第二神经网络代码及第二权重参数，将相对重要的第二神经网络代码与第二权重参数存储在第二片内存储单元，将相对不重要的第一神经网络代码与第一权重参数存储在片外存储器，可提升第二神经网络代码与第二权重参数的安全等级，NPU 可从第二片内存储单元读取神经网络代码与权重参数，可降低 NPU 访问片外存储器的带宽需求，同时将第二片内存储单元设置为仅允许 NPU 读取、CPU 写入，可以进一步提升第二神经网络代码与第二权重参数的安全等级，且可以避免设备厂商获取解密密钥、明文第二神经网络代码与第二权重参数。

如图 4 所示，以存储在密钥存储单元 15 内的密钥为解密密钥为例进行说明，不同输入地址的密文数据对应不同的解密密钥，输入的密文数据需要采用与其输入地址关联的解密密钥进行解密。解密单元 16 可以根据密文数据的输入地址，使用解密密钥对密文数据进行解密处理，得到明文数据及输出地址，其中输入地址作为解密处理的指示信息，输入地址与输出地址可以相同。密文数据的输入地址可以是指密文数据的存储地址，例如为存储在密钥存储单元 15 内的地址。

在一些实施例中，若所有密文数据共用一个解密密钥，解密单元 16 可以直接使用解密密钥对密文数据进行解密处理，得到明文数据。

参照图 5 所示，本申请实施例提供的一种神经网络参数部署方法，应用于 AI 计算装置 100。AI 计算装置 100 包括 AI 集成芯片 10 及片外存储器 20。AI 集成芯片 10 包括 CPU 11、

NPU 12、第一片内存单元 13 及第二片内存单元 14。

在一些实施例中，NPU 12 在数据处理过程中所需的网络代码被划分为第一神经网络代码与第二神经网络代码，在数据处理过程中所需的权重数据可以被划分为第一权重参数与第二权重参数。第二神经网络代码可以是指该网络代码中具有高价值的神经网络代码/神经模型代码，第二神经网络代码可以由开发者进行指定。第一神经网络代码可以是指该网络代码中除了第二神经网络代码之外的其余代码。第二权重参数可以是指该权重数据中具有高价值的权重参数，第二权重参数可以由开发者进行指定。第一权重参数可以是指该权重数据中除了第二权重参数之外的其余权重参数。本实施例中，神经网络参数部署方法可以包括：

步骤 50、将第一神经网络代码及第一权重参数存储至片外存储器 20。

在一些实施例中，可以通过 CPU 11 将第一神经网络代码及第一权重参数写入至片外存储器 20 进行存储。片外存储器 20 可以包括 DRAM。

步骤 51，将第二神经网络代码及第二权重参数存储至第二片内存单元 14。

在一些实施例中，可以通过 CPU 11 将第二神经网络代码及第二权重参数写入至第二片内存单元 14 进行存储。第二片内存单元 14 可以是 eFlash、eNVM、MRAM、或 RRAM 中的任意一种，第二片内存单元 14 可以设置有权仅允许 NPU 12 读取的权限，及仅允许 CPU 11 写入的权限，进而可以提高存储在第二片内存单元 14 内的第二神经网络代码与第二权重参数的安全性。

在一些实施例中，AI 集成芯片 10 还可以包括第三片内存单元 17，NPU 12 从片外存储器 20 读取第一神经网络代码与第一权重参数时，读取的第一神经网络代码与第一权重参数可以缓存在第三片内存单元 17。第三片内存单元 17 可以是 eFlash、eNVM、MRAM、或 RRAM 中的任意一种，通过第三片内存单元 17 可以加快 NPU 12 读取第一神经网络代码与第一权重参数的速度，提升数据处理效率。

在一些实施例中，第二片内存单元 14 存储的第二神经网络代码与第二权重参数可以选择性进行密码学加密处理或数学运算加扰处理，以提高第二神经网络代码与第二权重参数的安全性。加密处理或加扰处理的算法可以根据实际需求进行选定，本申请对此不做限定。加密处理或加扰处理可以在 AI 集成芯片 10 外部离线处理，例如可以采用独立于 AI 集成芯片 10 之外的加密装置对第二神经网络代码与第二权重参数进行加密或加扰处理，得到密文数据，加密装置可以将密文数据发送给 AI 集成芯片 10。密文数据的解密或解扰处理可以在 AI 集成芯片 10 内部进行。

AI 集成芯片 10 可以在密文数据写入第二片内存单元 14 时对密文数据进行解密或解扰处理，实现将第二神经网络代码与第二权重参数以明文形式存储在第二片内存单元 14。AI 集成芯片 10 也可以在密文数据文件写入第二片内存单元 14 时不进行解密或解扰处理，将第二神经网络代码与第二权重参数以密文形式存储在第二片内存单元 14，NPU 12 从第二片内存单元 14 读取密文数据时再对密文数据进行解密或解扰处理。

在一些实施例中，若第二片内存单元 14 能够存储 NPU 12 所需的所有网络代码和权重数据，即可以将第一神经网络代码、第二神经网络代码、第一权重参数及第二权重参数存储在第二片内存单元 14。

上述神经网络参数部署方法，增设第二片内存单元来存储与 NPU 关联的第二神经网络代码及第二权重参数，将相对重要的第二神经网络代码与第二权重参数存储在第二片内存

单元，将相对不重要的第一神经网络代码与第一权重参数存储在片外存储器，可提升第二神经网络代码与第二权重参数的安全等级，NPU 可从第二片内存储单元读取神经网络代码与权重参数，可降低 NPU 访问片外存储器的带宽需求，同时将第二片内存储单元设置为仅允许 NPU 读取、CPU 写入，可以进一步提升第二神经网络代码与第二权重参数的安全等级，且可以避免设备厂商获取解密密钥、明文的第二神经网络代码与第二权重参数。

参照图 6 所示，本申请实施例提供的一种神经网络参数部署方法，应用于第一处理端、第二处理端及第三处理端。第一处理端可以是指芯片厂商侧的电子装置，第二处理端可以是指 AI 厂商侧的电子装置，第三处理端可以是设备厂商侧的电子装置。芯片厂商提供 AI 集成芯片 10 的硬件模块，例如硬件模块可以包括 CPU 11、NPU 12、第一片内存储单元 13、第二片内存储单元 14、密钥存储单元 15、解密单元 16 等。AI 厂商提供 AI 集成芯片 10 的软件模块，例如软件模块包括 AI 集成芯片 10 中的 NPU 12 所需的神经网络代码及权重参数。设备厂商基于芯片厂商提供的 AI 集成芯片 10 的硬件模块与 AI 厂商提供 AI 集成芯片 10 的软件模块加工得到包含 AI 集成芯片 10 的终端。本实施例中，神经网络参数部署方法包括：

