

图像篡改检测

技术领域

[01]本说明书涉及计算机技术领域，尤其涉及一种图像篡改检测方法、装置、存储介质以及电子设备。

背景技术

[02]消费金融场景中，金融消费者和金融机构间涉及到大量带有用户身份信息的金融凭证，如薪水证明、职业凭证和身份证件等。通常，这些材料多以图像形式由用户上传至金融机构平台中，待金融机构平台完成材料验真及信息审核后，即可办理相关事务。

发明内容

[03]本说明书提供了一种图像篡改检测方法、装置、存储介质以及电子设备。所述技术方案如下。

[04]第一方面，本说明书提供了一种图像篡改检测方法，应用于电子设备，所述方法包括：确定本地图像篡改检测模型对应的第一模型梯度矩阵，基于所述第一模型梯度矩阵进行局部梯度差分隐私计算处理得到第二模型梯度矩阵；将所述第二模型梯度矩阵上传至服务平台，所述第二模型梯度矩阵用于指示所述服务平台基于至少一个所述电子设备上传的所述第二模型梯度矩阵确定目标模型参数；接收所述服务平台发送的所述目标模型参数，基于所述目标模型参数更新所述本地图像篡改检测模型得到目标图像篡改检测模型；基于所述目标图像篡改检测模型对目标隐私图像进行图像检测处理得到所述目标隐私图像对应的目标热力检测图，对所述目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果。

[05]第二方面，本说明书提供了一种图像篡改检测方法，应用于服务平台，所述方法包括：接收至少一个电子设备上传的第二模型梯度矩阵，所述第二模型梯度矩阵由所述电子设备基于本地图像篡改检测模型对应的第一模型梯度矩阵进行局部梯度差分隐私计算处理得到；基于各所述第二模型梯度矩阵确定目标模型参数；将所述目标模型参数发送至各所述电子设备，所述目标模型参数用于指示所述电子设备基于所述目标模型参数更新本地图像篡改检测模型得到目标图像篡改检测模型，并基于所述目标图像篡改检测模型对目标隐私图像进行图像检测处理得到所述目标隐私图像对应的目标热力检测图，对所述目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果。

[06]第三方面，本说明书提供了一种图像篡改检测装置，应用于电子设备，所述装置包括：梯度处理模块，用于确定本地图像篡改检测模型对应的第一模型梯度矩阵，基于所述第一模型梯度矩阵进行局部梯度差分隐私计算处理得到第二模型梯度矩阵；梯度发送模块，用于将所述第二模型梯度矩阵上传至服务平台，所述第二模型梯度矩阵用于指示所述服务平台基于至少一个所述电子设备上传的所述第二模型梯度矩阵确定目标模型参数；数据接收模块，用于接收所述服务平台发送的所述目标模型参数，基于所述目标模型参数更新所述本地图像篡改检测模型得到目标图像篡改检测模型；图像处理模块，用于基于所述目标图像篡改检测模型对目标隐私图像进行图像检测处理得到所述目标隐私图像对应的目标热力检测图，对所述目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果。

[07]第四方面，本说明书提供了一种图像篡改检测装置，应用于服务平台，所述装置包括：数据接收模块，用于接收至少一个电子设备上传的第二模型梯度矩阵，所述第二模型梯度矩阵由所述电子设备基于本地图像篡改检测模型对应的第一模型梯度矩阵进行局部梯度差分隐私计算处理得到；数据处理模块，用于基于各所述第二模型梯度矩阵确定目标模型参数；数据发送模块，用于将所述目标模型参数发送至各所述电子设备，所述目标模型参数用于指示所述电子设备基于所述目标模型参数更新本地图像篡改检测模型得到目标图像篡改检测模型，并基于所述目标图像篡改检测模型对目标隐私图像进行图像检测处理得到所述目标隐私图像对应的目标热力检测图，对所述目标热力检测图

进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果。

[08]第五方面,本说明书提供了一种计算机存储介质,所述计算机存储介质有多条指令,所述指令适于由处理器加载并执行上述的方法步骤。

[09]第六方面,本说明书提供了一种计算机程序产品,所述计算机程序产品存储有至少一条指令,所述至少一条指令由处理器加载并执行上述的方法步骤。

[10]第七方面,本说明书提供了一种电子设备,可包括:存储器和处理器;其中,所述存储器存储有计算机程序,所述计算机程序适于由所述存储器加载并执行上述的方法步骤。

[11]本说明书提供的技术方案带来的有益效果至少包括:本说明书实施例中,由电子设备确定本地图像篡改检测模型对应的第一模型梯度矩阵,然后根据第一模型梯度矩阵进行局部梯度差分隐私计算处理得到第二模型梯度矩阵,再由电子设备将第二模型梯度矩阵上传至服务平台,第二模型梯度矩阵用于服务平台根据至少一个电子设备上传的第二模型梯度矩阵来确定目标模型参数,之后电子设备接收服务平台发送的目标模型参数,电子设备根据目标模型参数更新本地图像篡改检测模型得到目标图像篡改检测模型,而后电子设备根据目标图像篡改检测模型对目标隐私图像进行图像检测处理得到目标隐私图像对应的目标热力检测图,再对目标热力检测图进行篡改识别处理得到目标隐私图像对应的篡改识别结果。采用上述方法步骤,电子设备通过对即将上传至服务平台的梯度数据进行前处理,以提升上传的梯度数据的抗逆向攻击能力,由于电子设备对隐私图像进行图像检测处理所采用的目标图像篡改检测模型,是根据服务平台结合多个电子设备的梯度数据加工后得到的模型参数所得到的模型,该模型在训练时结合了多方数据,因此该模型对隐私图像的识别准确率得到提升,因而再对该模型的识别结果进行篡改识别处理,提升了隐私图像的篡改识别结果的准确率。

附图说明

[12]为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[13]图1是本说明书实施例提供的一种图像篡改检测方法的系统架构示意图。

[14]图2是本说明书实施例提供的一种图像篡改检测方法的流程示意图。

[15]图3是本说明书实施例提供的又一种图像篡改检测方法的流程示意图。

[16]图4是本说明书实施例提供的一种目标热力检测图的示意图。

[17]图5是本说明书实施例提供的一种图像篡改检测方法的流程示意图。

[18]图6是本说明书实施例提供的又一种图像篡改检测方法的流程示意图。

[19]图7是本说明书实施例提供的一种图像篡改检测装置的结构示意图。

[20]图8是本说明书实施例提供的一种图像篡改检测装置的结构示意图。

[21]图9是本说明书实施例提供的一种电子设备的结构示意图。

[22]图10是本说明书实施例提供的一种服务平台的结构示意图。

具体实施方式

[23]为使得本说明书实施例的发明目的、特征、优点能够更加的明显和易懂,下面将结合本说明书实施例中的附图,对本说明书实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本说明书一部分实施例,而非全部实施例。基于本说明书中的实施例,本领域技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本说明书保护的范围。

[24]在本说明书的描述中,需要理解的是,术语“第一”、“第二”等仅用于描述目的,而不能理解为指示或暗示相对重要性。在本说明书的描述中,需要说明的是,除非另有明确的规定和限定,“包括”和“具有”以及它们任何变形,意图在于覆盖不排他的包含。例

如包含了一系列步骤或单元的过程、方法、系统、产品或设备没有限定于已列出的步骤或单元，而是可选地还包括没有列出的步骤或单元，或可选地还包括对于这些过程、方法、产品或设备固有的其他步骤或单元。对于本领域的普通技术人员而言，可以具体情况理解上述术语在本说明书中的具体含义。此外，在本说明书的描述中，除非另有说明，“多个”是指两个或两个以上。“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 这三种情况。字符“/”一般表示前后关联对象是一种“或”的关系。

[25]在消费金融场景中，需要金融机构对用户上传的材料，如带有用户身份信息的金融凭证进行审核，并在审核通过为用户办理相关事务。金融机构对用户上传的材料进行篡改检测是金融凭证审核的关键。相关技术中，金融机构依靠自身训练得到的篡改检测模型对用户上传的材料进行篡改检测，而单个金融机构拥有的数据有限，且其拥有的数据也不允许流出，单个金融机构可利用的训练篡改检测模型的训练数据则较少，就导致金融机构的篡改检测模型存在篡改识别效果较差的问题。

[26]为了解决上述技术问题，下面结合具体的实施例对本说明书进行详细说明。

[27]请参见图 1，为本说明书实施例提供的一种图像篡改检测方法的系统架构示意图。

[28]如图 1 所示，图 1 包括服务平台和电子设备，图 1 所示电子设备的数量和服务平台中的节点的数量仅为示例性的，本说明书实施例并不限制它们的数量。

[29]在本说明书实施例中，服务平台可指由一个节点构成的服务器或多个节点构成的服务器集群。

[30]当服务平台为多个节点构成的服务器集群时，服务平台中的每个节点可以是单独的服务器设备，例如：机架式、刀片、塔式、或者机柜式的服务器设备，或采用工作站、大型计算机等具备较强计算能力硬件设备；也可以是采用多个服务器组成的服务器集群，服务集群中的各服务器可以是对称方式组成的，其中每台服务器在事务链路中功能等价、地位等价，各服务器均可单独对外提供服务，单独对外提供服务可以理解为无需另外的服务器的辅助。

[31]在本说明书实施例中，电子设备可以是具有对隐私图像进行图像检测处理和篡改识别处理等功能的计算机设备。不同电子设备为不同消费金融机构的设备。

[32]需要说明的是，电子设备与服务平台中的至少一个节点通过网络建立通信连接进行交互通信。其中，网络可以是无线网络，也可以是有线网络，无线网络包括但不限于蜂窝网络、无线局域网、红外网络或蓝牙网络，有线网络包括但不限于以太网、通用串行总线（universal serial bus，USB）或控制器局域网。在说明书一个或多个实施例中，使用包括超文本标记语言（Hyper Text Mark-up Language，HTML）、可扩展标记语言（Extensible Markup Language，XML）等的技术和/或格式来代表通过网络交换的数据（如目标压缩包）。此外还可以使用诸如安全套接字层（Secure Socket Layer，SSL）、传输层安全（Transport Layer Security，TLS）、虚拟专用网络（Virtual Private Network，VPN）、网际协议安全（Internet Protocol Security，IPsec）等常规加密技术来加密所有或者一些链路。在另一些实施例中，还可以使用定制和/或专用数据通信技术取代或者补充上述数据通信技术。

