



(12) 发明专利

(10) 授权公告号 CN 108229591 B

(45) 授权公告日 2020.09.22

(21) 申请号 201810215337.2

G06N 3/02 (2006.01)

(22) 申请日 2018.03.15

(56) 对比文件

(65) 同一申请的已公布的文献号
申请公布号 CN 108229591 A

WO 2017106998 A1, 2017.06.29
CN 106709933 A, 2017.05.24
CN 105069472 A, 2015.11.18

(43) 申请公布日 2018.06.29

朱陶等. 一种基于前向无监督卷积神经网络的人脸表示学习方法.《计算机科学》.2016,第43卷(第6期),303-307.

(73) 专利权人 北京市商汤科技开发有限公司
地址 100084 北京市海淀区中关村东路1号院3号楼7层710-712房间

审查员 祝梦花

(72) 发明人 庞家昊 孙文秀 杨成熙 任思捷
肖瑞超 林惊

(74) 专利代理机构 北京思源智汇知识产权代理有限公司 11657
代理人 毛丽琴

(51) Int. Cl.

G06K 9/62 (2006.01)

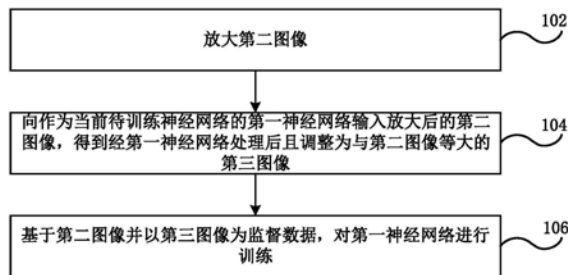
权利要求书3页 说明书13页 附图3页

(54) 发明名称

神经网络自适应训练方法和装置、设备、程序和存储介质

(57) 摘要

本发明实施例公开了一种神经网络自适应训练方法和装置、电子设备、计算机程序和存储介质。其中,方法包括:放大第二图像;向作为当前待训练神经网络的第一神经网络输入放大后的第二图像,得到经第一神经网络处理后且调整为与第二图像等大的第三图像;基于所述第二图像并以所述第三图像为监督数据,对所述第一神经网络进行训练。本发明实施例可以获得更有效的自适应效果。



1. 一种神经网络自适应训练方法,其特征在于,包括:

放大第二图像;

向作为当前待训练神经网络的第一神经网络输入放大后的第二图像,得到经所述第一神经网络处理后且调整为与所述第二图像等大的第三图像;其中,所述第一神经网络具有尺度分集现象;其中,所述尺度分集现象表示所述第三图像,相比较于直接将第二图像输入第一神经网络经过第一神经网络处理后输出的图像,具有更丰富的有意义的高频细节信息;

基于所述第二图像并以所述第三图像为监督数据,对所述第一神经网络进行训练;

所述第一神经网络在作为所述当前待训练的神经网络之前,用于处理针对第一场景的图像处理任务;所述第二图像为对应第二场景的图像;其中,所述第一场景和所述第二场景不同。

2. 根据权利要求1所述的方法,其特征在于,还包括:

基于第一图像并以所述第一图像的标注信息为监督数据,对所述第一神经网络进行训练。

3. 根据权利要求2所述的方法,其特征在于,所述当前待训练的神经网络训练完成后作为针对第二场景的图像处理任务的神经网络。

4. 根据权利要求3所述的方法,其特征在于,所述第一图像为对应所述第一场景的图像。

5. 根据权利要求1所述的方法,其特征在于,所述第二图像为在所述第二场景下的采集图像。

6. 根据权利要求2-4任一项所述的方法,其特征在于,还包括:

获取训练图像集,所述训练图像集包括:至少一所述第一图像,和/或,至少一所述第二图像。

7. 根据权利要求6所述的方法,其特征在于,

所述第一神经网络的训练为多次迭代训练,

任一迭代训练的训练图像集包括:至少一所述第一图像,和/或,至少一所述第二图像;

各次迭代的训练图像集的合集包括:至少一所述第一图像,和至少一所述第二图像。

8. 根据权利要求6所述的方法,其特征在于,所述获取训练图像集包括:

从无标注图像集中选取至少一所述第二图像;和/或,

从标注图像集中选取至少一所述第一图像。

9. 根据权利要求6所述的方法,其特征在于,

所述放大第二图像之前,还包括:确定所述第二图像为无标注图像;和/或,

所述基于第一图像并以所述第一图像的标注为监督数据,对所述第一神经网络进行训练之前,还包括:确定所述第一图像为标注图像。

10. 根据权利要求1-5任一项所述的方法,其特征在于,所述向第一神经网络输入放大后的第二图像,得到经所述第一神经网络处理后且调整为与所述第二图像等大的第三图像,包括:

向第一神经网络输入放大后的第二图像;

经所述第一神经网络对放大后的第二图像进行图像处理;

将图像处理结果调整为与所述第二图像等大的第三图像。

11. 根据权利要求10所述的方法,其特征在於,所述图像处理包括:像素级图像处理。

12. 根据权利要求11所述的方法,其特征在於,所述像素级图像处理包括下列中的任意一种:双目立体匹配、光流估计和图像语义分割。

13. 根据权利要求1-5任一项所述的方法,其特征在於,所述基于所述第二图像并以所述第三图像为监督数据,对所述第一神经网络进行训练,包括:

基于包括有正则化约束项的损失函数确定所述第二图像和所述第三图像之间的损失,基于确定的损失调整所述第一神经网络的网络参数。

14. 根据权利要求13所述的方法,其特征在於,所述正则化约束包括下列中的一种或多种:拉普拉斯矩阵约束和全变分约束。

15. 一种神经网络自适应训练装置,其特征在於,包括:

放大单元,用于放大第二图像;

处理单元,用于向作为当前待训练神经网络的第一神经网络输入放大后的第二图像,得到经所述第一神经网络处理后且调整为与所述第二图像等大的第三图像;其中,所述第一神经网络具有尺度分集现象;其中,所述尺度分集现象表示所述第三图像,相比较于直接将第二图像输入第一神经网络经过第一神经网络处理后输出的图像,具有更丰富的有意义的高频细节信息;

训练单元,用于基于所述第二图像并以所述第三图像为监督数据,对所述第一神经网络进行训练;

所述第一神经网络在作为所述当前待训练的神经网络之前,用于处理针对第一场景的图像处理任务;所述第二图像为对应第二场景的图像;其中,所述第一场景和所述第二场景不同。

16. 根据权利要求15所述的装置,其特征在於,

所述训练单元,还用于基于第一图像并以所述第一图像的标注信息为监督数据,对所述第一神经网络进行训练。

17. 根据权利要求16所述的装置,其特征在於,所述当前待训练的神经网络训练完成后作为针对第二场景的图像处理任务的神经网络。

18. 根据权利要求16所述的装置,其特征在於,所述第一图像为对应所述第一场景的图像。

19. 根据权利要求15所述的装置,其特征在於,所述第二图像为在所述第二场景下的采集图像。

20. 根据权利要求16-18任一项所述的装置,其特征在於,还包括:

获取单元,用于获取训练图像集,所述训练图像集包括:至少一所述第一图像,和/或,至少一所述第二图像。

21. 根据权利要求20所述的装置,其特征在於,

所述第一神经网络的训练为多次迭代训练,

任一迭代训练的训练图像集包括:至少一所述第一图像,和/或,至少一所述第二图像;

各次迭代的训练图像集的合集包括:至少一所述第一图像,和至少一所述第二图像。

22. 根据权利要求20所述的装置,其特征在於,所述获取单元具体用于:

从无标注图像集中选取至少一所述第二图像;和/或,
从标注图像集中选取至少一所述第一图像。

23. 根据权利要求20所述的装置,其特征在于,还包括:

确定单元,用于在所述放大第二图像之前,确定所述第二图像为无标注图像;和/或,在所述基于第一图像并以所述第一图像的标注为监督数据,对所述第一神经网络进行训练之前,确定所述第一图像为标注图像。

24. 根据权利要求15-19任一项所述的装置,其特征在于,所述处理单元,包括:

输入模块,用于向第一神经网络输入放大后的第二图像;

处理模块,用于经所述第一神经网络对放大后的第二图像进行图像处理;

调整模块,用于将图像处理结果调整为与所述第二图像等大的第三图像。

25. 根据权利要求24所述的装置,其特征在于,所述图像处理包括:像素级图像处理。

26. 根据权利要求25所述的装置,其特征在于,所述像素级图像处理包括下列中的任意一种:双目立体匹配、光流估计和图像语义分割。

27. 根据权利要求15-19任一项所述的装置,其特征在于,

所述训练单元,还用于基于包括有正则化约束项的损失函数确定所述第二图像和所述第三图像之间的损失,基于确定的损失调整所述第一神经网络的网络参数。

28. 根据权利要求27所述的装置,其特征在于,所述正则化约束包括下列中的一种或多种:拉普拉斯矩阵约束和全变分约束。

29. 一种电子设备,其特征在于,包括权利要求15-28任一项所述的装置。

30. 一种电子设备,其特征在于,包括:

存储器,用于存储可执行指令;以及

处理器,用于与所述存储器通信以执行所述可执行指令从而完成权利要求1至14任一项所述的方法。

31. 一种计算机存储介质,用于存储计算机可读的指令,其特征在于,所述指令被执行时执行权利要求1至14任一项所述的方法。

神经网络自适应训练方法和装置、设备、程序和存储介质

技术领域

[0001] 本发明属于计算机视觉技术领域,特别是涉及一种神经网络自适应训练方法和装置、电子设备、计算机程序和存储介质。

背景技术

[0002] 随着深度学习技术的不断发展,卷积神经网络在像素级图像处理问题上的应用越来越广泛,这类像素级图像处理问题的特点是:输入的是一幅或者两幅图像,经过图像处理输出的也是一幅图像,并且输出图像与输入图像之间存在着一定的关联。例如:双目立体匹配,其输入的是双目视觉系统采集得到的校准过的左图和右图,输出的是对应的视差图;又例如:图像语义分割,其输入的是一幅图像,输出的是对应的含有语义标注的图像;再例如:光流估计,其输入的是两幅连续的图像,输出的是对应的光流场图像。

发明内容

[0003] 本发明实施例提供一种神经网络自适应训练技术方案。

[0004] 根据本发明实施例的一个方面,提供一种神经网络自适应训练方法,包括:

[0005] 放大第二图像;

[0006] 向作为当前待训练神经网络的第一神经网络输入放大后的第二图像,得到经所述第一神经网络处理后且调整为与所述第二图像等大的第三图像;

[0007] 基于所述第二图像并以所述第三图像为监督数据,对所述第一神经网络进行训练。

[0008] 可选地,在本发明上述任一方法实施例中,还包括:

[0009] 基于第一图像并以所述第一图像的标注信息为监督数据,对所述第一神经网络进行训练。

[0010] 可选地,在本发明上述任一方法实施例中,所述当前待训练的神经网络训练完成后作为针对第二场景的图像处理任务的神经网络;

[0011] 所述第一神经网络在作为所述当前待训练的神经网络之前,用于处理针对第一场景的图像处理任务,其中,所述第一场景和所述第二场景不同。

[0012] 可选地,在本发明上述任一方法实施例中,所述第一图像为对应所述第一场景的图像,所述第二图像为对应所述第二场景的图像。

[0013] 可选地,在本发明上述任一方法实施例中,所述第二图像为在所述第二场景下的采集图像。

[0014] 可选地,在本发明上述任一方法实施例中,还包括:

[0015] 获取训练图像集,所述训练图像集包括:至少一所述第一图像,和/或,至少一所述第二图像

[0016] 可选地,在本发明上述任一方法实施例中,所述第一神经网络的训练为多次迭代训练,

- [0017] 任一迭代训练的训练图像集包括:至少一所述第一图像,和/或,至少一所述第二图像;
- [0018] 各次迭代的训练图像集的合集包括:至少一所述第一图像,和至少一所述第二图像。
- [0019] 可选地,在本发明上述任一方法实施例中,所述获取训练图像集包括:
- [0020] 从无标注图像集中选取至少一所述第二图像;和/或,
- [0021] 从标注图像集中选取至少一所述第一图像。
- [0022] 可选地,在本发明上述任一方法实施例中,所述放大第二图像之前,还包括:确定所述第二图像为无标注图像;和/或,
- [0023] 所述基于第一图像并以所述第一图像的标注为监督数据,对所述第一神经网络进行训练之前,还包括:确定所述第一图像为标注图像。
- [0024] 可选地,在本发明上述任一方法实施例中,所述向第一神经网络输入放大后的第二图像,得到经所述第一神经网络处理后且调整为与所述第二图像等大的第三图像,包括:
- [0025] 向第一神经网络输入放大后的第二图像;
- [0026] 经所述第一神经网络对放大后的第二图像进行图像处理;
- [0027] 将图像处理结果调整为与所述第二图像等大的第三图像。
- [0028] 可选地,在本发明上述任一方法实施例中,所述图像处理包括:像素级图像处理。
- [0029] 可选地,在本发明上述任一方法实施例中,所述像素级图像处理包括下列中的一种或任意组合:双目立体匹配、光流估计和图像语义分割。
- [0030] 可选地,在本发明上述任一方法实施例中,所述基于所述第二图像并以所述第三图像为监督数据,对所述第一神经网络进行训练,包括:
- [0031] 基于包括有正则化约束项的损失函数确定所述第二图像和所述第三图像之间的损失,基于确定的损失调整所述第一神经网络的网络参数。
- [0032] 可选地,在本发明上述任一方法实施例中,所述正则化约束包括下列中的一种或多种:拉普拉斯矩阵约束和全变分约束。
- [0033] 根据本发明实施例的另一个方面,提供一种神经网络自适应训练装置,包括:
- [0034] 放大单元,用于放大第二图像;
- [0035] 处理单元,用于向作为当前待训练神经网络的第一神经网络输入放大后的第二图像,得到经所述第一神经网络处理后且调整为与所述第二图像等大的第三图像;
- [0036] 训练单元,用于基于所述第二图像并以所述第三图像为监督数据,对所述第一神经网络进行训练。
- [0037] 可选地,在本发明上述任一装置实施例中,所述训练单元,还用于基于第一图像并以所述第一图像的标注信息为监督数据,对所述第一神经网络进行训练。可选地,在本发明上述任一装置实施例中,所述当前待训练的神经网络训练完成后作为针对第二场景的图像处理任务的神经网络;
- [0038] 所述第一神经网络在作为所述当前待训练的神经网络之前,用于处理针对第一场景的图像处理任务,其中,所述第一场景和所述第二场景不同。
- [0039] 可选地,在本发明上述任一装置实施例中,所述第一图像为对应所述第一场景的图像,所述第二图像为对应所述第二场景的图像。

[0040] 可选地,在本发明上述任一装置实施例中,所述第二图像为在所述第二场景下的采集图像。

[0041] 可选地,在本发明上述任一装置实施例中,还包括:

[0042] 获取单元,用于获取训练图像集,所述训练图像集包括:至少一所述第一图像,和/或,至少一所述第二图像。

[0043] 可选地,在本发明上述任一装置实施例中,所述第一神经网络的训练为多次迭代训练,

[0044] 任一迭代训练的训练图像集包括:至少一所述第一图像,和/或,至少一所述第二图像;