步骤 61、第一处理端将解密密钥或解扰密钥写入至 AI 集成芯片 10 的密钥存储单元 15。

在一些实施例中，芯片厂商与 AI 厂商可以协商解密密钥或解扰密钥，可以是 AI 厂商自己生成后提供给芯片厂商，也可以是芯片厂商生成后提供给 AI 厂商。密钥存储单元 15 可以是 OTP 器件，芯片厂商可以通过第一处理端将解密密钥或解扰密钥写入至 AI 集成芯片 10 的 OTP 器件。芯片厂商可以将植入有解密密钥或解扰密钥的 AI 集成芯片 10 的硬件模块提供给设备厂商。例如，第一处理端可以为烧录设备，芯片厂商使用烧录设备将解密密钥或解扰密钥写入至 AI 集成芯片 10 的 OTP 器件。

步骤 62、第二处理端对与 NPU 12 关联的神经网络代码及权重参数进行加密，得到密文数据。

在一些实施例中，当 AI 厂商完成与 NPU 12 关联的神经网络代码及权重参数的开发后，AI 厂商对神经网络代码及权重参数进行加密（加密所使用的加密密钥可以与解密密钥相同），得到密文数据。例如，AI 厂商可以使用个人电脑（Personal Computer, PC）或服务器对神经网络代码及权重参数进行加密，得到密文数据。当 AI 厂商得到密文数据后，AI 厂商可以将密文数据通过线上或线下方式提供给设备厂商。比如，设备厂商可以通过线下方式从 AI 厂商购买密文数据，设备厂商也可以于 AI 厂商进行线上交易，如向 AI 厂商的云服务器请求密文数据，AI 厂商的云服务器在确认线上交易后，发送密文数据给设备厂商（设备厂商侧的 PC、或服务器等设备）。

步骤 63、第三处理端将密文数据写入至 AI 集成芯片 10 的第二片内存储单元 14。

在一些实施例中，设备厂商可以通过第三处理端将密文数据写入至 AI 集成芯片 10 的第二片内存储单元 14，进而 AI 集成芯片 10 后续在进行 AI 推理时，解密单元 16 可以使用内置的解密密钥或解扰密钥对密文数据进行解密得到明文的神经网络代码及权重参数。第二片内存储单元 14 可以包括 eFlash、eNVM、MRAM 或 RRAM 中的任意一种。例如，第三处理端可以为烧录设备，设备厂商使用烧录设备将密文数据写入至 AI 集成芯片 10 的第二片内存储单元 14。

在一些实施例中，在 AI 集成芯片 10 上电后，解密单元 16 可以自动对存储在密钥存储单元 15 内的解密密钥或解扰密钥进行读取。例如，OTP 器件包括 OTP 存储器与 OTP 控制器，当 AI 集成芯片 10 上电后，OTP 控制器自动控制 OTP 存储器将解密密钥或解扰密钥传送给解

密单元 16,使得解密单元 16 可以在 AI 集成芯片 10 上电后自动完成解密密钥或解扰密钥的读取。同时,OTP 器件可以设置仅允许解密单元 16 读取的权限,避免解密密钥或解扰密钥的泄露。例如密钥存储单元 15 仅设置一个读取端口,该读取端口电连接于解密单元 16。

在一些实施例中,如果 AI 厂商需要升级神经网络代码及权重参数时,可以通过第二处理端对升级后的神经网络代码与权重参数进行加密,并将新的密文数据提供给设备厂商,设备厂商可以通过第三处理端将新的密文数据写入至 AI 集成芯片 10 的第二片内存单元 14。

上述神经网络参数部署方法,相比现有方案由设备厂商将解密密钥及明文的神经网络代码与权重参数写入至 AI 集成芯片,改由芯片厂商将解密密钥写入至 AI 集成芯片,再将储存有解密密钥的 AI 集成芯片提供给设备厂商,由 AI 厂商对神经网络代码及权重参数进行加密,AI 厂商将神经网络代码及权重参数以密文形式提供给设备厂商,由设备厂商将密文数据写入至第二片内存单元,可以避免设备厂商获取解密密钥及明文的神经网络代码与权重参数,再者将密文数据存储在片内 eFlash、eNVM、MRAM 或 RRAM 中,相比现有方案将神经网络代码与权重参数存储在片外存储器,可以进一步提升神经网络代码与权重参数的安全等级。

参照图 7 所示,本申请实施例提供的一种神经网络参数部署方法,应用于第三处理端。第三处理端可以是设备厂商侧的电子设备。本实施例中,神经网络参数部署方法包括:

步骤 70、获取密文数据。

在一些实施例中,密文数据可以是指对与 NPU 12 关联的神经网络代码及权重参数进行加密得到的数据。例如,可以由 AI 厂商对神经网络代码及权重参数进行加密得到密文数据,并将密文数据提供给设备厂商。

在一些实施例中,当 AI 厂商得到密文数据后,AI 厂商可以将密文数据通过线上或线下方式提供给设备厂商。比如,设备厂商可以通过线下方式从 AI 厂商购买密文数据,设备厂商也可以与 AI 厂商进行线上交易,如向 AI 厂商的云服务器请求密文数据,AI 厂商的云服务器在确认线上交易后,发送密文数据给设备厂商(设备厂商侧的 PC、或服务器等设备)。

步骤 71、将密文数据写入至 AI 集成芯片 10 的第二片内存单元 14。

在一些实施例中,当设备厂商得到密文数据是,设备厂商可以通过第三处理端将密文数据写入至 AI 集成芯片 10 的第二片内存单元 14。第二片内存单元 14 可以包括 eFlash、eNVM、MRAM 或 RRAM 中的任意一种。例如,第三处理端可以为烧录设备,设备厂商使用烧录设备将密文数据写入至 AI 集成芯片 10 的第二片内存单元 14。

在一些实施例中,AI 集成芯片的一次性可编程 OTP 器件中预先写入有对密文数据进行解密的解密密钥。例如,可以由芯片厂商将解密密钥或解扰密钥写入至 AI 集成芯片 10 的 OTP 器件,芯片厂商再将植入有解密密钥或解扰密钥的 AI 集成芯片 10 提供给设备厂商。