[33]本说明书实施例中，由电子设备确定本地图像篡改检测模型对应的第一模型梯度矩阵，然后根据第一模型梯度矩阵进行局部梯度差分隐私计算处理得到第二模型梯度矩阵，再由电子设备将第二模型梯度矩阵上传至服务平台，第二模型梯度矩阵用于服务平台根据至少一个电子设备上传的第二模型梯度矩阵来确定目标模型参数，之后电子设备接收服务平台发送的目标模型参数，电子设备根据目标模型参数更新本地图像篡改检测模型得到目标图像篡改检测模型，而后电子设备根据目标图像篡改检测模型对目标隐私图像进行图像检测处理得到目标隐私图像对应的目标热力检测图，再对目标热力检测图进行篡改识别处理得到目标隐私图像对应的篡改识别结果。采用上述方法步骤，电子设备通

过对即将上传至服务平台的梯度数据进行前处理,以提升上传的梯度数据的抗逆向攻击能力,由于电子设备对隐私图像进行图像检测处理所采用的目标图像篡改检测模型,是根据服务平台结合多个电子设备的梯度数据加工后得到的模型参数所得到的模型,该模型在训练时结合了多方数据,因此该模型对隐私图像的识别准确率得到提升,因而再对该模型的识别结果进行篡改识别处理,提升了隐私图像的篡改识别结果的准确率。

[34]请参见图 2,为本说明书实施例提供的一种图像篡改检测方法的流程示意图。本说明实施例所述方法的执行主体为电子设备,如图 2 所示,本说明书实施例的所述方法可以包括以下步骤。

[35]S202,确定本地图像篡改检测模型对应的第一模型梯度矩阵,基于第一模型梯度矩阵进行局部梯度差分隐私计算处理得到第二模型梯度矩阵。

[36]易于理解的,本地图像篡改检测模型指的是,基于机器学习模型得到的具有图像检测能力的模型。本地图像篡改检测模型所具有的图像检测能力,可理解为,一种识别图像中可能存在的篡改区域的能力。图像中的篡改区域指的是,经过一种或多种图像操作对图像中原本的图像区域进行修改所得到的区域;图像操作可以包括图像合成、复制粘贴、删改、修饰等操作。具体地,本说明书实施例本地图像篡改检测模型所识别的图像可以为个人身份证件图像、带有用户身份信息的凭证材料的图像等隐私图像;凭证材料可为薪水证明材料、职业证明材料、医疗证明材料等凭证。

[37]第一模型梯度矩阵指的是,本地图像篡改检测模型的模型参数的梯度所构成的矩阵。第一模型梯度矩阵中的各梯度可使用训练数据对本地图像篡改检测模型进行训练得到。模型参数是模型内部用于描述数据的变量,这些变量通常是在模型训练期间学习到的,以最小化模型预测与真实数据之间的误差。在电子设备中,本地图像篡改检测模型的模型参数和模型结构,可从服务平台获取到。具体来说,模型参数可以是权重、偏差或者其他可调整的变量。这些参数通过在训练数据上进行迭代来不断调整,以使模型更准确地预测新的数据。例如,在神经网络中,模型参数是每个神经元的权重和偏差。在线性回归中,模型参数是回归系数和截距。在支持向量机中,模型参数是决策边界和支持向量。

[38]第二模型梯度矩阵,用于指示服务平台基于至少一个电子设备上传的第二模型梯度矩阵确定目标模型参数。第二模型梯度矩阵是对第一模型梯度矩阵中的局部梯度进行差分隐私计算处理所得到的梯度矩阵。

[39]在本说明书一个或多个实施例中,执行确定本地图像篡改检测模型对应的第一模型梯度矩阵的步骤,具体可以是:根据样本隐私图像对本地图像篡改检测模型进行模型训练处理得到第一模型梯度矩阵。其中,样本隐私图像可以包括在历史时间段内采集的已获用户授权的第一隐私图像,还可以包括对已获用户授权的第一隐私图像进行数据扩充处理得到的第二隐私图像。对单个电子设备来说,其采集的已获用户授权的第一隐私图像的数量和种类有限,那么可以通过对第一隐私图像进行诸如拼接、删改、缩放等数据扩充处理,以得到第二隐私图像,以达到扩充样本数据的效果。在利用样本隐私图像对本地图像篡改检测模型进行模型训练处理时,可以通过损失函数和反向传播算法对本地图像篡改检测模型进行多轮训练,得到本地图像篡改检测模型的模型参数的第一模型梯度矩阵。

[40]执行基于第一模型梯度矩阵进行局部梯度差分隐私计算处理得到第二模型梯度矩阵的步骤,具体可以是:确定第一模型梯度矩阵中的第一梯度子集,对第一梯度子集中的所有梯度进行加噪处理得到第二梯度子集,基于第二梯度子集对第一模型梯度矩阵进行更新得到第二模型梯度矩阵。

[41]在确定第一模型梯度矩阵中的第一梯度子集时,可将第一模型梯度矩阵中至少一个指定位置的梯度作为构成第一梯度子集的梯度。其中,指定位置可根据指定参数在网络结构中的位置确定,网络结构指的是本地图像篡改检测模型所采用的网络的结构,指定

参数指的是本地图像篡改检测模型所采用的网络中的网络参数。指定参数可通过查询网络参数映射关系得到。网络参数映射关系中可存储参考模型与参考参数之间的对应关系，当本地图像篡改检测模型作为参考模型时，其对应的参考参数即为指定参数。网络参数映射关系中存储的参考模型与参考参数之间的对应关系可由技术人员根据在先经验进行配置得到。

[42]在对第一梯度子集中的所有梯度进行加噪处理得到第二梯度子集时，可利用高斯噪声等随机噪声对第一梯度子集中的所有梯度进行加噪处理，以得到第二梯度子集。

[43]在基于第二梯度子集对第一模型梯度矩阵进行更新处理得到第二模型梯度矩阵时，可先确定第二梯度子集中的每个第二梯度在第一模型梯度矩阵中的对应位置，然后将第一模型梯度矩阵中每个对应位置处的第一梯度替换为与该对应位置相对应的第二梯度，即可得到第二模型梯度矩阵。

[44]S204，将第二模型梯度矩阵上传至服务平台。

[45]易于理解的，服务平台指的是与电子设备构成联邦学习架构的服务器。联邦学习架构指的是图1所示的系统架构，即联邦学习架构包括一个服务平台和多个电子设备。在联邦学习架构中，电子设备向服务平台发送第二模型梯度矩阵，服务平台对各电子设备发送的第二模型梯度矩阵进行聚合处理，服务平台再将经聚合处理得到的模型参数发送至各电子设备。在服务平台与各电子设备的通信过程中，各电子设备用于计算得到第二模型梯度矩阵的图像数据不会流出到电子设备的外部，即电子设备各自使用的图像数据仅在各电子设备内部，因此采用联邦学习架构实现了在保证数据不出域的情况下，联合训练模型。

[46]在本说明书一个或多个实施例中，执行S204的步骤，具体可以包括：采用预设加密密钥对第二模型梯度矩阵进行加密，得到加密梯度数据，将加密梯度数据上传至服务平台。其中，预设加密密钥可以是电子设备与服务平台预先在可信环境中约定的加密密钥；预设加密密钥也可以是服务平台的公钥。由于电子设备上传的是加密后的第二模型梯度矩阵，那么服务平台可采用与预设加密密钥相对应的预设解密密钥对加密梯度数据进行解密，以得到第二模型梯度矩阵。当预设加密密钥为电子设备与服务平台预先在可信环境中约定的加密密钥时，服务平台采用的预设解密密钥则是与预设加密密钥相对应的解密密钥。当预设加密密钥为服务平台的公钥时，服务平台采用的预设解密密钥则可以是服务平台的私钥。如此，通过将加密的第二模型梯度矩阵发送至服务平台，避免了第二模型梯度矩阵被泄露而存在不安全的问题。

[47]可选地，在上述发送加密的第二模型梯度矩阵至服务平台的基础上，还可在当前时间到达预设发送时间时，将加密的第二模型梯度矩阵发送至服务平台。预设发送时间可为与服务平台约定的发送时间，如预设发送时间为每日20点、22点等时间。如此，通过在特定时间发送第二模型梯度的矩阵至服务平台，以保证模型的联合训练效率。

[48]S206，接收服务平台发送的目标模型参数，基于目标模型参数更新本地图像篡改检测模型得到目标图像篡改检测模型。

[49]易于理解的，目标模型参数指的是服务平台基于各电子设备发送的第二模型梯度矩阵进行聚合处理，得到的本地图像篡改检测模型最新的模型参数。

[50]在本说明书一个或多个实施例中，接收服务平台发送的目标模型参数，可以是服务平台主动向电子设备发送目标模型参数，使得电子设备可接收目标模型参数；还可以是电子设备向服务平台发送参数查询请求，服务平台基于该参数查询请求向电子设备发送目标模型参数，使得电子设备可接收目标模型参数。

[51]具体地，电子设备接收服务平台发送的目标模型参数，可以包括：电子设备接收服务平台发送的加密参数数据，加密参数数据为服务平台对目标模型参数进行加密得到的加密数据，电子设备对加密参数数据进行解密得到目标模型参数。在服务平台对目标模型参数进行加密得到加密参数数据时，所采用的加密密钥可以是电子设备与服务平台预

先在可信环境中约定的预设加密密钥，还可以是电子设备的公钥。相应地，服务平台对目标模型参数进行加密，电子设备则需要对加密参数数据进行解密。当加密密钥为上述预设加密密钥时，电子设备所使用的解密密钥是与该预设加密密钥对应的预设解密密钥。当加密密钥为电子设备的公钥时，电子设备所使用的解密密钥是电子设备的私钥。如此，通过接收服务平台发送的对目标模型参数加密得到的数据，避免了目标模型参数被泄露而存在不安全的问题。