[0045] 各次迭代的训练图像集的合集包括:至少一所述第一图像,和至少一所述第二图像。

[0046] 可选地,在本发明上述任一装置实施例中,所述获取单元具体用于:

[0047] 从无标注图像集中选取至少一所述第二图像;和/或,

[0048] 从标注图像集中选取至少一所述第一图像。

[0049] 可选地,在本发明上述任一装置实施例中,还包括:

[0050] 确定单元,用于在所述放大第二图像之前,确定所述第二图像为无标注图像;和/或,在所述基于第一图像并以所述第一图像的标注为监督数据,对所述第一神经网络进行训练之前,确定所述第一图像为标注图像。

[0051] 可选地,在本发明上述任一装置实施例中,所述处理单元,包括:

[0052] 输入模块,用于向第一神经网络输入放大后的第二图像;

[0053] 处理模块,用于经所述第一神经网络对放大后的第二图像进行图像处理;

[0054] 调整模块,用于将图像处理结果调整为与所述第二图像等大的第三图像。

[0055] 可选地,在本发明上述任一装置实施例中,所述图像处理包括:像素级图像处理。

[0056] 可选地,在本发明上述任一装置实施例中,所述像素级图像处理包括下列中的一种或任意组合:双目立体匹配、光流估计和图像语义分割。

[0057] 可选地,在本发明上述任一装置实施例中,所述训练单元,还用于基于包括有正则化约束项的损失函数确定所述第二图像和所述第三图像之间的损失,基于确定的损失调整所述第一神经网络的网络参数。

[0058] 可选地,在本发明上述任一装置实施例中,所述正则化约束包括下列中的一种或多种:拉普拉斯矩阵约束和全变分约束。

[0059] 根据本发明实施例的又一个方面,提供一种电子设备,包括上述任一实施例所述的装置。

[0060] 根据本发明实施例的再一个方面,提供一种电子设备,包括:

[0061] 存储器,用于存储可执行指令;以及

[0062] 处理器,用于与所述存储器通信以执行所述可执行指令从而完成上述任一实施例所述的方法。

[0063] 根据本发明实施例的再一个方面,提供一种计算机程序,包括计算机可读代码,当所述计算机可读代码在设备上运行时,所述设备中的处理器执行用于实现上述任一实施例所述方法的指令。

[0064] 根据本发明实施例的再一个方面,提供的一种计算机程序产品,用于存储计算机可读指令,所述指令被执行时使得计算机执行上述任一实施例所述的方法。

[0065] 在一个可选实施方式中,所述计算机程序产品具体为计算机存储介质,在另一个可选实施方式中,所述计算机程序产品具体为软件产品,例如SDK等。

[0066] 基于本发明上述实施例提供的神经网络自适应训练方法和装置、电子设备、计算机程序和存储介质,利用尺度分集获得的具有高频细节信息的图像作为监督数据,对神经网络进行监督训练,可以实现在采集到的新场景下的图像不具有正确的标注信息的情况下,对神经网络的训练,从而可以将适用于某一场景的神经网络泛化到另一场景,并获得良好的效果,由于不需要借助外部的辅助,本发明实施例相比于以往的方法,可以获得更有效的自适应效果。

附图说明

[0067] 构成说明书的一部分的附图描述了本发明的实施例,并且连同描述一起用于解释本发明的原理。

[0068] 参照附图,根据下面的详细描述,可以更加清楚地理解本发明,其中:

[0069] 图1是本发明一些实施例提供的神经网络自适应训练方法的流程图。

[0070] 图2是本发明另一些实施例提供的神经网络自适应训练方法的流程图。

[0071] 图3是本发明一些实施例提供的神经网络自适应训练装置的结构示意图。

[0072] 图4是本发明另一些实施例提供的神经网络自适应训练装置的结构示意图。

[0073] 图5是本发明一些实施例提供的电子设备的结构示意图。

具体实施方式

[0074] 现在将参照附图来详细描述本发明的各种示例性实施例。应注意到:除非另外具体说明,否则在这些实施例中阐述的部件的相对布置、数字表达式和数值不限制本发明的范围。

[0075] 同时,应当明白,为了便于描述,附图中所示出的各个部分的尺寸并不是按照实际的比例关系绘制的。

[0076] 以下对至少一个示例性实施例的描述实际上仅仅是说明性的,决不作为对本发明及其应用或使用的任何限制。

[0077] 对于相关领域普通技术人员已知的技术、方法和设备可能不作详细讨论,但在适当情况下,所述技术、方法和设备应当被视为说明书的一部分。

[0078] 应注意到:相似的标号和字母在下面的附图中表示类似项,因此,一旦某一项在一个附图中被定义,则在随后的附图中不需要对其进行进一步讨论。

[0079] 本发明实施例可以应用于计算机系统/服务器,其可与众多其它通用或专用计算机系统环境或配置一起操作。适于与计算机系统/服务器一起使用的众所周知的计算系统、环境和/或配置的例子包括但不限于:个人计算机系统、服务器计算机系统、瘦客户机、厚客户机、手持或膝上设备、基于微处理器的系统、机顶盒、可编程消费电子产品、网络个人电脑、小型计算机系统、大型计算机系统和包括上述任何系统的分布式云计算技术环境,等等。

[0080] 计算机系统/服务器可以在由计算机系统执行的计算机系统可执行指令(诸如程

序模块)的一般语境下描述。通常,程序模块可以包括例程、程序、目标程序、组件、逻辑、数据结构等等,它们执行特定的任务或者实现特定的抽象数据类型。计算机系统/服务器可以在分布式云计算环境中实施,分布式云计算环境中,任务是由通过通信网络链接的远程处理设备执行的。在分布式云计算环境中,程序模块可以位于包括存储设备的本地或远程计算机系统存储介质上。

[0081] 下面将结合附图对本发明实施例提供的神经网络自适应训练方法的相关流程进行描述。

[0082] 图1是本发明一些实施例提供的神经网络自适应训练方法的流程图。应理解,图1所示的例子仅仅是为了帮助本领域技术人员更好地理解本发明的技术方案,而不理解成对本发明的限定。本领域技术人员可以在图1的基础上进行各种变换,而这种变换也应理解成本发明技术方案的一部分。

[0083] 如图1所示,该方法包括:

[0084] 102,放大第二图像。

[0085] 在一个可选的例子中,第二图像可以是采集图像,可以是在神经网络当前应用的场景下采集的图像。在另一个可选的例子中,第二图像可以是计算机合成的图像,可以通过计算机图形学技术合成的图像,例如:使用游戏引擎合成的图像。本发明实施例对第二图像的类型不做限定。

[0086] 104,向作为当前待训练神经网络的第一神经网络输入放大后的第二图像,得到经第一神经网络处理后且调整为与第二图像等大的第三图像。

[0087] 可选地,第一神经网络可以是卷积神经网络,或者是除卷积神经网络外的其它神经网络,本发明实施例对第一神经网络的类型不做限定。在一个或多个可选的例子中,第一神经网络可以是适用于某一场景的神经网络,或者是根据适用于某一场景的神经网络得到的训练的中间模型。

[0088] 可选地,可以向第一神经网络输入放大后的第二图像,经第一神经网络对放大后的第二图像进行图像处理,将图像处理结果调整为与第二图像等大的第三图像。在一个或多个可选的例子中,可以将图像处理结果缩小与第二图像放大倍数相同倍数,得到第三图像,其中,第二图像的放大倍数可以根据具体情况设定,本发明实施例对此不做限定。在一个可选例子中,缩小图像处理结果的调整,可以在神经网络处理过程中通过降采样处理来实现。在另一个可选的例子中,缩小图像处理结果的调整,可以通过在神经网络输出结果的基础上再进行降采样处理来实现。