在一些实施例中,存储在 OTP 器件内的解密密钥或解扰密钥,可以由设置在 CPU 11 或 NPU 12 内的程序所制定的读取逻辑控制解密单元 16 完成对解密密钥或解扰密钥的读取,且读取操作可以设置为对应用软件不可见。同时,OTP 器件可以设置仅允许解密单元 16 读取的权限,避免解密密钥或解扰密钥的泄露。例如密钥存储单元 15 仅设置一个读取端口,该读取端口电连接于解密单元 16。

上述神经网络参数部署方法,对于设备厂商而言,解密密钥是有芯片厂商写入至 AI 集成芯片的 OTP 器件,再将储存有解密密钥的 AI 集成芯片提供给设备厂商,同时 AI 厂商将神经网络代码及权重参数以密文形式提供给设备厂商,设备厂商无法获取解密密钥及明文

的神经网络代码与权重参数,再者将密文数据存储在片内 eFlash、eNVM、MRAM 或 RRAM 中,相比现有方案将神经网络代码与权重参数存储在片外存储器,可以进一步提升神经网络代码与权重参数的安全等级。

参考图 8,为本申请实施例提供的电子设备 1000 的硬件结构示意图。如图 8 所示,电子设备 1000 可以包括第一处理器 1001、第一存储器 1002 及第一通信总线 1003。第一存储器 1002 用于存储一个或多个计算机程序 1004。一个或多个计算机程序 1004 被配置为被该第一处理器 1001 执行。该一个或多个计算机程序 1004 包括指令,上述指令可以用于实现在电子设备 1000 中执行如图 5 所述的神经网络参数部署方法。

可以理解的是,本实施例示意的结构并不构成对电子设备 1000 的具体限定。在另一些实施例中,电子设备 1000 可以包括比图示更多或更少的部件,或者组合某些部件,或者拆分某些部件,或者不同的部件布置。

第一处理器 1001 可以包括一个或多个处理单元,例如:第一处理器 1001 可以包括应用处理器(application processor, AP),调制解调器,图形处理器(graphics processing unit, GPU),图像信号处理器(image signal processor, ISP),控制器,视频编解码器,数字信号处理器(digital signal processor, DSP),基带处理器,和/或神经网络处理器(neural-network processing unit, NPU)等。其中,不同的处理单元可以是独立的器件,也可以集成在一个或多个处理器中。

第一处理器 1001 还可以设置有存储器,用于存储指令和数据。在一些实施例中,第一处理器 1001 中的存储器为高速缓冲存储器。该存储器可以保存第一处理器 1001 刚用过或循环使用的指令或数据。如果第一处理器 1001 需要再次使用该指令或数据,可从该存储器中直接调用。避免了重复存取,减少了第一处理器 1001 的等待时间,因而提高了系统的效率。

在一些实施例中,第一处理器 1001 可以包括一个或多个接口。接口可以包括集成电路(inter-integrated circuit, I2C)接口,集成电路内置音频(inter-integrated circuit sound, I2S)接口,脉冲编码调制(pulse code modulation, PCM)接口,通用异步收发传输器(universal asynchronous receiver/transmitter, UART)接口,移动产业处理器接口(mobile industry processor interface, MIPI),通用输入输出(general-purpose input/output, GPIO)接口, SIM 接口,和/或 USB 接口等。

在一些实施例中,第一存储器 1002 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,例如硬盘、内存、插接式硬盘,智能存储卡(Smart Media Card, SMC),安全数字(Secure Digital, SD)卡,闪存卡(Flash Card)、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。

本实施例还提供一种计算机存储介质,该计算机存储介质中存储有计算机指令,当该计算机指令在电子设备上运行时,使得电子设备执行上述相关方法步骤实现如图 5 所述的神经网络参数部署方法。

本实施例还提供了一种计算机程序产品,当该计算机程序产品在计算机上运行时,使得计算机执行上述相关步骤,以实现如图 5 所述的神经网络参数部署方法。

其中,本实施例提供的电子设备、计算机存储介质或计算机程序产品均用于执行上文所提供的对应的方法,因此,其所能达到的有益效果可参考上文所提供的对应的方法中的有益效果,此处不再赘述。

通过以上的实施方式的描述,所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,仅以上述各功能模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能

分配由不同的功能模块完成，即将装置的内部结构划分成不同的功能模块，以完成以上描述的全部或者部分功能。

在本申请所提供的几个实施例中，应该理解到，所揭露的装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例是示意性的，例如，该模块或单元的划分，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装置，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。

该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一个物理单元或多个物理单元，即可以位于一个地方，或者也可以分布到多个不同地方。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。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

该集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用，可以存储在一个可读取存储介质中。基于这样的理解，本申请实施例的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该软件产品存储在一个存储介质中，包括若干指令用以使得一个设备（可以是单片机，芯片等）或处理器（processor）执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U盘、移动硬盘、只读存储器（Read-Only Memory, ROM）、随机存取存储器（Random Access Memory, RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

以上所述，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何在本申请揭露的技术范围内的变化或替换，都应涵盖在本申请的保护范围之内。

权 利 要 求 书

1. 一种人工智能 AI 计算装置，包括 AI 集成芯片及片外存储器，所述 AI 集成芯片包括中央处理单元 CPU、神经网络处理单元 NPU 及第一片内存储单元，其特征在于，所述片外存储器用于存储与所述 NPU 关联的第一神经网络代码及第一权重参数；

所述 AI 集成芯片还包括第二片内存储单元，所述第二片内存储单元用于存储与所述 NPU 关联的第二神经网络代码及第二权重参数，所述 NPU 用于基于所述第一神经网络代码、所述第二神经网络代码、所述第一权重参数及所述第二权重参数对指定数据进行处理；

其中，所述第二片内存储单元设置有仅允许所述 NPU 读取的权限，及仅允许所述 CPU 写入的权限。

2. 如权利要求 1 所述的 AI 计算装置，其特征在于，所述 AI 集成芯片还包括第三片内存储单元，所述第三片内存储单元用于缓存从所述片外存储器读取的所述第一神经网络代码及所述第一权重参数的至少部分。