[52]基于目标模型参数更新本地图像篡改检测模型得到目标图像篡改检测模型，具体可以是，将本地图像篡改检测模型的模型参数更新为目标模型参数，得到目标图像篡改检测模型。

[53]S208，基于目标图像篡改检测模型对目标隐私图像进行图像检测处理得到目标隐私图像对应的目标热力检测图，对目标热力检测图进行篡改识别处理得到目标隐私图像对应的篡改识别结果。

[54]易于理解的，目标隐私图像可指经用户授权的待检测的隐私图像。目标隐私图像可以包括个人身份证件的图片、带有用户身份信息的凭证材料的图像等隐私图像；凭证材料可为薪水证明材料、职业证明材料、医疗证明材料等凭证。

[55]目标热力检测图，可指通过使用颜色编码来显示篡改区域的分布情况的图像数据。具体地，在目标热力检测图中，可通过不同颜色来显示不同概率的篡改区域，该概率指的是经目标图像篡改检测模型检测的篡改区域发生篡改的概率。

[56]篡改识别结果可包括目标隐私图像是否为篡改图像的结果。

[57]在本说明书一个或多个实施例中，将目标隐私图像输入至目标图像篡改检测模型，输出目标热力检测图。进一步地，可采用设备专有分类器对目标热力检测图进行篡改识别处理得到目标隐私图像对应的篡改识别结果。其中，设备专有分类器可基于不同电子设备拥有的隐私图像的数据特征确定，不同电子设备内的设备专有分类器可不完全相同。例如，不同电子设备可采用相同网络结构的分类器，不同电子设备可根据自身的隐私图像的数据特征在该分类器中配置不同的判别阈值，以实现不同电子设备对公用模型识别出来的结果进行针对性判别。篡改识别结果可为目标隐私图像为篡改图像，篡改识别结果也可为目标隐私图像不为篡改图像。如此，通过针对不同电子设备配置不同的设备专有分类器，以准确确定各电子设备需检测的隐私图像的篡改识别结果。

[58]可选地，篡改识别结果可包括目标隐私图像是否为篡改图像的结果，还可包括目标隐私图像中存在的篡改区域。此种情况下，执行对目标热力检测图进行篡改识别处理得到目标隐私图像对应的篡改识别结果的步骤，具体可以是：基于预设篡改校验方法对目标热力检测图中的参考篡改区域进行校验处理，得到各参考篡改区域对应的真实篡改结果；基于真实篡改结果确定目标隐私图像对应的篡改识别结果。

[59]其中，预设篡改校验方法可为电子设备基于其拥有的隐私图像的数据特征确定的篡改校验方法，不同电子设备具有不同的预设篡改校验方法。

[60]基于真实篡改结果确定目标隐私图像对应的篡改识别结果，可理解为，当至少一个真实篡改结果为存在篡改类型的第一篡改结果时，将第一篡改结果对应的参考篡改区域作为目标隐私图像中的篡改区域，生成包含篡改区域的篡改识别结果；当所有真实篡改结果为不存在篡改类型的第二篡改结果时，生成目标隐私图像不为篡改图像的篡改识别结果。如此，不同电子设备采用各自的预设篡改校验方法对目标热力检测图中可能存在的篡改区域进行校验，可准确确定各电子设备需检测的隐私图像的篡改识别结果。

[61]本说明书实施例采用上述方法步骤，电子设备通过对即将上传至服务平台的梯度数据进行前处理，以提升上传的梯度数据的抗逆向攻击能力，由于电子设备对隐私图像进行图像检测处理所采用的目标图像篡改检测模型，是根据服务平台结合多个电子设备的梯度数据加工后得到的模型参数所得到的模型，该模型在训练时结合了多方数据，因此该模型对隐私图像的识别准确率得到提升，因而再对该模型的识别结果进行篡改识别处

理，提升了隐私图像的篡改识别结果的准确率。

[62]请参见图 3，为本说明书实施例提供的一种图像篡改检测方法的流程示意图。本说明实施例所述方法的执行主体为电子设备，如图 3 所示，本说明书实施例的所述方法可以包括以下步骤。

[63]S302，确定本地图像篡改检测模型对应的第一模型梯度矩阵。

[64]在本说明书一个或多个实施例中，执行 S302，具体可以是：获取历史隐私图像；基于历史隐私图像对本地图像篡改检测模型进行模型训练处理得到第一模型梯度矩阵。

[65]其中，历史隐私图像可以是金融反欺诈识别场景中经用户授权采集的涉及身份信息、带有用户身份信息的凭证材料的图像。身份信息可以指个人身份证件。凭证材料可以指薪水证明材料、职业证明材料、医疗证明材料等材料。

[66]获取历史隐私图像，可以是：确定当前训练事务场景，从预设数据库中获取当前训练事务场景对应的历史隐私图像。预设数据库中存储的隐私图像，可以为多种训练事务场景分别对应的隐私图像，每种训练事务场景对应的隐私图像可以是用户终端在历史时间段内上传的隐私图像，历史时间段可以是过去 24 小时、过去 7 天等时间段。

[67]当前训练事务场景指的是，各电子设备与服务平台约定的本次联合训练的事务场景。本说明书实施例涉及的场景为金融反欺诈识别场景，金融反欺诈识别场景具体又可以包括多种不同的事务场景，如贫困证明识别场景、医疗证明识别场景、贷款证明识别场景等场景，当前训练事务场景可以为金融反欺诈识别场景下的任意一个场景。在各电子设备与服务平台采用联邦学习架构联合训练模型的过程中，可以分场景联合训练模型，如联合训练的次数达到预设轮次可以完成一个事务场景的模型训练。本说明书实施例通过当前训练事务场景来获取隐私图像，保证了各电子设备联合训练模型时所使用的数据内容的一致性，利于提升后续联合建模得到的模型的识别精度。

[68]基于历史隐私图像对本地图像篡改检测模型进行模型训练处理得到第一模型梯度矩阵，可理解为，对历史隐私图像进行数据扩充处理得到参考隐私图像，对参考隐私图像进行标注处理得到训练数据，采用训练数据对本地图像篡改检测模型进行模型训练处理得到第一模型梯度矩阵。

[69]具体地，对历史隐私图像进行数据扩充处理得到参考隐私图像，可以是采用数据扩充方法对历史隐私图像进行处理得到参考隐私图像。数据扩充方法可以包括但不限于图像复制粘贴、图像切割拼接、图像生成、图像重采样与缩放等方法。图像复制粘贴指的是，将历史隐私图像中某个位置的数据复制并粘贴到该历史隐私图像中的另一位置。图像切割拼接指的是，从多个历史隐私图像中分别切割出部分图像，将切割得到的部分图像进行拼接。图像生成指的是，利用数字图像处理技术，将历史隐私图像中某个数据替换为与该相似的相似数据；或者，将历史隐私图像输入至生成式模型得到参考隐私图像，参考隐私图像中某位置的数据是历史隐私图像中该位置的初始数据的相似数据。图像重采样与缩放指的是，通过改变历史隐私图像的分辨率或尺寸或采样率，以达到调整历史隐私图像的清晰度或比例或视觉效果的目的。本说明书实施例通过上述数据扩充方法增大了模型的样本数量，进而提升了模型的泛化能力。

[70]具体地，对参考隐私图像进行标注处理得到训练数据，可以包括标注参考隐私图像中的篡改区域、标注参考隐私图像的篡改标签、标注参考隐私图像中的隐私区域等标注操作，以得到包含参考隐私图像、参考隐私图像对应的篡改区域、参考隐私图像对应的篡改标签、参考隐私图像对应的隐私区域等标注信息的训练数据。标注参考隐私图像的篡改标签，可使用预设字符进行表示，如预设字符为 1 和 0，当篡改标签为 1 时，可表示该参考隐私图像中存在篡改区域；当篡改标签为 0 时，可表示该参考隐私图像中不存在篡改区域。

[71]具体地，采用训练数据对本地图像篡改检测模型进行模型训练处理得到第一模型梯度矩阵，可理解为，将训练数据作为批数据输入至本地图像篡改检测模型进行前向传播

计算,得到本地图像篡改检测模型的输出,然后基于本地图像篡改检测模型的输出采用误差反向传播算法计算本地图像篡改检测模型的模型参数的梯度,由这些梯度构成梯度矩阵,即可得到第一模型梯度矩阵。

[72]S304,确定第一模型梯度矩阵中的初始局部梯度子集。

[73]在本说明书一个或多个实施例中,执行 S304 的步骤,具体可以是:A2,确定本地图像篡改检测的模型特征;A4,基于模型特征确定本地图像篡改检测模型对应的候选网络参数;A6,在第一模型梯度矩阵中确定候选网络参数对应的初始局部梯度子集。

[74]其中,模型特征指的是模型采用的网络结构的特征。例如,模型特征可以包括网络结构中网络层的特征,诸如网络层的数量、网络层的功能等特征。

[75]候选网络参数指的是,根据模型特征从本地图像篡改检测模型采用的网络结构中选取的网络参数。

[76]初始局部梯度子集指的是,由第一模型梯度矩阵的部分梯度所组成的梯度集合。

[77]具体地,在执行步骤 A2 时,可以是:从模型配置文件中获取本地图像篡改检测模型的模型特征。模型配置文件中可存储模型的模型特征、模型参数等模型配置信息。

[78]具体地,在执行步骤 A4 时,可以是:a2,基于模型特征确定本地图像篡改检测模型中的候选网络层;a4,确定样本隐私图像对应的样本掩码图像,将样本掩码图像输入至本地图像篡改检测模型得到候选网络层中的可学习网络参数;a6,基于可学习网络参数确定本地图像篡改检测模型对应的候选网络参数。