[0089] 在一个可选的例子中,第一神经网络可以对图像进行像素级图像处理,例如:像素级图像处理可以是双目立体匹配、光流估计和图像语义分割中的一种或者任意组合。在另一个可选的例子中,第一神经网络可以对图像进行除像素级图像处理外的其它图像处理。本发明实施例对第一神经网络的图像处理方式不做限定。

[0090] 在本发明实施例中,通过将第二图像放大后输入第一神经网络,再将经过第一神经网络处理后输出的图像缩小与放大倍数相同的倍数,得到的第三图像,相比较于直接将第二图像输入第一神经网络,经过第一神经网络处理后输出的图像,通常具有更为丰富的有意义的高频细节信息,我们可以将这一现象称为“尺度分集”(scale diversity)。

[0091] 106,基于第二图像并以第三图像为监督数据,对第一神经网络进行训练。

[0092] 可选地,第一神经网络训练完成后可以作为针对第二场景的图像处理任务的神经网络,第一神经网络在作为当前待训练的神经网络之前,可以用于处理针对第一场景的图像处理任务,其中,第一场景和第二场景不同,例如:第一场景可以是人像照片场景,第二场景可以是自动驾驶场景。

[0093] 通常将适用于某一场景(第一场景)的神经网络直接应用于另一场景(第二场景),很难获得良好的效果,我们可以称这种现象为“泛化误差”(generalization error),需要采集大量带有正确标注信息的新场景下的图像作为训练数据,对神经网络进行监督训练,以得到适用于新场景的神经网络。例如:将应用于人像照片场景下的双目立体匹配卷积神经网络直接应用于自动驾驶场景。然而,在实际操作中,要采集大量带有正确标注信息的新场景下的图像通常非常困难,过程相当繁琐且昂贵。

[0094] 在本发明实施例中,通过将第二图像通过尺度分集获得的第三图像作为监督数据,利用对应于第二场景的第二图像对第一神经网络进行监督训练,可以使第一神经网络学习自身输出的更多高频细节信息,以适用于新场景,从而可以将适用于某一场景的神经网络应用于另一场景,并尽可能地抑制泛化误差。因此,本发明实施例提供的神经网络自适应训练方法也可以称为ZOOM AND LEARN。

[0095] 可选地,第二图像可以是在第二场景下的采集图像,以便于实现在线自适应训练。

[0096] 基于本发明上述实施例提供的神经网络自适应训练方法,通过放大第二图像,向作为当前待训练神经网络的第一神经网络输入放大后的第二图像,得到经第一神经网络处理后且调整为与第二图像等大的第三图像,基于第二图像并以第三图像为监督数据,对第一神经网络进行训练,利用尺度分集获得的具有高频细节信息的图像作为监督数据,对神经网络进行监督训练,可以实现在采集到的新场景下的图像不具有正确的标注信息的情况下,对神经网络的训练,从而可以将适用于某一场景的神经网络泛化到另一场景,并获得良好的效果,由于不需要借助外部的辅助,本发明实施例相比于以往的方法,可以获得更有效的自适应效果。

[0097] 在上述各实施例中,在第一神经网络训练的过程中,还可以基于第一图像并以第一图像的标注信息为监督数据,对第一神经网络进行训练。可选地,第一图像可以是计算机合成的图像,利用这类图像可以使神经网络学习在没有任何噪声和物理误差的理想化情况下的图像处理方法,或者第一图像可以是在神经网络之前应用的场景下采集的图像,在一个可选的例子中,第一图像可以是对应于第一场景的图像,本发明实施例对第一图像的类型不做限定。

[0098] 可选地,对第一神经网络的训练可以为多次迭代训练,对于任一迭代训练的训练图像集可以包括至少一第一图像和/或至少一第二图像,各次迭代的训练图像集的合集可以包括至少一第一图像和至少一第二图像,采用两类训练用图像相结合的方式对神经网络进行训练,可以极大的提高训练的效率,通过实验表明:对于传统采用单一类训练用图像进行监督训练,需要几十万次迭代才能够完成的神经网络,采用两类训练用图像相结合的方式训练,大约一万次左右即可完成。

[0099] 下面将结合图2详细描述对第一神经网络进行一次迭代的相关流程。

[0100] 应理解,图2所示的例子仅仅是为了帮助本领域技术人员更好地理解本发明的技术方案,而不应理解成对本发明的限定。本领域技术人员可以在图2的基础上进行各种变

换,而这种变换也应理解成本发明技术方案的一部分。

[0101] 如图2所述,该方法包括:

[0102] 202,获取训练图像集,训练图像集包括:至少一第一图像和/或至少一第二图像。

[0103] 可选地,可以从无标注图像集中选取至少一第二图像,和/或从标注图像集中选取至少一第一图像。在一个可选的例子中,训练图像集可以仅包括第一图像,此时可以从标注图像集中选取第一图像。在另一个可选的例子中,训练图像集可以仅包括第二图像,此时可以从无标注图像集中选取第二图像。在又一个可选的例子中,训练图像集可以同时包括第一图像和第二图像,此时可以从标注图像集中选取第一图像,从无标注图像集中选取第二图像。

[0104] 204,确定第二图像为无标注图像和/或确定第一图像为标注图像。

[0105] 在具体训练之前,需要对训练图像集中的每一个图像是否为标注图像进行确定;若确定训练图像集中的图像为标注图像,可以确定该图像为第一图像,则执行操作206;若确定训练图像集中的图像为无标注图像,可以确定该图像为第二图像,则执行操作208至212。

[0106] 206,基于第一图像并以第一图像的标注信息为监督数据,对第一神经网络进行训练。

[0107] 208,放大第二图像。

[0108] 210,向作为当前待训练神经网络的第一神经网络输入放大后的第二图像,得到经第一神经网络处理后且调整为与第二图像等大的第三图像。

[0109] 212,基于第二图像并以第三图像为监督数据,对第一神经网络进行训练。

[0110] 在上述各实施例中,在第一神经网络训练的过程中,还可以基于包括有正则化约束项的损失函数确定第二图像和第三图像之间的损失,并基于确定的损失调整第一神经网络的网络参数,包括有正则化约束项的损失函数可以使得神经网络可以有区分地学习第三图像中有用的高频细节,同时抑制错误的泛化误差。可选地,正则化约束可以是拉普拉斯矩阵约束(graph Laplacian regularizer)和全变分约束中的一种或多种,本发明实施例对此不做限定。

[0111] 具体地,拉普拉斯矩阵约束是根据一组图像 S_1, S_2, \dots, S_n 构造的,它具有这样一个特性:若同一条边或图案(称为E)同时出现在图像 S_1, S_2, \dots, S_n 中的大多数图像的同位置,那么最小化这个拉普拉斯矩阵约束将使最后得到的结果图像也具有这条边或图案E。例如:对于双目立体匹配神经网络,基于下面两个观察:

[0112] a. 将一组左图L和右图R放大后输入双目立体匹配神经网络,再将神经网络处理得到的图像缩小,得到图像A;将同样一组左图L和右图R直接输入双目立体匹配神经网络,得到图像B;图像A往往比图像B包含更多的细节信息。

[0113] b. 若将新场景下的图像直接输入到原来场景训练好的神经网络中,得到的结果通常会存在很多错误的细节。

[0114] 拉普拉斯矩阵约束可以由新场景下的左图L、具有精细结果的图像A和具有粗糙结果的图像B一起构造。那么会有下面两种情况产生:

[0115] 1. 若由于前述a,一条正确的边缘C出现在精细结果图像A中,同时边缘C也出现在左图L中(因为C是正确的物体边缘,那么C一定会在左图中出现),那么拉普拉斯矩阵约束将

会使最后的结果具有这条(正确的)边缘C;

[0116] 2.若由于前述b,一条错误的边缘D孤立地出现在精细结果的图像A或粗糙结果的图像B中,由于边缘D只出现了一次,那么拉普拉斯矩阵约束将不会使最后的结果具有这条边缘D。