3. 如权利要求 1 或 2 所述的 AI 计算装置，其特征在于，所述 CPU 还用于接收密文数据，并将所述密文数据写入至所述第二片内存储单元，所述密文数据为对所述第二神经网络代码及所述第二权重参数进行加密得到的。

4. 如权利要求 3 所述的 AI 计算装置，其特征在于，所述 AI 集成芯片还包括密钥存储单元及解密单元，解密密钥在所述密文数据写入至所述第二片内存储单元之前存储至所述密钥存储单元，所述密钥存储单元设置有仅允许所述解密单元读取的权限，所述 NPU 还用于从所述第二片内存储单元读取所述密文数据，并调用所述解密单元使用所述解密密钥对所述密文数据进行解密，得到所述第二神经网络代码及所述第二权重参数。

5. 如权利要求 1 或 2 所述的 AI 计算装置，其特征在于，所述 AI 集成芯片还包括密钥存储单元及解密单元，所述密钥存储单元用于存储解密密钥，所述密钥存储单元设置有仅允许所述解密单元读取的权限，所述 CPU 用于接收密文数据，并调用所述解密单元使用所述解密密钥对所述密文数据进行解密，所述密文数据为对所述第二神经网络代码及所述第二权重参数进行加密得到的，所述 CPU 还用于将解密得到的第二神经网络代码及第二权重参数写入至所述第二片内存储单元。

6. 如权利要求 4 或 5 所述的 AI 计算装置，其特征在于，所述密钥存储单元包括一次性可编程 OTP 器件。

7. 如权利要求 1 至 6 任一项所述的 AI 计算装置，其特征在于，所述片外存储器还用于存储所述指定数据及所述指定数据的处理结果，所述第一片内存储单元用于缓存从所述片外存储器读取的所述指定数据。

8. 如权利要求 1 至 7 中任意一项所述的 AI 计算装置，其特征在于，所述第一片内存储单元包括静态随机访问存储器 SRAM，所述片外存储器包括动态随机访问存储器 DRAM，所述第二片内存储单元包括嵌入式闪存 eFalsh、嵌入式非易失性存储器 eNVM、磁性随机访问存储器 MRAM 或阻性随机访问存储器 RRAM 中的任意一种，所述第三片内存储单元包括 eFalsh、eNVM、MRAM 或 RRAM 中的任意一种。

9. 一种人工智能 AI 集成芯片，包括中央处理单元 CPU、神经网络处理单元 NPU 及第一片内存储单元，其特征在于，所述 AI 集成芯片还包括第二片内存储单元，所述第二片内存储

单元用于存储与所述 NPU 关联的神经网络代码及权重参数，所述 NPU 用于基于所述神经网络代码及所述权重参数对指定数据进行处理；

其中，所述第二片内存储单元包括嵌入式闪存 eFlash、嵌入式非易失性存储器 eNVM、磁性随机访问存储器 MRAM 或阻性随机访问存储器 RRAM 中的任意一种。

10. 如权利要求 9 所述的 AI 集成芯片，其特征在于，所述第二片内存储单元仅设置一个读端口与一个写端口，所述读端口电连接于所述 NPU，所述写端口电连接于所述 CPU。

11. 一种神经网络参数部署方法，应用于人工智能 AI 计算装置，其特征在于，所述 AI 计算装置包括 AI 集成芯片及片外存储器，所述 AI 集成芯片包括中央处理单元 CPU、神经网络处理单元 NPU、第一片内存储单元及第二片内存储单元，所述神经网络参数部署包括：

将与所述 NPU 关联的第一神经网络代码、第一权重参数存储至所述片外存储器；

将与所述 NPU 关联的第二神经网络代码及第二权重参数存储至所述第二片内存储单元；

其中，所述 NPU 用于基于所述第一神经网络代码、所述第二神经网络代码、所述第一权重参数及所述第二权重参数对指定数据进行处理，所述第二片内存储单元设置有仅允许所述 NPU 读取的权限，及仅允许所述 CPU 写入的权限。

12. 如权利要求 11 所述的神经网络参数部署方法，其特征在于，所述 AI 集成芯片还包括第三片内存储单元，所述神经网络参数部署方法还包括：

从所述片外存储器读取所述第一神经网络代码及所述第一权重参数后缓存在所述第三片内存储单元中。

13. 如权利要求 11 或 12 所述的神经网络参数部署方法，其特征在于，所述神经网络参数部署方法还包括：

接收加密装置发送的密文数据，并将所述密文数据写入至所述第二片内存储单元；

其中，所述密文数据为对所述第二神经网络代码及所述第二权重参数进行加密得到的。

14. 如权利要求 13 所述的神经网络参数部署方法，其特征在于，所述神经网络参数部署方法还包括：

当所述 NPU 从所述第二片内存储单元读取所述密文数据时，调用预先存储的解密密钥对所述密文数据进行解密，得到所述第二神经网络代码及所述第二权重参数。

15. 如权利要求 11 或 12 所述的神经网络参数部署方法，其特征在于，所述神经网络参数部署方法还包括：

接收加密装置发送的密文数据，并调用预先存储的解密密钥对所述密文数据进行解密；

将解密得到的第二神经网络代码及第二权重参数写入至所述第二片内存储单元；

其中，所述密文数据为对所述第二神经网络代码及所述第二权重参数进行加密得到的。

16. 一种神经网络参数部署方法，其特征在于，包括：

第一处理端将解密密钥写入至人工智能 AI 集成芯片的一次性可编程 OTP 器件；

第二处理端对神经网络代码及权重参数进行加密，得到密文数据；

第三处理端将所述密文数据写入至所述 AI 集成芯片的片内存储单元；

其中，所述解密密钥用于对所述密文数据进行解密，所述片内存储单元包括嵌入式闪存 eFlash、嵌入式非易失性存储器 eNVM、磁性随机访问存储器 MRAM 或阻性随机访问存储器 RRAM 中的任意一种。

17. 一种神经网络参数部署方法，其特征在于，包括：

获取密文数据，所述密文数据为对神经网络代码及权重参数进行加密得到的；

将所述密文数据写入至人工智能 AI 集成芯片的片内存储单元；

其中，所述 AI 集成芯片的一次性可编程 OTP 器件中预先写入有对所述密文数据进行解密的解密密钥，所述片内存储单元包括嵌入式闪存 eFlash、嵌入式非易失性存储器 eNVM、磁性随机访问存储器 MRAM 或阻性随机访问存储器 RRAM 中的任意一种。