[79]在执行步骤 a2 时,获取网络层映射关系,网络层映射关系用于存储参考模型特征与参考候选网络层之间的对应关系,在网络层映射关系中查找模型特征对应的目标候选网络层,目标候选网络层即为本说明实施例中本地图像篡改检测模型的候选网络层。网络层映射关系可由模型网络专家配置生成,也可由技术人员根据在先经验配置生成。举例来说,如本地图像篡改检测模型采用的网络是 CATNet,网络层映射关系中 CATNet 的模型特征对应的参考候选网络层可为 RGB 分支中的第一个卷积层、RGB 分支中的最后一个卷积层、DCT 分支中的第一个卷积层。当根据 CATNet 的模型特征查询网络层映射关系时,即可得到候选网络层为 RGB 分支中的第一个卷积层、RGB 分支中的最后一个卷积层、DCT 分支中的第一个卷积层。

[80]在执行步骤 a4 时,确定样本隐私图像对应的样本掩码图像,可理解为,从电子设备全量的隐私图像中选取预设数量的样本隐私图像,预设数量可根据实际需求进行设定,然后将样本隐私图像中的至少一个隐私区域作为掩码区域,对掩码区域进行掩码操作,得到仅包含掩码区域的样本掩码图像。将样本掩码图像输入至本地图像篡改检测模型得到候选网络层中的可学习网络参数,可理解为,将样本掩码图像输入至本地图像篡改检测模型,得到候选网络层中的网络参数,从这些网络参数中确定可学习网络参数,此处的可学习网络参数包含可学习网络参数的名称和可学习网络参数的参数值。可学习网络参数指的是在训练过程中学习参数值的参数,即参数值为可变数值的参数。可学习参数的参数值为可变数值,可以理解为,其参数值通常从设定为一组随机值开始,然后随着网络的学习,以迭代的方式更新这些值。例如,可学习参数可以为卷积层的参数,包括卷积核的权值、各通道的偏置量等。

[81]在执行步骤 a6 时,基于可学习网络参数确定本地图像篡改检测模型对应的候选网络参数,可理解为,将可学习网络参数中绝对值最大的参数作为本地图像篡改检测模型对应的候选网络参数。这是因为,候选网络参数对隐私图像中隐私区域的输入有强响应输出,候选网络参数可用于定位隐私区域对应的梯度。

[82]具体地,在执行步骤 A4 时,可以是:b2,在第一模型梯度矩阵中确定各候选网络参数对应的至少一个梯度;b4,对至少一个梯度进行过滤处理,得到候选网络参数对应的目标梯度;b6,基于目标梯度确定候选网络参数对应的初始局部梯度子集。

[83]在执行步骤 b2 时,在获得候选网络参数后,容易确定候选网络参数在本地图像篡改

检测模型采用的网络中的位置信息,由于第一模型梯度矩阵中不同位置的梯度与网络中不同位置的参数具有对应关系,因此可以根据候选网络参数在网络中的位置信息从第一模型梯度矩阵中选取每个候选网络参数对应的至少一个梯度。

[84]在执行步骤 b4 时,针对每个候选网络参数对应的至少一个梯度,可从梯度过滤映射表中获取该候选网络参数对应的梯度阈值,使用梯度阈值对至少一个梯度进行过滤处理,即将至少一个梯度中小于该梯度阈值的梯度进行滤除,得到该候选网络参数对应的目标梯度。其中,梯度过滤映射表中预先配置有不同候选网络参数对应的梯度阈值。梯度阈值的作用在于过滤掉候选网络参数对应的梯度中较小的梯度。

[85]在执行步骤 b6 时,针对每个候选网络参数,其对应目标梯度,将所有候选网络参数对应的目标梯度存储在一梯度集合得到初始局部梯度子集。

[86]S306,对初始局部梯度子集进行加噪处理得到目标局部梯度子集,基于目标局部梯度子集对第一模型梯度矩阵进行更新处理得到第二模型梯度矩阵。

[87]在本说明书一个或多个实施例中,执行对初始局部梯度子集进行加噪处理得到目标局部梯度子集,具体可以是: B2,对初始局部梯度子集进行梯度裁剪处理得到参考局部梯度子集; B4,对参考局部梯度子集进行加噪处理得到目标局部梯度子集。

[88]在执行步骤 B2 时,可通过梯度边界对初始局部梯度子集进行梯度裁剪处理的,得到参考局部梯度子集。梯度裁剪公式可以为:

$$(G) = \frac{G}{\max \left(1, \frac{\|G\|_F^2}{C} \right)}$$

[89]其中, C 表示梯度边界, G 表示初始局部梯度子集, (G) 表示参考局部梯度子集, $\|G\|_F^2$ 表示二范数。C 的取值可根据实际应用进行设定。通过上述梯度裁剪公式计算得到的梯度都不会大于 C,也就是参考局部梯度子集中的梯度都不会大于 C。本说明书实施例通过对梯度子集进行裁剪处理,保证了模型在训练过程中的稳定性。

[90]在执行步骤 B4 时,对参考局部梯度子集中的每一个梯度都进行加噪处理,所有加噪后的梯度构成目标局部梯度子集。例如,加噪的公式可以为:

$$(g) = g + N \left(0, \frac{\sigma^2 C^2}{m} \right)$$

[91]其中, m 为步骤 a4 中选取样本隐私图像的数量, σ^2 为高斯噪声的方差, C 为上述梯度边界, g 表示参考局部梯度子集, (g) 表示目标局部梯度子集。设定隐私预算 ϵ 和松弛项

$$\sigma = \sqrt{2 \log \frac{1.25}{\delta}}$$

δ 后,可知当 ϵ 时,通过隐私区域定位的局部梯度子集 G 达到 (ϵ, δ) 的差分隐私保护级别。隐私预算 ϵ 是控制差分隐私保护效果的超参数,其取值越小,隐私保护效果越好,但引入的噪声就越大。

[92]在执行基于目标局部梯度子集对第一模型梯度矩阵进行更新处理得到第二模型梯度矩阵时,具体可以是,确定目标局部梯度子集中各梯度在第一模型梯度矩阵中的对应位置,将第一模型梯度矩阵中这些对应位置处的梯度分别替换为目标局部梯度子集中与该对应位置相对应的梯度,得到第二模型梯度矩阵。例如,目标局部梯度子集中的一个梯度为 g1, g1 在第一模型梯度矩阵中的对应位置为第 3 行第 4 列,第一模型梯度矩阵中第 3 行第 4 列的梯度为 G1,将第一模型梯度矩阵中的 G1 替换为 g1。依次类推,将第一模型梯度矩阵中相应位置的梯度替换为目标局部梯度子集中的对应梯度,即可得到第二模型梯度矩阵。

[93]S308,将第二模型梯度矩阵上传至服务平台。

[94]S310, 接收服务平台发送的目标模型参数, 基于目标模型参数更新本地图像篡改检测模型得到目标图像篡改检测模型。

[95]具体地, S308 和 S310 的实现方式具体可参见图 2 所示实施例中 S204 和 S206 的解释说明, 在此不再赘述。

[96]S312, 基于目标图像篡改检测模型对目标隐私图像进行图像检测处理得到目标隐私图像对应的目标热力检测图。

[97]目标隐私图像, 目标热力检测图的释义具体可参见图 2 所示实施例中 S208 的解释说明, 在此不再赘述。

[98]在本说明书一个或多个实施例中, 将目标隐私图像输入至目标图像篡改检测模型, 得到目标图像篡改检测模型输出的目标热力检测图。例如, 可参见图 4 所示的目标热力检测图的示意图。图 4 中标记“红色区域”、“橙色区域”、“黄色区域”是目标图像篡改检测模型检测出可能存在篡改的篡改区域, 不同颜色区分的篡改区域表示不同的篡改概率, 如“红色区域”的篡改概率大于“橙色区域”的篡改概率, “橙色区域”的篡改概率大于“黄色区域”的篡改概率。

[99]S314, 获取设备专有分类器, 基于设备专有分类器对目标热力检测图进行篡改识别处理得到目标隐私图像对应的篡改识别结果。

[100]易于理解的, 设备专有分类器指的是针对电子设备持有的隐私图像的数据特征所设定的专有分类器。不同电子设备的设备专有分类器不全相同。

[101]由于不同电子设备持有的隐私图像的数据特征不全相同, 不同电子设备对本地图像篡改检测模型的召回率和准确率的偏好有所区别, 那么不同电子设备可设定不同判别阈值的设备专有分类器, 以适配不同电子设备对隐私图像的篡改识别需求。例如, 目标热力检测图指示某个字符可能存在篡改, 而对于某个电子设备来说, 连续的字符出现篡改才认为是发生了篡改, 单个字符的篡改不认为是发生了篡改, 因此, 需要不同电子设备配置适配自身识别需求的分类器来确定最终的篡改识别结果。

[102]在本说明书一个或多个实施例中, 可从电子设备本地配置的模型库中选取设备专有分类器, 以获取设备专有分类器, 进一步地, 将目标热力检测图输入至设备专有分类器, 得到输出的篡改识别结果。篡改识别结果可为二分类结果, 即篡改识别结果可包括有篡改和无篡改两种结果, 有篡改表示目标隐私图像存在篡改, 无篡改可表示目标隐私图像不存在篡改。

[103]可选地, 本说明书实施例的设备专有分类器可由一个 3 层的全连接网络构成, 且分类器使用 sigmoid 激活函数得到最终的分类输出。电子设备通过二分类交叉熵损失和误差反向传播算法对分类器进行训练。使用 sigmoid 激活函数的模型输出值分布区间为 0-1, 模型输出值可用于表示图像存在篡改的概率。不同电子设备可依据事务开展情况和数据验真的重要性选择合适的判别阈值, 以得到篡改识别结果。

[104]例如, 当设备专有分类器的判别阈值设置为 0.7 时, 模型输出值等于或者大于 0.7 的目标隐私图像会被识别为存在篡改的图像。沿用图 4 所示的目标热力检测图的示意图, 假如本说明书实施例中的设备分类器对“红色区域”、“橙色区域”、“黄色区域”的模型输出值分别为 0.81、0.67、0.32, 那么对于该设备分类器来说, “红色区域”的篡改识别结果为有篡改, “橙色区域”的篡改识别结果为无篡改, “黄色区域”的篡改识别结果为无篡改。