[0117] 可见,通过使用拉普拉斯矩阵约束可以尽可能地利用a带来的好处,同时抑制b的弱点,而无论是上述哪一种情况,拉普拉斯矩阵约束都可以为最后的训练结果带来好处。

[0118] 在一个或多个可选的例子中,上述各实施例提供的神经网络自适应训练方法可以采用下面的操作来实现:

Algorithm 1 Zoom and learn (ZOLE)

```

1: Input: Pre-trained deep stereo model  $S(\cdot; \Theta^{(0)})$ , training data
    $\{P_i\}_{i=1}^{N_{\text{dom}}}$  and  $\{P_i, D_i\}_{i=N_{\text{dom}}+1}^N$ 
2: Shuffle the training data to form a list  $l$ 
3: for  $k = 0$  to  $k_{\text{max}} - 1$  do
[0119] 4:   for  $b = 1$  to  $n$  do
5:     Draw an index  $i$  from list  $l$ 
6:     if  $i \leq N_{\text{dom}}$  then
7:       Compute  $D_i$ , then compute  $S(P_i; \Theta^{(k)})$  and hence the
         graph Laplacian matrices  $\mathbf{L}_{ij}^{(k)}$ 
8:     end if
9:     Insert  $\{P_i, D_i\}$  to the current batch
10:   end for
11:   Use the formed training batch and the pre-computed Laplacian
[0120]   matrices to perform a step of gradient descent
12:   if  $\text{mod}(k + 1, t) = 0$  then
13:     Perform validation, update  $\Theta^{(\text{bst})}$  and  $v^{(\text{bst})}$  if needed
14:   end if
15: end for
16: Output: model parameter  $\Theta^{(\text{bst})}$ 

```

[0121] 下面对上述操作中的一些关键步骤进行说明:

[0122] Step 1:一共有N组训练数据,其中前 N_{dom} 组训练数据为 $\{P_i\}_{i=1}^{N_{\text{dom}}}$,是新场景(new domain)下的训练数据, P_i 是第*i*个图像,例如:对于双目立体匹配,每一组训练数据包含双目立体匹配用到的一张左图和一张右图,假设 P_i 对应的真实数据(ground-truth)的视差图(disparity map)为 D_i ,而 D_i 是未知的。后 $N - N_{\text{dom}}$ 组训练数据为 $\{P_i, D_i\}_{i=N_{\text{dom}}+1}^N$,其中一组训练数据的左图、右图以及对应的视差图均是已知的。

[0123] Step 3: k_{max} 为神经网络训练的最大迭代次数, k 为当前的迭代次数。

[0124] Step 4: n 为每一个训练批次(batch)中所包含数据的数量。

[0125] Step 7:随机抽取一组新场景下的训练数据,由于其真实数据(ground-truth)的

视差图是未知的,因此按照本发明实施例提供的方法(即放大、CNN处理、缩小的方式)为其生成伪真实数据(ground-truth)的视差图 D_i 。此外, $S(P_i; \Theta^{(k)})$ 是将左图和右图(即 P_i)直接输入神经网络得到的一个估算的视差图,这个直接估算的视差图、伪真实数据(ground-truth)的视差图 D_i 以及对应的左图,三者一起可以用于构建拉普拉斯矩阵 $L_{ij}^{(k)}$ 。

[0126] Step 9:无论在之前的步骤中是否随机抽取到新场景下的训练数据,都会有左图和右图及对应的真实数据(ground-truth)的视差图,于是将当前的这组数据 $\{P_i, D_i\}$ 加入到当前的训练批次中。

[0127] Step 11:使用当前一个批次(batch)的训练数据,以及预先确定好的拉普拉斯矩阵 $L_{ij}^{(k)}$ 对神经网络进行一次梯度下降(或称一次迭代),对应的神经网络参数 $\Theta^{(k)}$ 将会更新。

[0128] Step 13:每经过 t 次迭代,就对当前得到的神经网络参数 $\Theta^{(k)}$ 进行一次验证,得到一个验证的指标 v 。假设当前最优的网络参数为 $\Theta^{(bst)}$,对应的(最优的)效果指标为 $v^{(bst)}$ 。那么在对新的模型 $\Theta^{(k)}$ 验证之后,如果发现指标 v 是优于 $v^{(bst)}$ 的话,就把 $\Theta^{(k)}$ 赋值给 $\Theta^{(bst)}$,以及把 v 赋值给 $v^{(bst)}$ 。

[0129] 其中,根据不同的问题和需要,可以采用不同的验证方法进行验证,例如:对于双目立体匹配,如果验证图像只包含第二场景下的左图和右图(或者主图和辅图),可以利用神经网络的输出进行图像间一致性的验证;如果验证图像包括第二场景下有标注的图像,可以直接利用有标注的图像进行验证,本发明实施例对验证过程采用的验证方法不做限定。在一个可选的例子中,可以利用验证数据通过当前神经网络获得视差图,再将该视差图与已知的右图合成出左图,通过将合成出来的左图与真正的左图进行比较,计算平均的信噪比数值进行验证,若新的信噪比比现存的最优信噪比更好,我们将更新最优信噪比并保存当前神经网络的参数。

[0130] 本发明实施例提供的神经网络自适应训练方法,可以应用于具有双目摄像头的智能手机等移动终端,训练神经网络用于估计场景深度,也可以应用于自动驾驶中求取深度的双摄像头模块,训练神经网络估计场景深度;还可以应用于智能手机的摄像头,训练神经网络用于估计光流的场景;以及自动驾驶系统中的摄像头,训练神经网络用于图像语义分割。

[0131] 图3是本发明一些实施例提供的神经网络自适应训练装置的结构示意图。应理解,图3所示的例子仅仅是为了帮助本领域技术人员更好地理解本发明的技术方案,而不应理解成对本发明的限定。本领域技术人员可以在图3的基础上进行各种变换,而这种变换也应理解成本发明技术方案的一部分。

[0132] 如图3所示,该装置包括:放大单元310、处理单元320和训练单元330。其中,放大单元310,用于放大第二图像。

[0133] 在一个可选的例子中,第二图像可以是采集图像,可以是在神经网络当前应用的场景下采集的图像。在另一个可选的例子中,第二图像可以是计算机合成的图像,可以通过计算机图形学技术合成的图像,例如:使用游戏引擎合成的图像。本发明实施例对第二图像的类型不做限定。

[0134] 处理单元320,用于向作为当前待训练神经网络的第一神经网络输入放大后的第二图像,得到经第一神经网络处理后且调整为与第二图像等大的第三图像。

[0135] 可选地,第一神经网络可以是卷积神经网络,或者是除卷积神经网络外的其它神经网络,本发明实施例对第一神经网络的类型不做限定。在一个或多个可选的例子中,第一神经网络可以是适用于某一场景的神经网络,或者是根据适用于某一场景的神经网络得到的训练的中间模型。

[0136] 可选地,处理单元320可以包括:输入模块、处理模块和调整模块,输入模块可以用于向第一神经网络输入放大后的第二图像,处理模块可以用于经第一神经网络对放大后的第二图像进行图像处理,调整模块可以用于将图像处理结果调整为与第二图像等大的第三图像。在一个或多个可选的例子中,可以将图像处理结果缩小与第二图像放大倍数相同倍数,得到第三图像,其中,第二图像的放大倍数可以根据具体情况设定,本发明实施例对此不做限定。在一个可选例子中,缩小图像处理结果的调整,可以在神经网络处理过程中通过降采样处理来实现。在另一个可选的例子中,缩小图像处理结果的调整,可以通过在神经网络输出结果的基础上再进行降采样处理来实现。

[0137] 在一个可选的例子中,第一神经网络可以对图像进行像素级图像处理,例如:像素级图像处理可以是双目立体匹配、光流估计和图像语义分割中的一种或者任意组合。在另一个可选的例子中,第一神经网络可以对图像进行除像素级图像处理外的其它图像处理。本发明实施例对第一神经网络的图像处理方式不做限定。