18. 一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储计算机指令，当所述计算机指令在计算机或处理器上运行时，使得所述计算机或处理器执行如权利要求 11 至权利要求 15 中任一项所述的神经网络参数部署方法。

19. 一种电子设备，其特征在于，所述电子设备包括如权利要求 1 至 8 中任意一项所述的人工智能 AI 计算装置，或者如权利要求 9 或 10 所述的 AI 集成芯片。

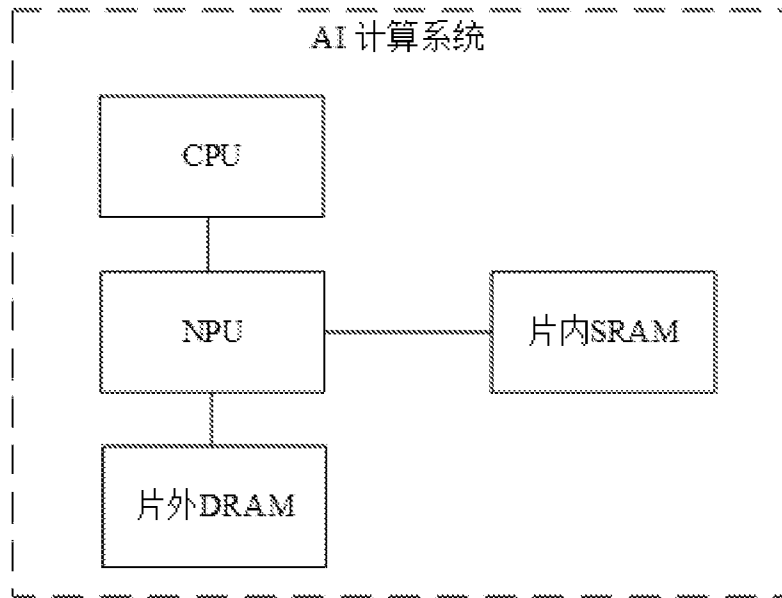


图 1

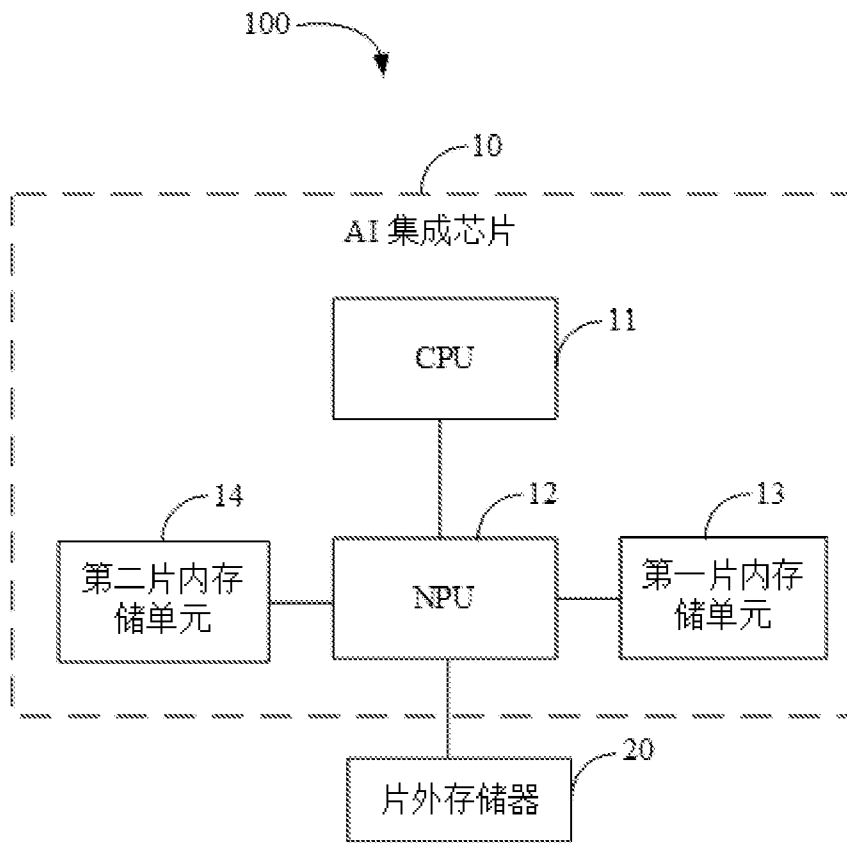


图 2

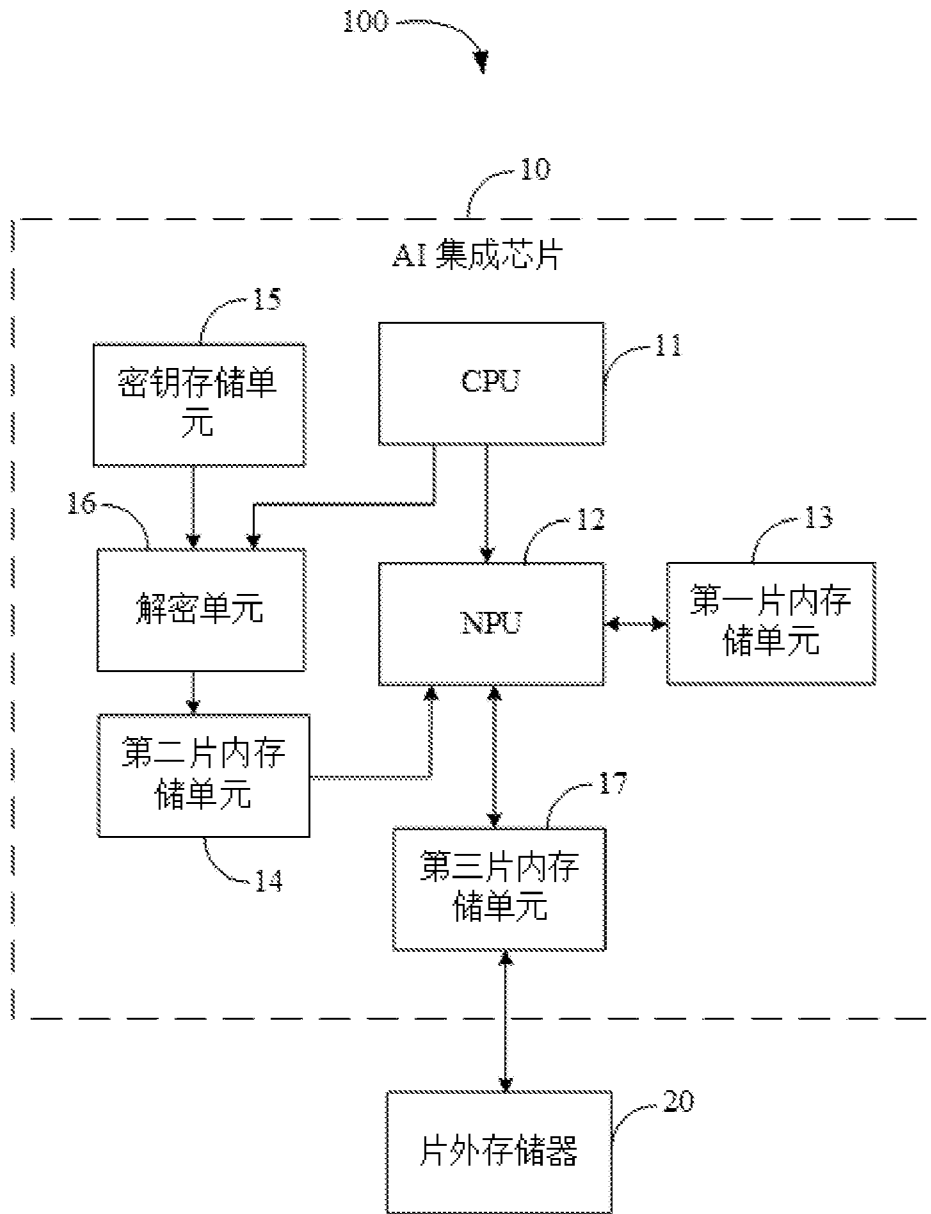


图 3a

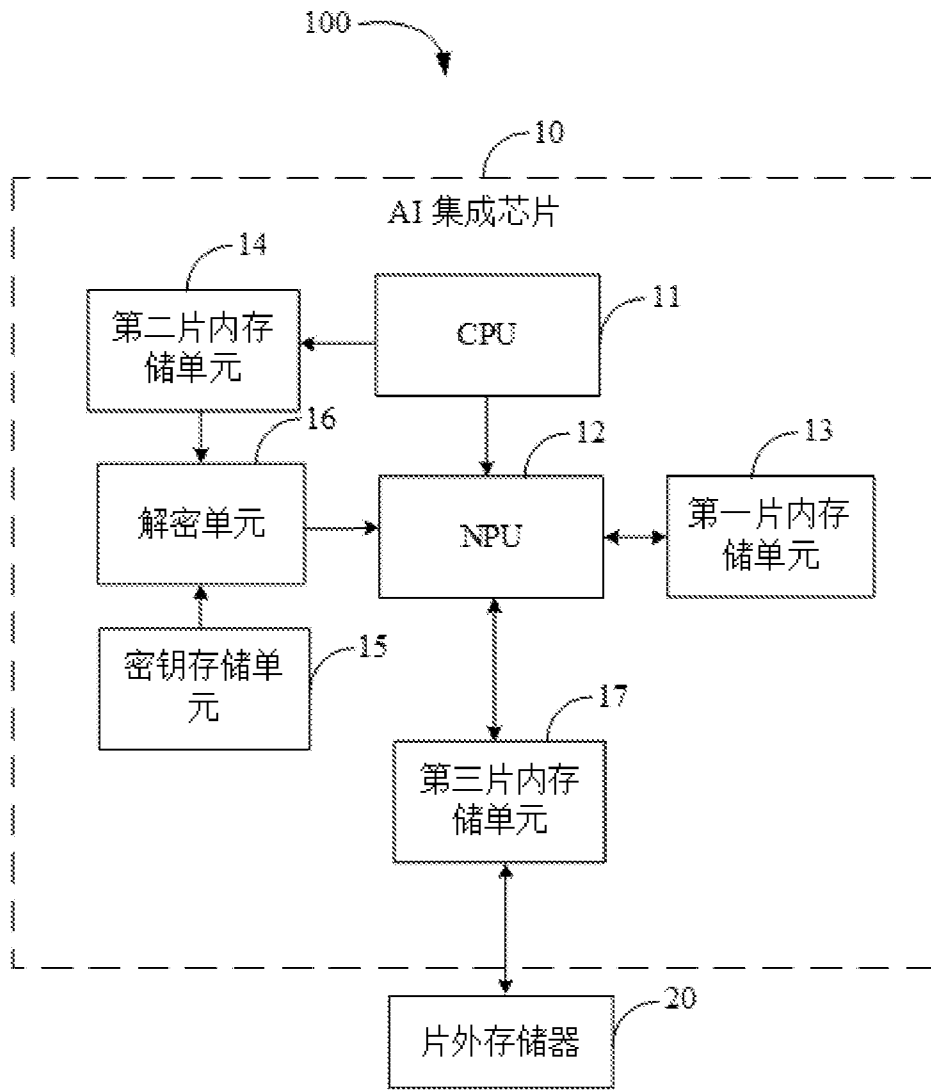


图 3b

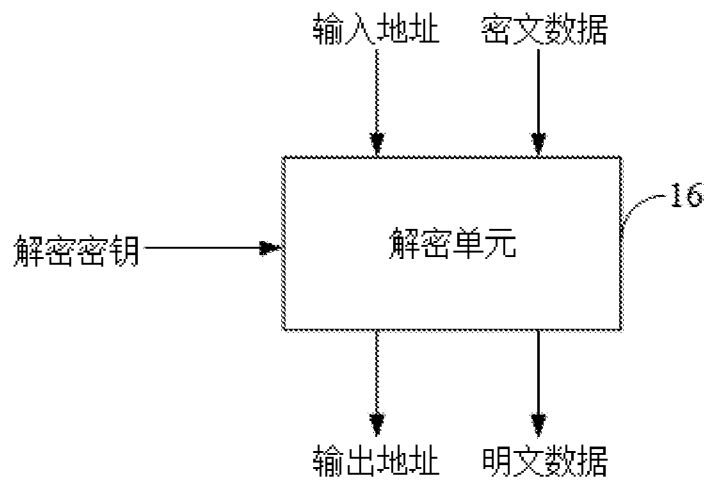


图 4

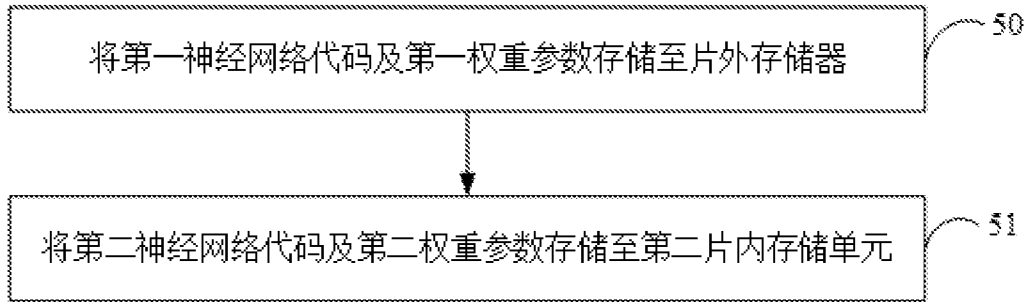


图 5

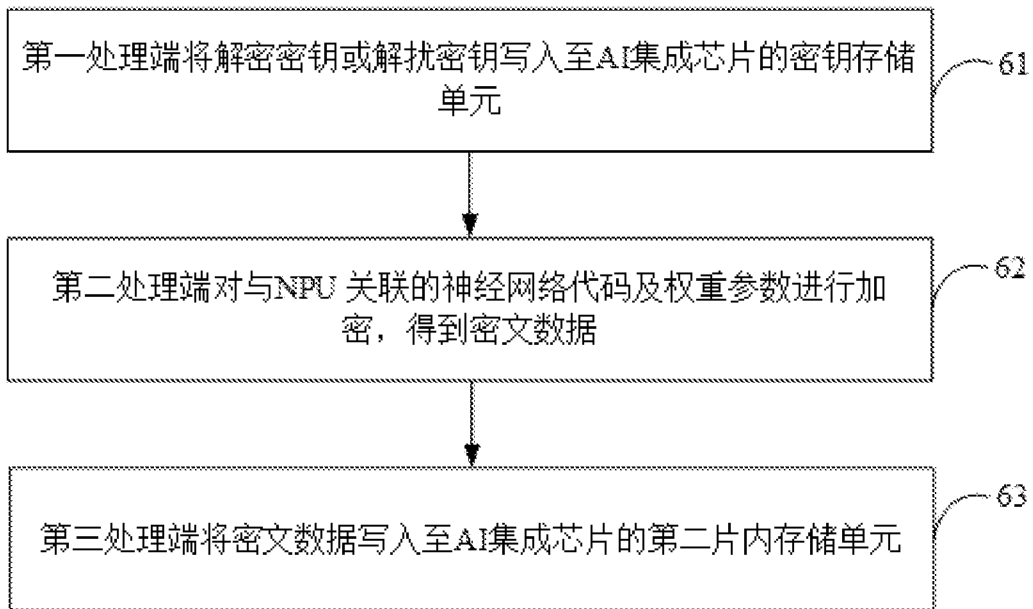


图 6

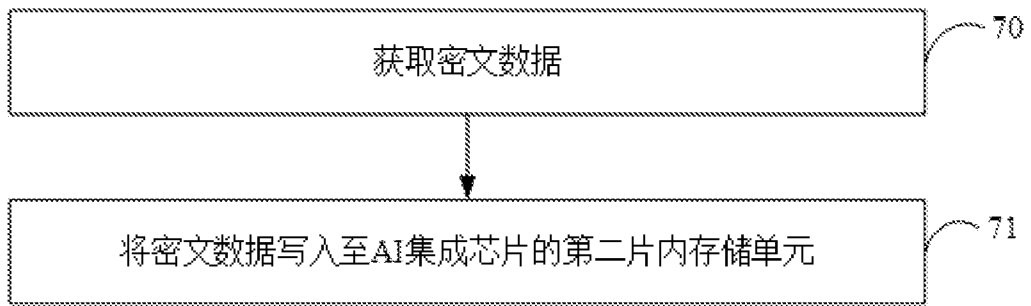


图 7

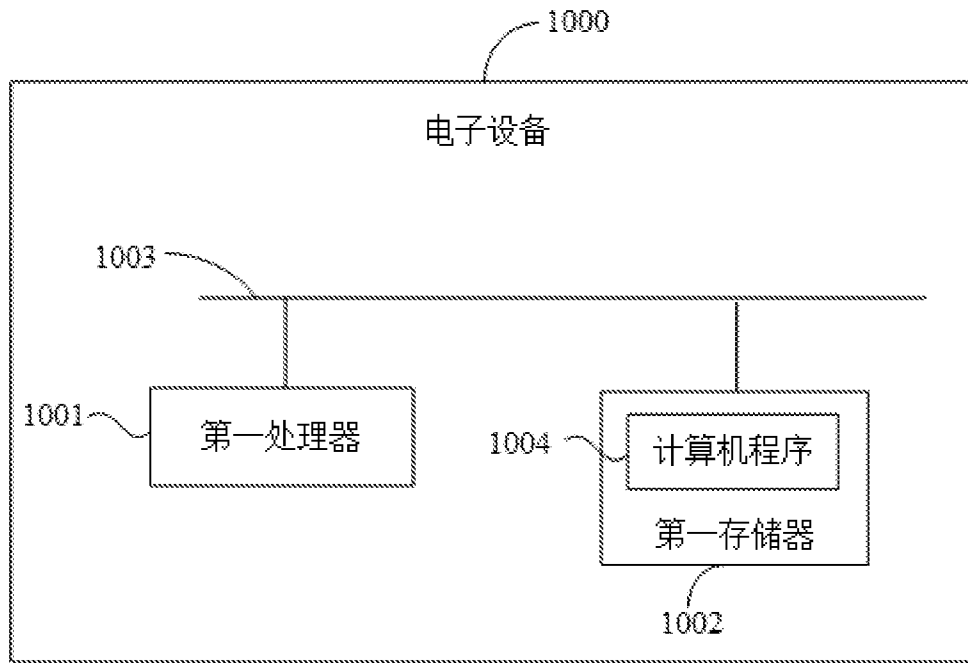


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/094495

A. CLASSIFICATION OF SUBJECT MATTER		
G06N 3/02(2006.01)i; G06N 3/063(2006.01)i; H04L 9/08(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; H04L		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) CNABS, CNTXT, ENTXTC, CNKI, DWPI, ENTXT: 神经网络, 权重, 参数, 保护, 安全, 片内, 片上, 片外, 加密, 部分; neural network, weight, parameter, protect, security, on-chip, off-chip, encrypt, part, portion.		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 107086910 A (INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) 22 August 2017 (2017-08-22) description, paragraphs 0041-0110 and 0127	9, 10, 16, 17, 19
Y	CN 107086910 A (INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) 22 August 2017 (2017-08-22) description, paragraphs 0041-0110 and 0127	1-8, 11-15, 18