[105]在本说明书实施例中, 设备专有分类器判别阈值设置的越大, 设备专有分类器识别的准确度越高, 召回率越低。

[106]在本说明书实施例中, 在电子设备本地, 电子设备先确定本地图像篡改检测模型的第一模型梯度矩阵, 然后对第一模型梯度矩阵中对隐私区域感知明显的梯度进行加噪处理, 以得到第二模型梯度矩阵, 这样, 通过上传至服务平台进行联合训练的梯度矩阵中添加噪声扰动, 以防御外部攻击者通过窃取上传的梯度实施逆向攻击, 达到提升隐私

数据的安全性的效果。另外，在接收服务平台发送的目标模型参数对模型进行更新以得到最新的目标图像篡改检测模型，还通过针对电子设备的设备专有分类器对目标图像篡改检测模型对目标隐私图像的检测结果进行二次篡改识别处理，在适配不同电子设备的篡改检测需求的同时，也提升了各电子设备对隐私图像的篡改识别结果的准确率。

[107]请参见图 5，为本说明书实施例提供的一种图像篡改检测方法的流程示意图。本说明实施例所述方法的执行主体为服务平台，如图 5 所示，本说明书实施例的所述方法可以包括以下步骤。

[108]S502，接收至少一个电子设备上传的第二模型梯度矩阵。

[109]第二模型梯度矩阵由电子设备基于本地图像篡改检测模型对应的第一模型梯度矩阵进行局部梯度差分隐私计算处理得到。各电子设备由第一模型梯度矩阵计算得到第二模型梯度矩阵的步骤具体可参见图 2 所示实施例中 S202 的描述，在此不再赘述。

[110]在一些实施例中，执行 S502 的步骤，具体可以包括：接收至少一个电子设备上传的加密梯度数据，使用预设解密密钥对各加密梯度数据进行解密，以得到各电子设备分别对应的第二模型梯度矩阵。对不同电子设备发送的加密梯度数据进行解密，所采用的预设解密密钥可以不同。其中，预设解密密钥可以是电子设备与服务平台预先在可信环境中约定的解密密钥；预设解密密钥还可以是电子设备的公钥。服务平台对加密梯度数据进行解密，相应地，则需要电子设备对第二模型梯度矩阵进行加密。当预设解密密钥为电子设备与服务平台预先在可信环境中约定的解密密钥时，电子平台采用的预设加密密钥则是与预设解密密钥相对应的加密密钥。当预设解密密钥为电子设备的公钥时，电子设备采用的加密密钥则可以是电子设备的私钥。如此，通过将加密的第二模型梯度矩阵发送至服务平台，避免了第二模型梯度矩阵被泄露而存在不安全的问题。

[111]可选地，在执行 S502 之前，服务平台可向各电子设备共享本地图像篡改检测模型，具体地，服务平台可向电子设备发送本地图像篡改检测模型的模型架构和初始模型参数，本地图像篡改检测模型可为服务平台基于开源的隐私数据训练集进行模型训练得到的模型。

[112]S504，基于各第二模型梯度矩阵确定目标模型参数。

[113]易于理解的，目标模型参数指的是服务平台基于各电子设备发送的第二模型梯度矩阵进行聚合处理，得到的本地图像篡改检测模型最新的模型参数。

[114]在本说明书一个或多个实施例中，服务平台可对相同批次的各第二模型梯度矩阵进行梯度聚合处理得到聚合梯度，根据聚合梯度对本地图像篡改检测模型的模型参数进行更新得到目标模型参数。相同批次的各第二模型梯度矩阵可指相同时间段内接收的各第二模型梯度矩阵。服务平台与各电子设备可事先约定每一轮上传第二模型梯度矩阵的时间段，在每一轮计算目标模型参数的过程中，即可利用该时间段内的第二模型梯度矩阵进行计算，以得到目标模型参数。

[115]可选地，服务平台得到聚合梯度的算法可采用联邦平均算法(Federated Average，简称 FedAvg)。

[116]S506，将目标模型参数发送至各电子设备。

[117]目标模型参数用于指示电子设备基于目标模型参数更新本地图像篡改检测模型得到目标图像篡改检测模型，并基于目标图像篡改检测模型对目标隐私图像进行图像检测处理得到目标隐私图像对应的目标热力检测图，对目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果。各电子设备基于目标模型参数更新本地图像篡改检测模型得到目标图像篡改检测模型，以及根据目标图像篡改检测模型进行后续的处理过程，可参见图 2 所示实施例中 S206-S208 的描述，在此不再赘述。

[118]在一些实施例中，执行 S506 的步骤，具体可以包括：采用预设加密密钥对目标模型参数进行加密得到加密参数数据，将加密参数数据发送至各电子设备。发送至不同电子设备的加密参数数据，可采用不同的加密密钥。其中，服务平台采用的预设加密密钥

可以是电子设备与服务平台预先在可信环境中约定的加密密钥；预设加密密钥还可以是该电子设备的公钥。服务平台对目标模型参数进行加密，相应地，则需要电子设备对加密参数数据进行解密以得到目标模型参数。当服务平台采用的预设加密密钥为电子设备与服务平台预先在可信环境中约定的加密密钥，电子设备使用的解密密钥可以是与该预设加密密钥对应的预设解密密钥；当服务平台所采用的预设加密密钥为电子设备的公钥时，电子设备解密所使用的解密密钥可以是电子设备的私钥。如此，服务平台将加密的目标模型参数发送至电子设备，避免了目标模型参数被泄露而存在不安全的问题。

[119]需要说明的是，服务平台可通过多个批次的第二模型梯度矩阵计算多轮目标模型参数，以实现对各电子设备共享的本地图像篡改检测模型不断进行更新，直至本地图像篡改检测模型收敛。在服务平台每一次更新本地图像篡改检测模型时，都可将本地图像篡改检测模型最新的模型参数，即目标模型参数发送至各电子设备。

[120]在本说明书实施例中，服务平台通过接收各电子设备上传的第二模型梯度矩阵，第二模型梯度矩阵是电子设备通过局部差分隐私计算后得到的梯度矩阵，然后根据第二模型梯度矩阵确定目标模型参数，达到了利用多方数据整合训练模型的效果，进而提升了模型的识别能力。由于电子设备上传的第二模型梯度矩阵的隐私级别更高，很好地抵御了外部攻击者的逆向攻击，保证了联合训练模型过程中数据的安全性。服务平台再将目标模型参数发送至各电子设备，可使电子设备根据目标模型参数对服务平台共享的本地图像篡改检测模型进行更新，以得到识别能力更好的目标图像篡改检测模型，因此提升了电子设备利用目标图像篡改检测模型对于隐私图像的检测效果，进而提升了电子设备对利用目标图像篡改检测模型的检测结果进行二次识别得到的识别结果的准确率。

[121]请参见图 6，为本说明书实施例提供的一种图像篡改检测方法的流程示意图。本说明实施例所述方法的执行主体为服务平台，如图 6 所示，本说明书实施例的所述方法可以包括以下步骤。

[122]S602，接收至少一个电子设备上传的第二模型梯度矩阵。

[123]第二模型梯度矩阵由电子设备基于本地图像篡改检测模型对应的第一模型梯度矩阵进行局部梯度差分隐私计算处理得到。各电子设备由第一模型梯度矩阵计算得到第二模型梯度矩阵的步骤具体可参见图 3 所示实施例中 S302-S306 的描述，在此不再赘述。

[124]S604，基于各第二模型梯度矩阵计算目标梯度。

[125]在一些实施例中，在执行 S604 时，具体可以包括：C2，获取各电子设备上传的矩阵训练数据量，矩阵训练数据量为模型训练阶段生成第二模型梯度矩阵时的训练数据消耗值；C4，基于所有矩阵训练数据量确定数据量总和；C6，确定矩阵训练数据量和数据量总和的比值，将比值作为电子设备对应的梯度权重；C8，基于梯度权重和第二模型梯度矩阵进行加权求和，得到目标梯度。

[126]其中，矩阵训练数据量为电子设备在得到第二模型梯度矩阵时的训练数据消耗值，训练数据消耗值指的是电子设备在得到第二模型梯度矩阵时所采用的样本隐私图像的数量，由于第二模型梯度矩阵由第一模型梯度矩阵，训练数据消耗值也可理解为电子设备在训练得到第一模型梯度矩阵时所使用的样本隐私图像的数量。

[127]在执行步骤 C4 时，可以将所有矩阵训练数据量的和值，得到数据量总和。

[128]在执行步骤 C6 时，对于每一个矩阵训练数据量，可以计算其与数据量总和的比值，得到电子设备对应的梯度权重。

[129]在执行步骤 C8 时，对于每一个电子设备，可以计算其对应的梯度权重和其上传的第二模型梯度矩阵的乘积，然后对这些乘积进行求和计算，得到目标梯度。

[130]本申请实施例通过加权求和计算梯度均值的方式，以对各第二模型梯度矩阵进行综合评估，保证用于更新模型参数的梯度的准确性。

[131]S606，获取本地图像篡改检测模型对应的学习率，计算学习率和目标梯度的乘积。

[132]S608，获取本地图像篡改模型对应的初始模型参数，计算乘积和初始模型参数的和

值, 将和值作为目标模型参数。

[133]易于理解的, 初始模型参数指的是服务平台与电子设备在上一轮聚合训练中得到的模型参数。

[134]在本说明书实施例中, 计算目标模型参数的公式可以为:

$$w_{t+1} = w_t + \eta \sum_{k=1}^K \frac{n_k}{n} G_i^t *$$

[135]其中, w_t 表示初始模型参数, w_{t+1} 表示目标模型参数, η 表示学习率, K 表示电子设备的总数量, n_k 表示第 k 个电子设备上传的矩阵训练数据量, n 表示数据量总和, $G_i^t *$ 表示第 i 个电子设备的第二模型梯度矩阵。