[0138] 训练单元330,用于基于第二图像并以第三图像为监督数据,对第一神经网络进行训练。

[0139] 可选地,第一神经网络训练完成后可以作为针对第二场景的图像处理任务的神经网络,第一神经网络在作为当前待训练的神经网络之前,可以用于处理针对第一场景的图像处理任务,其中,第一场景和第二场景不同,例如:第一场景可以是人像照片场景,第二场景可以是自动驾驶场景。

[0140] 可选地,第二图像可以是在第二场景下的采集图像,以便于实现在线自适应训练。

[0141] 基于本发明上述实施例提供的神经网络自适应训练装置,通过放大第二图像,向作为当前待训练神经网络的第一神经网络输入放大后的第二图像,得到经第一神经网络处理后且调整为与第二图像等大的第三图像,基于第二图像并以第三图像为监督数据,对第一神经网络进行训练,利用尺度分集获得的具有高频细节信息的图像作为监督数据,对神经网络进行监督训练,可以实现在采集到的新场景下的图像不具有正确的标注信息的情况下,对神经网络的训练,从而可以将适用于某一场景的神经网络泛化到另一场景,并获得良好的效果,由于不需要借助外部的辅助,本发明实施例相比于以往的方法,可以获得更有效的自适应效果。

[0142] 在上述各实施例中,训练单元还可以基于第一图像并以第一图像的标注信息为监督数据,对第一神经网络进行训练。可选地,第一图像可以是计算机合成的图像,利用这类图像可以使神经网络学习在没有任何噪声和物理误差的理想化情况下的图像处理方法,或者第一图像可以是在神经网络之前应用的场景下采集的图像,在一个可选的例子中,第一图像可以是对应于第一场景的图像,本发明实施例对第一图像的类型不做限定。

[0143] 可选地,对第一神经网络的训练可以为多次迭代训练,对于任一迭代训练的训练

图像集可以包括至少一第一图像和/或至少一第二图像,各次迭代的训练图像集的合集可以包括至少一第一图像和至少一第二图像,采用两类训练用图像相结合的方式对神经网络进行训练,可以极大的提高训练的效率,通过实验表明:对于传统采用单一类训练用图像进行监督训练,需要几十万次迭代才能够完成的神经网络,采用两类训练用图像相结合的方式训练,大约一万次左右即可完成。

[0144] 图4是本发明另一些实施例提供的神经网络自适应训练装置的结构示意图。应理解,图4所示的例子仅仅是为了帮助本领域技术人员更好地理解本发明的技术方案,而不应理解成对本发明的限定。本领域技术人员可以在图4的基础上进行各种变换,而这种变换也应理解成本发明技术方案的一部分。

[0145] 如图4所述,该装置包括:获取单元410、确定单元420、放大单元430、处理单元440和训练单元450。其中,获取单元410,用于获取训练图像集,训练图像集包括:至少一第一图像和/或至少一第二图像。

[0146] 可选地,获取单元410可以从无标注图像集中选取至少一第二图像,和/或从标注图像集中选取至少一第一图像。在一个可选的例子中,训练图像集可以仅包括第一图像,此时获取单元410可以从标注图像集中选取第一图像。在另一个可选的例子中,训练图像集可以仅包括第二图像,此时获取单元410可以从无标注图像集中选取第二图像。在又一个可选的例子中,训练图像集可以同时包括第一图像和第二图像,此时获取单元410可以从标注图像集中选取第一图像,从无标注图像集中选取第二图像。

[0147] 确定单元420,用于确定第二图像为无标注图像和/或确定第一图像为标注图像。

[0148] 在具体训练之前,确定单元420需要对训练图像集中的每一个图像是否为标注图像进行确定;若确定训练图像集中的图像为标注图像,可以确定该图像为第一图像,则通过训练单元450直接对神经网络进行监督训练;若确定训练图像集中的图像为无标注图像,可以确定该图像为第二图像,则通过执行放大单元430和处理单元440获取第三图像,利用第三图像通过训练单元450对神经网络进行监督训练。

[0149] 放大单元430,用于放大第二图像。

[0150] 处理单元440,用于向作为当前待训练神经网络的第一神经网络输入放大后的第二图像,得到经第一神经网络处理后且调整为与第二图像等大的第三图像。

[0151] 训练单元450,用于基于第二图像并以第三图像为监督数据,对第一神经网络进行训练;以及基于第一图像并以第一图像的标注信息为监督数据,对第一神经网络进行训练。

[0152] 在上述各实施例中,训练单元还可以基于包括有正则化约束项的损失函数确定第二图像和第三图像之间的损失,并基于确定的损失调整第一神经网络的网络参数,包括有正则化约束项的损失函数可以使得神经网络可以有区分地学习第三图像中有用的高频细节,同时抑制错误的泛化误差。可选地,正则化约束可以是拉普拉斯矩阵约束(graph Laplacian regularizer)和全变分约束中的一种或多种,本发明实施例对此不做限定。

[0153] 本发明实施例提供的神经网络自适应训练装置,可以应用于具有双目摄像头的智能手机等移动终端,训练神经网络用于估计场景深度,也可以应用于自动驾驶中求取深度的双摄像头模块,训练神经网络估计场景深度;还可以应用于智能手机的摄像头,训练神经网络用于估计光流的场景;以及自动驾驶系统中的摄像头,训练神经网络用于图像语义分割。

[0154] 另外,本发明实施例还提供了一种电子设备,例如可以是移动终端、个人确定机(PC)、平板电脑、服务器等,该电子设备设置有本发明上述任一实施例的图像处理装置。

[0155] 本发明实施例还提供了一种电子设备,例如可以是移动终端、个人确定机(PC)、平板电脑、服务器等。下面参考图5,其示出了适于用来实现本申请实施例的终端设备或服务器的电子设备500的结构示意图:如图5所示,确定机系统500包括一个或多个处理器、通信部等,所述一个或多个处理器例如:一个或多个中央处理单元(CPU)501,和/或一个或多个图像处理器(GPU)513等,处理器可以根据存储在只读存储器(ROM)502中的可执行指令或者从存储部分508加载到随机访问存储器(RAM)503中的可执行指令而执行各种适当的动作和处理。通信部512可包括但不限于网卡,所述网卡可包括但不限于IB(Infiniband)网卡。

[0156] 处理器可与只读存储器502和/或随机访问存储器530中通信以执行可执行指令,通过总线504与通信部512相连、并经通信部512与其他目标设备通信,从而完成本申请实施例提供的任一项方法对应的操作,例如,放大第二图像;向作为当前待训练神经网络的第一神经网络输入放大后的第二图像,得到经所述第一神经网络处理后且调整为与所述第二图像等大的第三图像;基于所述第二图像并以所述第三图像为监督数据,对所述第一神经网络进行训练。

[0157] 此外,在RAM503中,还可存储有装置操作所需的各种程序和数据。CPU501、ROM502以及RAM503通过总线504彼此相连。在有RAM503的情况下,ROM502为可选模块。RAM503存储可执行指令,或在运行时向ROM502中写入可执行指令,可执行指令使处理器501执行上述通信方法对应的操作。输入/输出(I/O)接口505也连接至总线504。通信部512可以集成设置,也可以设置为具有多个子模块(例如多个IB网卡),并在总线链接上。

[0158] 以下部件连接至I/O接口505:包括键盘、鼠标等的输入部分506;包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分507;包括硬盘等的存储部分508;以及包括诸如LAN卡、调制解调器等的网络接口卡的通信部分509。通信部分509经由诸如因特网的网络执行通信处理。驱动器510也根据需要连接至I/O接口505。可拆卸介质511,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器510上,以便于从其上读出的确定机程序根据需要被安装入存储部分508。