Y	CN 110601814 A (SHENZHEN QIANHAI WEBANK CO., LTD.) 20 December 2019 (2019-12-20) description, paragraphs 0048-0058	1-8, 11-15, 18
A	CN 107885509 A (HANGZHOU NATIONALCHIP SCIENCE & TECHNOLOGY CO., LTD.) 06 April 2018 (2018-04-06) entire document	1-19
A	CN 113127407 A (NANJING YOUCUN TECHNOLOGY CO., LTD.) 16 July 2021 (2021-07-16) entire document	1-19
A	US 10956584 B1 (AMAZON TECHNOLOGIES, INC.) 23 March 2021 (2021-03-23) entire document	1-19
<input 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 11 July 2022		Date of mailing of the international search report 28 July 2022
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/CN2022/094495

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107086910	A	22 August 2017	WO	2018171663	A1	27 September 2018
				US	2020019843	A1	16 January 2020
				CN	107086910	B	10 August 2018
<hr/>							
CN	110601814	A	20 December 2019	WO	2021056760	A1	01 April 2021
				CN	110601814	B	27 August 2021
<hr/>							
CN	107885509	A	06 April 2018	None			
<hr/>							
CN	113127407	A	16 July 2021	CN	214846708	U	23 November 2021
<hr/>							
US	10956584	B1	23 March 2021	None			
<hr/>							

国际检索报告

国际申请号

PCT/CN2022/094495

<p>A. 主题的分类</p> <p>G06N 3/02(2006.01)i; G06N 3/063(2006.01)i; H04L 9/08(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. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>G06N; H04L</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))</p> <p>CNABS, CNTXT, ENTXT, CNKI, DWPI, ENTXT: 神经网络, 权重, 参数, 保护, 安全, 片内, 片上, 片外, 加密, 部分; neural network, weight, parameter, protect, security, on-chip, off-chip, encrypt, part, portion.</p>																							