[136]S610, 将目标模型参数发送至各电子设备。

[137]目标模型参数用于指示电子设备基于目标模型参数更新本地图像篡改检测模型得到目标图像篡改检测模型, 并基于目标图像篡改检测模型对目标隐私图像进行图像检测处理得到目标隐私图像对应的目标热力检测图, 对目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果。各电子设备基于目标模型参数更新本地图像篡改检测模型得到目标图像篡改检测模型, 以及根据目标图像篡改检测模型进行后续的处理过程, 可参见图 3 所示实施例中 S310-S314 的描述, 在此不再赘述。

[138]S612, 确定本地图像篡改检测模型的当前训练轮次。

[139]S614, 若当前训练轮次大于或者等于预设轮次, 则向各电子设备发送训练结束指令。

[140]下面对 S612 和 S614 进行解释说明。

[141]易于理解的, 当前训练轮次指的是本地图像篡改检测模型当前已训练的轮数。

[142]预设轮次可指预先设定的模型的最大迭代轮次。

[143]训练结束指令用于指示各电子设备停止上传第二模型梯度矩阵。

[144]在本说明一个或多个实施例中, 确定本地图像篡改检测模型的当前训练轮次, 可理解为, 通过目标模型参数的当前计算次数确定当前训练轮次, 还可通过各电子设备对应的第二模型梯度矩阵的接收次数确定当前训练轮次。进一步地, 可判断当前训练轮次是否大于或者等于预设轮次。在当前训练轮次大于或者等于预设轮次时, 表示模型参数的迭代次数以达到最大迭代轮次, 服务平台与电子设备可停止对本地图像篡改检测模型的联合训练, 那么服务平台可向各电子设备发送训练结束指令, 以通过训练结束指令告知电子设备停止联合训练模型。在当前训练轮次小于预设轮次时, 表示模型参数的迭代次数还未达到最大迭代轮次, 服务平台与电子设备还需继续联合训练模型, 因此服务平台无需向电子设备发送训练结束指令。

[145]在本说明书实施例中, 服务平台通过接收各电子设备上传的第二模型梯度矩阵, 以根据各第二模型梯度矩阵确定目标梯度, 而后再根据目标梯度确定目标模型参数, 在达到利用多方数据整合训练模型的效果的同时, 还通过计算目标模型梯度均值达到了提升梯度的准确度的效果, 进而利用准确度更高的梯度计算准确度更高的目标模型参数, 以使用准确度更高的目标模型参数更新模型, 更好地提升了模型的识别能力。由于各电子设备上传的第二模型梯度矩阵中添加了噪声, 因而第二模型梯度矩阵具有较好的隐私性, 因而服务平台根据第二模型梯度矩阵计算得到的目标模型参数也具有较好的隐私性。服务平台再将准确度更高目标模型参数发送至各电子设备, 可使电子设备根据准确度更高的目标模型参数对服务平台共享的本地图像篡改检测模型进行更新, 以得到识别能力更好的目标图像篡改检测模型, 因此提升了电子设备利用目标图像篡改检测模型对于隐私图像的检测效果, 进而提升了电子设备对利用目标图像篡改检测模型的检测结果进行二次识别得到的篡改识别结果的准确率。

[146]下面将结合图 7, 对本申请实施例提供的图像篡改检测装置进行详细介绍。需要说明的是, 图 7 所示的图像篡改检测装置, 用于执行本申请图 2 和图 3 所示实施例的方法,

为了便于说明，仅示出了与本说明书实施例相关的部分，具体技术细节未揭示的，请参照本申请图 2 和图 3 所示的实施例。

[147]请参见图 7，其示出本说明书实施例的图像篡改检测装置的结构示意图。该图像篡改检测装置 1 可以通过软件、硬件或者两者的结合实现成为装置的全部或一部分。根据一些实施例，该图像篡改检测装置 1 包括梯度处理模块 11、梯度发送模块 12、数据接收模块 13 和图像处理模块 14，具体用于：梯度处理模块 11，用于确定本地图像篡改检测模型对应的第一模型梯度矩阵，基于所述第一模型梯度矩阵进行局部梯度差分隐私计算处理得到第二模型梯度矩阵；梯度发送模块 12，用于将所述第二模型梯度矩阵上传至服务平台，所述第二模型梯度矩阵用于指示所述服务平台基于至少一个所述电子设备上传的所述第二模型梯度矩阵确定目标模型参数；数据接收模块 13，用于接收所述服务平台发送的所述目标模型参数，基于所述目标模型参数更新所述本地图像篡改检测模型得到目标图像篡改检测模型；图像处理模块 14，用于基于所述目标图像篡改检测模型对目标隐私图像进行图像检测处理得到所述目标隐私图像对应的目标热力检测图，对所述目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果。

[148]可选地，梯度处理模块 11 包括：第一处理单元，用于确定所述第一模型梯度矩阵中的初始局部梯度子集；第二处理单元，用于对所述初始局部梯度子集进行加噪处理得到目标局部梯度子集，基于所述目标局部梯度子集对所述第一模型梯度矩阵进行更新处理得到第二模型梯度矩阵。

[149]可选地，第一处理单元包括：参数确定单元，用于确定所述本地图像篡改检测的模型特征，基于所述模型特征确定所述本地图像篡改检测模型对应的候选网络参数；梯度确定单元，用于在所述第一模型梯度矩阵中确定所述候选网络参数对应的初始局部梯度子集。

[150]可选地，参数确定单元，具体用于：基于所述模型特征确定所述本地图像篡改检测模型中的候选网络层；确定样本隐私图像对应的样本掩码图像，将所述样本掩码图像输入至所述本地图像篡改检测模型得到所述候选网络层中的可学习网络参数；基于所述可学习网络参数确定所述本地图像篡改检测模型对应的候选网络参数。

[151]可选地，梯度确定单元，具体用于：在所述第一模型梯度矩阵中确定各所述候选网络参数对应的至少一个梯度；对所述至少一个梯度进行过滤处理，得到所述候选网络参数对应的目标梯度；基于所述目标梯度确定所述候选网络参数对应的初始局部梯度子集。

[152]可选地，第二处理单元，具体用于：对所述初始局部梯度子集进行梯度裁剪处理得到参考局部梯度子集；对所述参考局部梯度子集进行加噪处理得到目标局部梯度子集。

[153]可选地，梯度处理模块 11，具体用于：获取历史隐私图像；基于所述历史隐私图像对本地图像篡改检测模型进行模型训练处理得到第一模型梯度矩阵。

[154]可选地，图像处理模块 14，具体用于：获取设备专有分类器，基于所述设备专有分类器对所述目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果；其中，不同电子设备的设备专有分类器不全相同。

[155]下面将结合图 8，对本申请实施例提供的图像篡改检测装置进行详细介绍。需要说明的是，图 8 所示的图像篡改检测装置，用于执行本申请图 5 和图 6 所示实施例的方法，为了便于说明，仅示出了与本说明书实施例相关的部分，具体技术细节未揭示的，请参照本申请图 5 和图 6 所示的实施例。

[156]请参见图 8，其示出本说明书实施例的图像篡改检测装置的结构示意图。该图像篡改检测装置 1 可以通过软件、硬件或者两者的结合实现成为装置的全部或一部分。根据一些实施例，该图像篡改检测装置 1 包括数据接收模块 11、数据处理模块 12 和数据发送模块 13，具体用于：数据接收模块 11，用于接收至少一个电子设备上传的第二模型梯度矩阵，所述第二模型梯度矩阵由所述电子设备基于本地图像篡改检测模型对应的第一模型梯度矩阵进行局部梯度差分隐私计算处理得到；数据处理模块 12，用于基于各所

述第二模型梯度矩阵确定目标模型参数；数据发送模块 13，用于将所述目标模型参数发送至各所述电子设备，所述目标模型参数用于指示所述电子设备基于所述目标模型参数更新本地图像篡改检测模型得到目标图像篡改检测模型，并基于所述目标图像篡改检测模型对目标隐私图像进行图像检测处理得到所述目标隐私图像对应的目标热力检测图，对所述目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果。

[157]可选地，数据处理模块 12 包括：第一计算单元，用于基于各所述第二模型梯度矩阵计算目标梯度；第二计算单元，用于获取本地图像篡改检测模型对应的学习率，计算所述学习率和所述目标梯度的乘积；第三计算单元，用于获取所述本地图像篡改模型对应的初始模型参数，计算所述乘积和所述初始模型参数的和值，将所述和值作为目标模型参数。

[158]可选地，第一计算单元，具体用于：获取各所述电子设备上传的矩阵训练数据量，所述矩阵训练数据量为模型训练阶段生成所述第二模型梯度矩阵时的训练数据消耗值；基于所有所述矩阵训练数据量确定数据量总和；确定所述矩阵训练数据量和所述数据量总和的比值，将所述比值作为所述电子设备对应的梯度权重；基于所述梯度权重和所述第二模型梯度矩阵进行加权求和，得到目标梯度。

[159]可选地，图像篡改检测装置 1 还用于：确定所述本地图像篡改检测模型的当前训练轮次；若所述当前训练轮次大于或者等于预设轮次，则向各所述电子设备发送训练结束指令，所述训练结束指令用于指示各所述电子设备停止上传所述第二模型梯度矩阵。

[160]请参考图 9，其示出了本说明书一个示例性实施例提供的电子设备的结构示意图。本说明书中的电子设备可以包括一个或多个如下部件：处理器 110、存储器 120、输入装置 130、输出装置 140 和总线 150。处理器 110、存储器 120、输入装置 130 和输出装置 140 之间可以通过总线 150 连接。