[0159] 需要说明的,如图5所示的架构仅为一种可选实现方式,在具体实践过程中,可根据实际需要对上述图5的部件数量和类型进行选择、删减、增加或替换;在不同功能部件设置上,也可采用分离设置或集成设置等实现方式,例如GPU和CPU可分离设置或者可将GPU集成在CPU上,通信部可分离设置,也可集成设置在CPU或GPU上,等等。这些可替换的实施方式均落入本发明公开的保护范围。

[0160] 特别地,根据本发明的实施例,上文参考流程图描述的过程可以被实现为计算机软件程序。例如,本发明的实施例包括一种计算机程序产品,其包括有形地包含在机器可读介质上的计算机程序,计算机程序包含用于执行流程图所示的方法的程序代码,程序代码可包括对应执行本申请实施例提供的方法步骤对应的指令,例如,放大第二图像;向作为当前待训练神经网络的第一神经网络输入放大后的第二图像,得到经所述第一神经网络处理后且调整为与所述第二图像等大的第三图像;基于所述第二图像并以所述第三图像为监督数据,对所述第一神经网络进行训练。在这样的实施例中,该计算机程序可以通过通信部分509从网络上被下载和安装,和/或从可拆卸介质511被安装。在该计算机程序被中央处理单

元(CPU)501执行时,执行本申请的方法中限定的上述功能。

[0161] 在一个或多个可选实施方式中,本发明实施例还提供了一种计算机程序产品,用于存储计算机可读指令,该指令被执行时使得计算机执行上述任一可能的实现方式中的神经网络自适应训练方法。

[0162] 该计算机程序产品可以具体通过硬件、软件或其结合的方式实现。在一个可选例子中,该计算机程序产品具体体现为计算机存储介质,在另一个可选例子中,该计算机程序产品具体体现为软件产品,例如软件开发包(Software Development Kit,SDK)等等。

[0163] 在一个或多个可选实施方式中,本发明实施例还提供了一种神经网络自适应训练方法及其对应的装置和电子设备、计算机存储介质、计算机程序以及计算机程序产品,其中,该方法包括:第一装置向第二装置发送训练指示,该指示使得第二装置执行上述任一可能的实施例中的神经网络自适应训练方法;第一装置接收第二装置发送的训练结果。

[0164] 在一些实施例中,该训练指示可以具体为调用指令,第一装置可以通过调用的方式指示第二装置执行训练,相应地,响应于接收到调用指令,第二装置可以执行上述神经网络自适应训练方法中的任意实施例中的步骤和/或流程。

[0165] 应理解,本发明实施例中的“第一”、“第二”等术语仅仅是为了区分,而不应理解成对本发明实施例的限定。

[0166] 还应理解,在本发明中,“多个”可以指两个或两个以上,“至少一个”可以指一个、两个或两个以上。

[0167] 还应理解,对于本发明中提及的任一部件、数据或结构,在没有明确限定或者在前后文给出相反启示的情况下,一般可以理解为一个或多个。

[0168] 还应理解,本发明对各个实施例的描述着重强调各个实施例之间的不同之处,其相同或相似之处可以相互参考,为了简洁,不再一一赘述。

[0169] 可能以许多方式来实现本发明的方法和装置、设备。例如,可通过软件、硬件、固件或者软件、硬件、固件的任何组合来实现本发明的方法和装置、设备。用于方法的步骤的上述顺序仅是为了进行说明,本发明的方法的步骤不限于以上具体描述的顺序,除非以其它方式特别说明。此外,在一些实施例中,还可将本发明实施为记录在记录介质中的程序,这些程序包括用于实现根据本发明的方法的机器可读指令。因而,本发明还覆盖存储用于执行根据本发明的方法的程序的记录介质。

[0170] 本发明的描述是为了示例和描述起见而给出的,而并不是无遗漏的或者将本发明限于所公开的形式。很多修改和变化对于本领域的普通技术人员而言是显然的。选择和描述实施例是为了更好说明本发明的原理和实际应用,并且使本领域的普通技术人员能够理解本发明从而设计适于特定用途的带有各种修改的各种实施例。