<p>C. 相关文件</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>X</td> <td>CN 107086910 A (中国科学院计算技术研究所) 2017年8月22日 (2017 - 08 - 22) 说明书0041-0110、0127段</td> <td>9, 10, 16, 17, 19</td> </tr> <tr> <td>Y</td> <td>CN 107086910 A (中国科学院计算技术研究所) 2017年8月22日 (2017 - 08 - 22) 说明书0041-0110、0127段</td> <td>1-8, 11-15, 18</td> </tr> <tr> <td>Y</td> <td>CN 110601814 A (深圳前海微众银行股份有限公司) 2019年12月20日 (2019 - 12 - 20) 说明书0048-0058段</td> <td>1-8, 11-15, 18</td> </tr> <tr> <td>A</td> <td>CN 107885509 A (杭州国芯科技股份有限公司) 2018年4月6日 (2018 - 04 - 06) 全文</td> <td>1-19</td> </tr> <tr> <td>A</td> <td>CN 113127407 A (南京优存科技有限公司) 2021年7月16日 (2021 - 07 - 16) 全文</td> <td>1-19</td> </tr> <tr> <td>A</td> <td>US 10956584 B1 (AMAZON TECHNOLOGIES, INC.) 2021年3月23日 (2021 - 03 - 23) 全文</td> <td>1-19</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 107086910 A (中国科学院计算技术研究所) 2017年8月22日 (2017 - 08 - 22) 说明书0041-0110、0127段	9, 10, 16, 17, 19	Y	CN 107086910 A (中国科学院计算技术研究所) 2017年8月22日 (2017 - 08 - 22) 说明书0041-0110、0127段	1-8, 11-15, 18	Y	CN 110601814 A (深圳前海微众银行股份有限公司) 2019年12月20日 (2019 - 12 - 20) 说明书0048-0058段	1-8, 11-15, 18	A	CN 107885509 A (杭州国芯科技股份有限公司) 2018年4月6日 (2018 - 04 - 06) 全文	1-19	A	CN 113127407 A (南京优存科技有限公司) 2021年7月16日 (2021 - 07 - 16) 全文	1-19	A	US 10956584 B1 (AMAZON TECHNOLOGIES, INC.) 2021年3月23日 (2021 - 03 - 23) 全文	1-19