[161]处理器 110 可以包括一个或者多个处理核心。处理器 110 利用各种接口和线路连接整个终端内的各个部分，通过运行或执行存储在存储器 120 内的指令、程序、代码集或指令集，以及调用存储在存储器 120 内的数据，执行终端 100 的各种功能和处理数据。可选地，处理器 110 可以采用数字信号处理（digital signal processing, DSP）、现场可编程门阵列（field-programmable gate array, FPGA）、可编程逻辑阵列（programmable logic Array, PLA）中的至少一种硬件形式来实现。处理器 110 可集成中心处理器（central processing unit, CPU）、图像处理器（graphics processing unit, GPU）和调制解调器等中的一种或几种的组合。其中，CPU 主要处理操作系统、用户界面和应用程序等；GPU 用于负责显示内容的渲染和绘制；调制解调器用于处理无线通信。可以理解的是，上述调制解调器也可以不集成到处理器 110 中，单独通过一块通信芯片进行实现。

[162]存储器 120 可以包括随机存储器（random Access Memory, RAM），也可以包括只读存储器（read-only memory, ROM）。可选地，该存储器 120 包括非瞬时性计算机可读介质（non-transitory computer-readable storage medium）。存储器 120 可用于存储指令、程序、代码、代码集或指令集。存储器 120 可包括存储程序区和存储数据区，其中，存储程序区可存储用于实现操作系统的指令、用于实现至少一个功能的指令（例如触控功能、声音播放功能、图像播放功能等）、用于实现下述各个方法实施例的指令等，该操作系统可以是安卓（Android）系统，包括基于 Android 系统深度开发的系统、苹果公司开发的 IOS 系统，包括基于 IOS 系统深度开发的系统或其它系统。

[163]为了使操作系统能够区分第三方应用程序的具体应用场景，需要打通第三方应用程序与操作系统之间的数据通信，使得操作系统能够随时获取第三方应用程序当前的场景信息，进而基于当前场景进行针对性的系统资源适配。

[164]其中，输入装置 130 用于接收输入的指令或数据，输入装置 130 包括但不限于键盘、鼠标、摄像头、麦克风或触控设备。输出装置 140 用于输出指令或数据，输出装置 140 包括但不限于显示设备和扬声器等。在一个示例中，输入装置 130 和输出装置 140 可以

合设，输入装置 130 和输出装置 140 为触摸显示屏。

[165]触摸显示屏可被设计成为全面屏、曲面屏或异型屏。触摸显示屏还可被设计成为全面屏与曲面屏的结合，异型屏与曲面屏的结合，本说明书实施例对此不加以限定。

[166]除此之外，本领域技术人员可以理解，上述附图所示出的电子设备的结构并不构成对电子设备的限定，电子设备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。比如，电子设备中还包括射频电路、输入单元、传感器、音频电路、无线保真（Wireless Fidelity, WiFi）模块、电源、蓝牙模块等部件，在此不再赘述。

[167]在图 9 所示的电子设备中，处理器 110 可以用于调用存储器 120 中存储的图像篡改检测方法的程序，并具体执行以下操作：确定本地图像篡改检测模型对应的第一模型梯度矩阵，基于所述第一模型梯度矩阵进行局部梯度差分隐私计算处理得到第二模型梯度矩阵；将所述第二模型梯度矩阵上传至服务平台，所述第二模型梯度矩阵用于指示所述服务平台基于至少一个所述电子设备上传的所述第二模型梯度矩阵确定目标模型参数；接收所述服务平台发送的所述目标模型参数，基于所述目标模型参数更新所述本地图像篡改检测模型得到目标图像篡改检测模型；基于所述目标图像篡改检测模型对目标隐私图像进行图像检测处理得到所述目标隐私图像对应的目标热力检测图，对所述目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果。

[168]在一个实施例中，处理器 110 在执行所述基于所述第一模型梯度矩阵进行局部梯度差分隐私计算处理得到第二模型梯度矩阵的步骤时，具体执行以下操作：确定所述第一模型梯度矩阵中的初始局部梯度子集；对所述初始局部梯度子集进行加噪处理得到目标局部梯度子集，基于所述目标局部梯度子集对所述第一模型梯度矩阵进行更新处理得到第二模型梯度矩阵。

[169]在一个实施例中，处理器 110 在执行所述确定所述第一模型梯度矩阵中的初始局部梯度子集的步骤时，具体执行以下操作：确定所述本地图像篡改检测的模型特征，基于所述模型特征确定所述本地图像篡改检测模型对应的候选网络参数；在所述第一模型梯度矩阵中确定所述候选网络参数对应的初始局部梯度子集。

[170]在一个实施例中，处理器 110 在执行所述基于所述模型特征确定所述本地图像篡改检测模型对应的候选网络参数的步骤时，具体执行以下操作：基于所述模型特征确定所述本地图像篡改检测模型中的候选网络层；确定样本隐私图像对应的样本掩码图像，将所述样本掩码图像输入至所述本地图像篡改检测模型得到所述候选网络层中的可学习网络参数；基于所述可学习网络参数确定所述本地图像篡改检测模型对应的候选网络参数。

[171]在一个实施例中，处理器 110 在执行所述在所述第一模型梯度矩阵中确定所述候选网络参数对应的初始局部梯度子集的步骤时，具体执行以下操作：在所述第一模型梯度矩阵中确定各所述候选网络参数对应的至少一个梯度；对所述至少一个梯度进行过滤处理，得到所述候选网络参数对应的目标梯度；基于所述目标梯度确定所述候选网络参数对应的初始局部梯度子集。

[172]在一个实施例中，处理器 110 在执行所述对所述初始局部梯度子集进行加噪处理得到目标局部梯度子集，具体执行以下操作：对所述初始局部梯度子集进行梯度裁剪处理得到参考局部梯度子集；对所述参考局部梯度子集进行加噪处理得到目标局部梯度子集。

[173]在一个实施例中，处理器 110 在执行所述确定本地图像篡改检测模型对应的第一模型梯度矩阵的步骤时，具体执行以下操作：获取历史隐私图像；基于所述历史隐私图像对本地图像篡改检测模型进行模型训练处理得到第一模型梯度矩阵。

[174]在一个实施例中，处理器 110 在执行所述对所述目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果的步骤时，具体执行以下操作：获取设备专有分类器，基于所述设备专有分类器对所述目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果；其中，不同电子设备的设备专有分类器不全相同。

[175]请参考图 10,其示出了本说明书一个示例性实施例提供的服务平台的结构示意图。本说明书中的服务平台可以包括一个或多个如下部件:处理器 110、存储器 120、输入装置 130、输出装置 140 和总线 150。处理器 110、存储器 120、输入装置 130 和输出装置 140 之间可以通过总线 150 连接。

[176]处理器 110 可以包括一个或者多个处理核心。处理器 110 利用各种接口和线路连接整个终端内的各个部分,通过运行或执行存储在存储器 120 内的指令、程序、代码集或指令集,以及调用存储在存储器 120 内的数据,执行终端 100 的各种功能和处理数据。可选地,处理器 110 可以采用数字信号处理(digital signal processing, DSP)、现场可编程门阵列(field-programmable gate array, FPGA)、可编程逻辑阵列(programmable logic Array, PLA)中的至少一种硬件形式来实现。处理器 110 可集成中心处理器(central processing unit, CPU)、图像处理器(graphics processing unit, GPU)和调制解调器等中的一种或几种的组合。其中,CPU 主要处理操作系统、用户界面和应用程序等;GPU 用于负责显示内容的渲染和绘制;调制解调器用于处理无线通信。可以理解的是,上述调制解调器也可以不集成到处理器 110 中,单独通过一块通信芯片进行实现。

[177]存储器 120 可以包括随机存储器(random Access Memory, RAM),也可以包括只读存储器(read-only memory, ROM)。可选地,该存储器 120 包括非瞬态性计算机可读介质(non-transitory computer-readable storage medium)。存储器 120 可用于存储指令、程序、代码、代码集或指令集。存储器 120 可包括存储程序区和存储数据区,其中,存储程序区可存储用于实现操作系统的指令、用于实现至少一个功能的指令(例如触控功能、声音播放功能、图像播放功能等)、用于实现下述各个方法实施例的指令等,该操作系统可以是安卓(Android)系统,包括基于 Android 系统深度开发的系统、苹果公司开发的 IOS 系统,包括基于 IOS 系统深度开发的系统或其它系统。

[178]为了使操作系统能够区分第三方应用程序的具体应用场景,需要打通第三方应用程序与操作系统之间的数据通信,使得操作系统能够随时获取第三方应用程序当前的场景信息,进而基于当前场景进行针对性的系统资源适配。

[179]其中,输入装置 130 用于接收输入的指令或数据,输入装置 130 包括但不限于键盘、鼠标、摄像头、麦克风或触控设备。输出装置 140 用于输出指令或数据,输出装置 140 包括但不限于显示设备和扬声器等。在一个示例中,输入装置 130 和输出装置 140 可以合设,输入装置 130 和输出装置 140 为触摸显示屏。

[180]触摸显示屏可被设计成为全面屏、曲面屏或异型屏。触摸显示屏还可被设计成为全面屏与曲面屏的结合,异型屏与曲面屏的结合,本说明书实施例对此不加以限定。

[181]除此之外,本领域技术人员可以理解,上述附图所示出的电子设备的结构并不构成对服务平台的限定,服务平台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。比如,电子设备中还包括射频电路、输入单元、传感器、音频电路、无线保真(Wireless Fidelity, WiFi)模块、电源、蓝牙模块等部件,在此不再赘述。

[182]在图 10 所示的服务平台中,处理器 110 可以用于调用存储器 120 中存储的图像篡改检测方法的程序,并具体执行以下操作:接收至少一个电子设备上传的第二模型梯度矩阵,所述第二模型梯度矩阵由所述电子设备基于本地图像篡改检测模型对应的第一模型梯度矩阵进行局部梯度差分隐私计算处理得到;基于各所述第二模型梯度矩阵确定目标模型参数;将所述目标模型参数发送至各所述电子设备,所述目标模型参数用于指示所述电子设备基于所述目标模型参数更新本地图像篡改检测模型得到目标图像篡改检测模型,并基于所述目标图像篡改检测模型对目标隐私图像进行图像检测处理得到所述目标隐私图像对应的目标热力检测图,对所述目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果。

[183]在一个实施例中,处理器 110 在执行所述基于各所述第二模型梯度矩阵确定目标模型参数的步骤时,具体执行以下操作:基于各所述第二模型梯度矩阵计算目标梯度;获