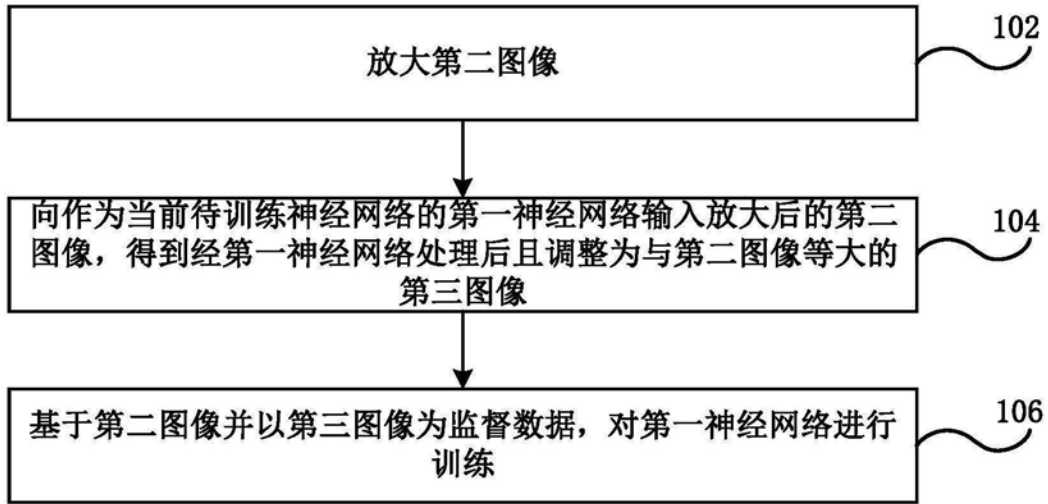


图1

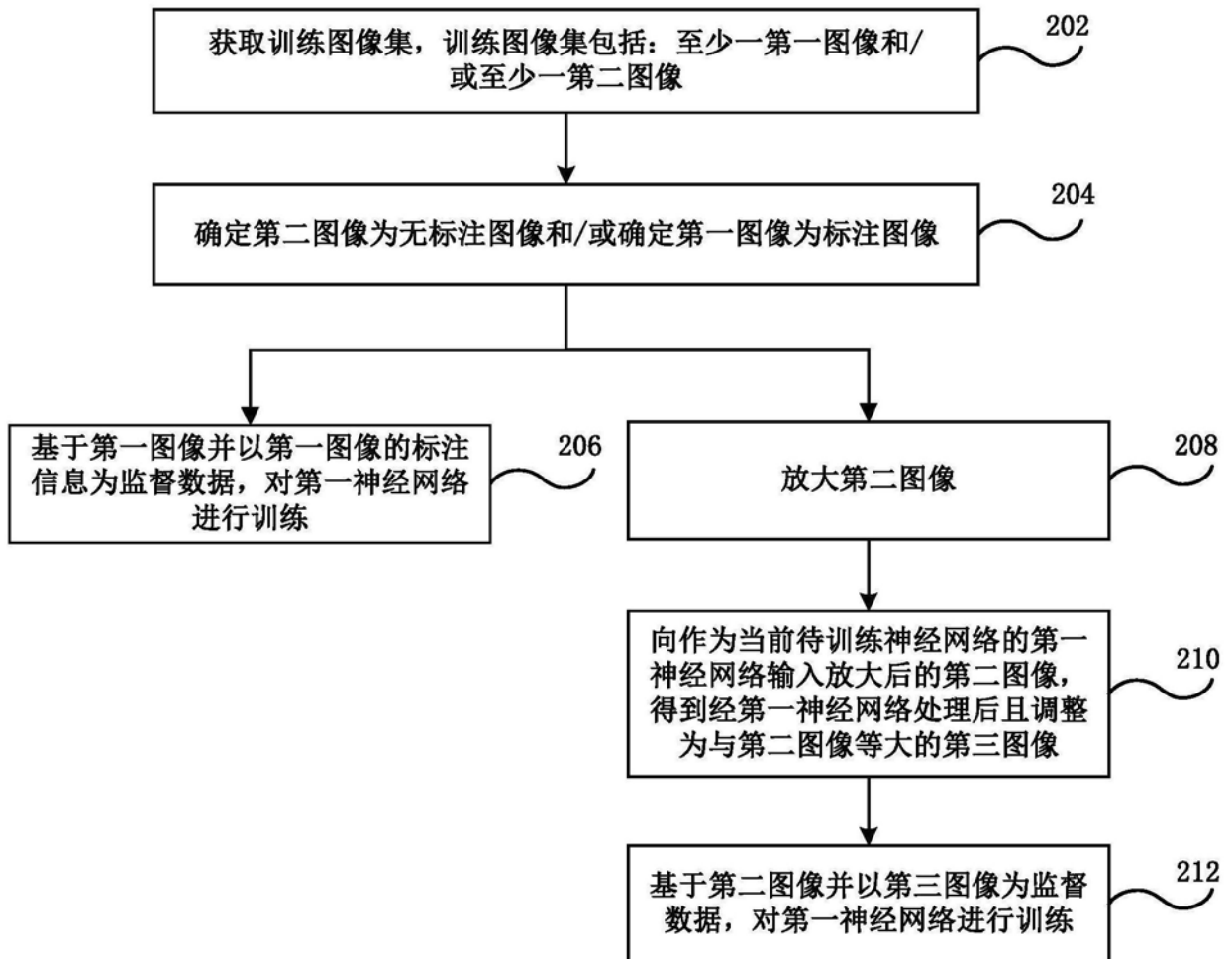


图2

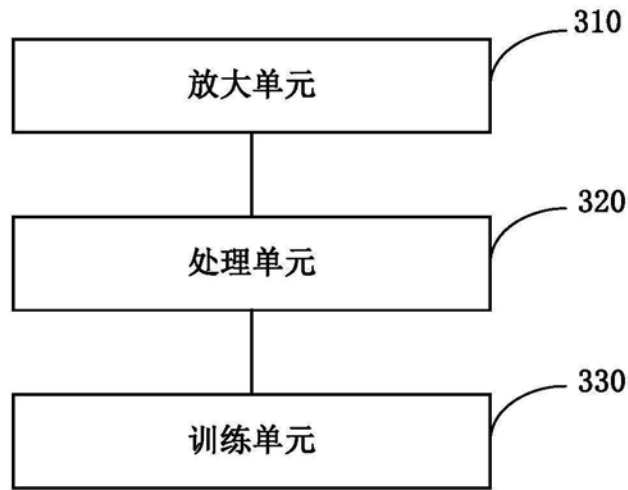


图3

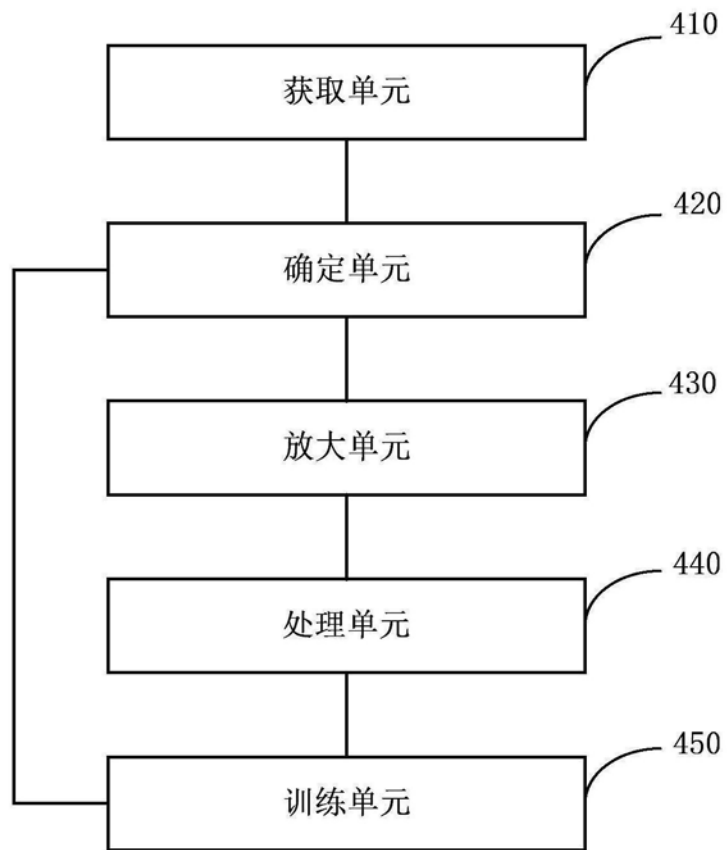


图4

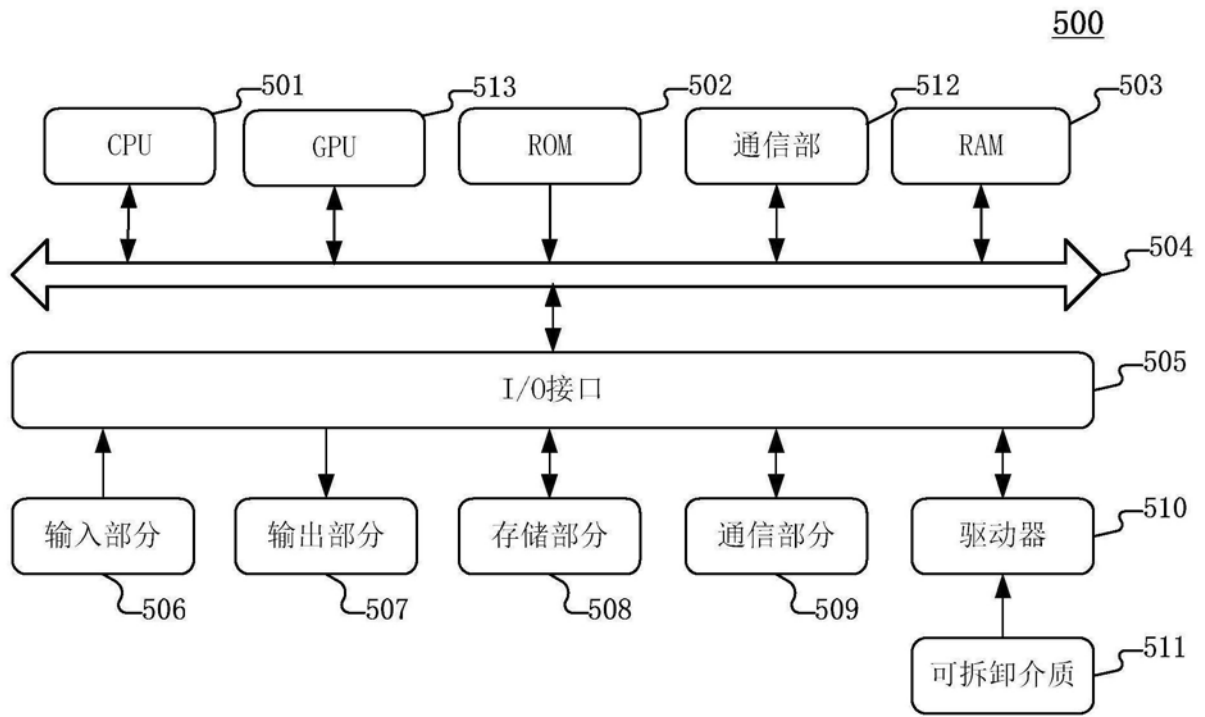


图5