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
X	CN 107086910 A (中国科学院计算技术研究所) 2017年8月22日 (2017 - 08 - 22) 说明书0041-0110、0127段	9, 10, 16, 17, 19																					
Y	CN 107086910 A (中国科学院计算技术研究所) 2017年8月22日 (2017 - 08 - 22) 说明书0041-0110、0127段	1-8, 11-15, 18																					
Y	CN 110601814 A (深圳前海微众银行股份有限公司) 2019年12月20日 (2019 - 12 - 20) 说明书0048-0058段	1-8, 11-15, 18																					
A	CN 107885509 A (杭州国芯科技股份有限公司) 2018年4月6日 (2018 - 04 - 06) 全文	1-19																					
A	CN 113127407 A (南京优存科技有限公司) 2021年7月16日 (2021 - 07 - 16) 全文	1-19																					
A	US 10956584 B1 (AMAZON TECHNOLOGIES, INC.) 2021年3月23日 (2021 - 03 - 23) 全文	1-19																					
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><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>“&” 同族专利的文件</p>																							
<p>国际检索实际完成的日期</p> <p>2022年7月11日</p>		<p>国际检索报告邮寄日期</p> <p>2022年7月28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>张会</p> <p>电话号码 62412328</p>																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/094495

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	107086910	A	2017年8月22日	WO	2018171663	A1	2018年9月27日
				US	2020019843	A1	2020年1月16日
				CN	107086910	B	2018年8月10日
CN	110601814	A	2019年12月20日	WO	2021056760	A1	2021年4月1日
				CN	110601814	B	2021年8月27日
CN	107885509	A	2018年4月6日	无			
CN	113127407	A	2021年7月16日	CN	214846708	U	2021年11月23日
US	10956584	B1	2021年3月23日	无			