取本地图像篡改检测模型对应的学习率，计算所述学习率和所述目标梯度的乘积；获取所述本地图像篡改模型对应的初始模型参数，计算所述乘积和所述初始模型参数的和值，将所述和值作为目标模型参数。

[184]在一个实施例中，处理器 110 在执行所述基于各所述第二模型梯度矩阵计算目标梯度的步骤时，具体执行以下操作：获取各所述电子设备上传的矩阵训练数据量，所述矩阵训练数据量为模型训练阶段生成所述第二模型梯度矩阵时的训练数据消耗值；基于所有所述矩阵训练数据量确定数据量总和；确定所述矩阵训练数据量和所述数据量总和的比值，将所述比值作为所述电子设备对应的梯度权重；基于所述梯度权重和所述第二模型梯度矩阵进行加权求和，得到目标梯度。

[185]在一个实施例中，处理器 110 还执行以下操作：确定所述本地图像篡改检测模型的当前训练轮次；若所述当前训练轮次大于或者等于预设轮次，则向各所述电子设备发送训练结束指令，所述训练结束指令用于指示各所述电子设备停止上传所述第二模型梯度矩阵。

[186]需要说明的是，本说明书实施例所涉及的信息（包括但不限于用户设备信息、用户个人信息等）、数据（包括但不限于用于分析的数据、存储的数据、展示的数据等）以及信号，均为经用户授权或者经过各方充分授权的，且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准。例如，本说明书中涉及的历史隐私图像、目标隐私图像等都是在充分授权的情况下获取的。

[187]本说明书实施例还提供了一种计算机程序产品，该计算机程序产品存储有至少一条指令，所述至少一条指令由所述处理器加载并执行以实现如上各个实施例所述的图像篡改检测方法。

[188]本领域技术人员应该可以意识到，在上述一个或多个示例中，本说明书实施例所描述的功能可以用硬件、软件、固件或它们的任意组合来实现。当使用软件实现时，可以将这些功能存储在计算机可读介质中或者作为计算机可读介质上的一个或多个指令或代码进行传输。计算机可读介质包括计算机存储介质和通信介质，其中通信介质包括便于从一个地方向另一个地方传送计算机程序的任何介质。存储介质可以是通用或专用计算机能够存取的任何可用介质。

[189]以上所述仅为本说明书的可选实施例，并不用以限制本说明书，凡在本说明书的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本说明书的保护范围之内。

[190]上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中，多任务处理和并行处理也是可以的或者可能是有利的。

权利要求书

- 1、一种图像篡改检测方法，应用于电子设备，所述方法包括：
确定本地图像篡改检测模型对应的第一模型梯度矩阵，基于所述第一模型梯度矩阵进行局部梯度差分隐私计算处理得到第二模型梯度矩阵；
将所述第二模型梯度矩阵上传至服务平台，所述第二模型梯度矩阵用于指示所述服务平台基于至少一个所述电子设备上传的所述第二模型梯度矩阵确定目标模型参数；
接收所述服务平台发送的所述目标模型参数，基于所述目标模型参数更新所述本地图像篡改检测模型得到目标图像篡改检测模型；
基于所述目标图像篡改检测模型对目标隐私图像进行图像检测处理得到所述目标隐私图像对应的目标热力检测图，对所述目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果。
- 2、根据权利要求1所述的方法，所述基于所述第一模型梯度矩阵进行局部梯度差分隐私计算处理得到第二模型梯度矩阵，包括：
确定所述第一模型梯度矩阵中的初始局部梯度子集；
对所述初始局部梯度子集进行加噪处理得到目标局部梯度子集，基于所述目标局部梯度子集对所述第一模型梯度矩阵进行更新处理得到第二模型梯度矩阵。
- 3、根据权利要求2所述的方法，所述确定所述第一模型梯度矩阵中的初始局部梯度子集，包括：
确定所述本地图像篡改检测的模型特征，基于所述模型特征确定所述本地图像篡改检测模型对应的候选网络参数；
在所述第一模型梯度矩阵中确定所述候选网络参数对应的初始局部梯度子集。
- 4、根据权利要求3所述的方法，所述基于所述模型特征确定所述本地图像篡改检测模型对应的候选网络参数，包括：
基于所述模型特征确定所述本地图像篡改检测模型中的候选网络层；
确定样本隐私图像对应的样本掩码图像，将所述样本掩码图像输入至所述本地图像篡改检测模型得到所述候选网络层中的可学习网络参数；
基于所述可学习网络参数确定所述本地图像篡改检测模型对应的候选网络参数。
- 5、根据权利要求3所述的方法，所述在所述第一模型梯度矩阵中确定所述候选网络参数对应的初始局部梯度子集，包括：
在所述第一模型梯度矩阵中确定各所述候选网络参数对应的至少一个梯度；
对所述至少一个梯度进行过滤处理，得到所述候选网络参数对应的目标梯度；
基于所述目标梯度确定所述候选网络参数对应的初始局部梯度子集。
- 6、根据权利要求2所述的方法，所述对所述初始局部梯度子集进行加噪处理得到目标局部梯度子集，包括：
对所述初始局部梯度子集进行梯度裁剪处理得到参考局部梯度子集；
对所述参考局部梯度子集进行加噪处理得到目标局部梯度子集。
- 7、根据权利要求1所述的方法，所述确定本地图像篡改检测模型对应的第一模型梯度矩阵，包括：
获取历史隐私图像；
基于所述历史隐私图像对本地图像篡改检测模型进行模型训练处理得到第一模型梯度矩阵。
- 8、根据权利要求1所述的方法，所述对所述目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果，包括：
获取设备专有分类器，基于所述设备专有分类器对所述目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果；其中，不同电子设备的设备专有分类器不全相同。

9、一种图像篡改检测方法，应用于服务平台，所述方法包括：

接收至少一个电子设备上传的第二模型梯度矩阵，所述第二模型梯度矩阵由所述电子设备基于本地图像篡改检测模型对应的第一模型梯度矩阵进行局部梯度差分隐私计算处理得到；

基于各所述第二模型梯度矩阵确定目标模型参数；

将所述目标模型参数发送至各所述电子设备，所述目标模型参数用于指示所述电子设备基于所述目标模型参数更新本地图像篡改检测模型得到目标图像篡改检测模型，并基于所述目标图像篡改检测模型对目标隐私图像进行图像检测处理得到所述目标隐私图像对应的目标热力检测图，对所述目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果。

10、根据权利要求9所述的方法，所述基于各所述第二模型梯度矩阵确定目标模型参数，包括：

基于各所述第二模型梯度矩阵计算目标梯度；

获取本地图像篡改检测模型对应的学习率，计算所述学习率和所述目标梯度的乘积；

获取所述本地图像篡改模型对应的初始模型参数，计算所述乘积和所述初始模型参数的和值，将所述和值作为目标模型参数。

11、根据权利要求10所述的方法，所述基于各所述第二模型梯度矩阵计算目标梯度，包括：

获取各所述电子设备上传的矩阵训练数据量，所述矩阵训练数据量为模型训练阶段生成所述第二模型梯度矩阵时的训练数据消耗值；

基于所有所述矩阵训练数据量确定数据量总和；

确定所述矩阵训练数据量和所述数据量总和的比值，将所述比值作为所述电子设备对应的梯度权重；

基于所述梯度权重和所述第二模型梯度矩阵进行加权求和，得到目标梯度。

12、根据权利要求9所述的方法，所述方法还包括：

确定所述本地图像篡改检测模型的当前训练轮次；

若所述当前训练轮次大于或者等于预设轮次，则向各所述电子设备发送训练结束指令，所述训练结束指令用于指示各所述电子设备停止上传所述第二模型梯度矩阵。

13、一种图像篡改检测装置，应用于电子设备，所述装置包括：

梯度处理模块，用于确定本地图像篡改检测模型对应的第一模型梯度矩阵，基于所述第一模型梯度矩阵进行局部梯度差分隐私计算处理得到第二模型梯度矩阵；

梯度发送模块，用于将所述第二模型梯度矩阵上传至服务平台，所述第二模型梯度矩阵用于指示所述服务平台基于至少一个所述电子设备上传的所述第二模型梯度矩阵确定目标模型参数；

数据接收模块，用于接收所述服务平台发送的所述目标模型参数，基于所述目标模型参数更新所述本地图像篡改检测模型得到目标图像篡改检测模型；

图像处理模块，用于基于所述目标图像篡改检测模型对目标隐私图像进行图像检测处理得到所述目标隐私图像对应的目标热力检测图，对所述目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果。

14、一种图像篡改检测装置，应用于服务平台，所述装置包括：

数据接收模块，用于接收至少一个电子设备上传的第二模型梯度矩阵，所述第二模型梯度矩阵由所述电子设备基于本地图像篡改检测模型对应的第一模型梯度矩阵进行局部梯度差分隐私计算处理得到；

数据处理模块，用于基于各所述第二模型梯度矩阵确定目标模型参数；

数据发送模块，用于将所述目标模型参数发送至各所述电子设备，所述目标模型参数用于指示所述电子设备基于所述目标模型参数更新本地图像篡改检测模型得到目标

图像篡改检测模型，并基于所述目标图像篡改检测模型对目标隐私图像进行图像检测处理得到所述目标隐私图像对应的目标热力检测图，对所述目标热力检测图进行篡改识别处理得到所述目标隐私图像对应的篡改识别结果。

15、一种计算机存储介质，所述计算机存储介质存储有多条指令，所述指令适于由处理器加载并执行如权利要求 1~8 或 9~12 任意一项的方法步骤。

16、一种计算机程序产品，所述计算机程序产品存储有至少一条指令，所述至少一条指令由处理器加载并执行如权利要求 1~8 或 9~12 任意一项的方法步骤。

17、一种电子设备，包括：处理器和存储器；其中，所述存储器存储有计算机程序，所述计算机程序适于由所述处理器加载并执行如权利要求 1~8 或 9~12 任意一项的方法步骤。