



(12) 发明专利申请

(10) 申请公布号 CN 117015968 A

(43) 申请公布日 2023. 11. 07

(21) 申请号 202280019006.6

(22) 申请日 2022.01.27

(30) 优先权数据

17/198,813 2021.03.11 US

(85) PCT国际申请进入国家阶段日

2023.09.04

(86) PCT国际申请的申请数据

PCT/US2022/014143 2022.01.27

(87) PCT国际申请的公布数据

W02022/191933 EN 2022.09.15

(71) 申请人 高通股份有限公司

地址 美国加利福尼亚

(72) 发明人 R·普里扎 T·S·科亨

(74) 专利代理机构 永新专利商标代理有限公司

72002

专利代理师 戴开良

(51) Int.Cl.

H04N 19/117 (2006.01)

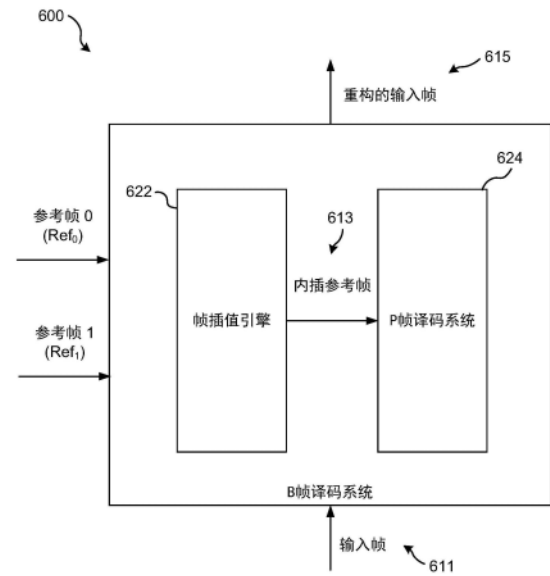
权利要求书6页 说明书38页 附图20页

(54) 发明名称

使用P帧译码系统的学习B帧译码

(57) 摘要

本文描述了用于处理视频数据的技术,例如通过使用单向译码系统和内插参考帧来执行学习的双向译码。例如,处理可以包括:获得第一参考帧和第二参考帧。该处理可以包括:至少部分地通过在第一参考帧和第二参考帧之间执行插值来生成第三参考帧。该处理可以包括:例如通过估计输入帧和第三参考帧之间的运动,基于第三参考帧对输入帧执行单向帧间预测,并至少部分地通过基于估计的运动,变形第三参考帧一个或多个像素来生成变形帧。该处理可以包括:基于变形帧和预测残差来生成表示输入帧的重构帧,重构帧包括双向预测帧。



1. 一种用于处理视频数据的系统,所述系统包括:
双向译码引擎,其包括插值引擎和单向译码引擎,其中:
所述插值引擎被配置为:
获得第一参考帧和第二参考帧;以及
至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成第三参考帧;以及
所述单向译码引擎被配置为:
基于所述第三参考帧,对输入帧执行单向帧间预测。
2. 根据权利要求1所述的系统,其中,所述单向译码引擎包括运动引擎和变形引擎,并且其中:
所述运动引擎被配置为估计所述输入帧和所述第三参考帧之间的运动;以及
所述变形引擎被配置为至少部分地通过基于所述估计的运动来变形所述第三参考帧的一个或多个像素,以生成变形帧。
3. 根据权利要求2所述的系统,其中,所述单向译码引擎还包括残差引擎,并且其中:
所述残差引擎被配置为:
至少部分地通过确定所述输入帧和所述变形帧之间的差异来确定残差;以及
使用所述残差来生成预测残差。
4. 根据权利要求3所述的系统,还包括存储器和耦合到所述存储器的至少一个处理器,其中,所述单向译码引擎被配置为:
基于所述变形帧和所述预测残差,生成表示所述输入帧的重构帧,所述重构帧包括双向预测帧;以及
其中,所述至少一个处理器被配置为使所述重构帧被存储在所述存储器中。
5. 根据权利要求4所述的系统,其中,所述至少一个处理器包括神经处理单元(NPU)。
6. 根据权利要求3所述的系统,其中,所述运动引擎包括第一神经网络,并且所述残差引擎包括第二神经网络。
7. 根据权利要求6所述的系统,其中,所述第一神经网络包括第一自动编码器,并且所述第二神经网络包括第二自动编码器。
8. 根据权利要求1所述的系统,其中,所述单向译码引擎被配置为:
获得第四参考帧;
估计另外输入帧和所述第四参考帧之间的运动;
至少部分地通过基于所述估计的运动来变形所述第四参考帧的一个或多个像素,生成另外的变形帧;以及
基于所述另外的变形帧和另外的预测残差,生成表示所述另外输入帧的重构帧,表示所述另外输入帧的所述重构帧包括单向预测帧。
9. 根据权利要求1所述的系统,其中,为了至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成所述第三参考帧,所述插值引擎被配置为:
确定表示从所述第一参考帧到所述第三参考帧的像素运动的第一组运动信息;
确定表示从所述第二参考帧到所述第三参考帧的像素运动的第二组运动信息;
至少部分地通过使用所述第一组运动信息对所述第一参考帧执行变形函数来生成第

一变形信息;

至少部分地通过使用所述第二组运动信息对所述第二参考帧执行所述变形函数来生成第二变形信息;以及

基于所述第一变形信息和所述第二变形信息生成所述第三参考帧。

10. 根据权利要求9所述的系统,其中,所述第一组运动信息和所述第二组运动信息是基于所述第一参考帧和所述第二参考帧之间的像素运动来确定的。

11. 根据权利要求9所述的系统,其中,所述第一组运动信息包括第一光流图,并且所述第二组运动信息包括第二光流图。

12. 根据权利要求9所述的系统,其中,所述变形函数包括双线性插值函数。

13. 根据权利要求2所述的系统,其中,所述输入帧与所述第三参考帧之间的所述估计的运动包括光流数据。

14. 根据权利要求2所述的系统,其中,所述输入帧与所述第三参考帧之间的所述估计的运动包括动态卷积数据。

15. 根据权利要求2所述的系统,其中,所述输入帧与所述第三参考帧之间的所述估计的运动包括基于块的运动数据。

16. 根据权利要求1所述的系统,其中,所述系统是移动设备、扩展现实设备或电视之一。

17. 根据权利要求1所述的系统,还包括:

被配置为捕获一个或多个视频帧的显示器和相机中的至少一个。

18. 一种用于处理视频数据的方法,所述方法包括:

获得第一参考帧和第二参考帧;

至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成第三参考帧;以及

基于所述第三参考帧对输入帧执行单向帧间预测。

19. 根据权利要求18所述的方法,还包括:

估计所述输入帧和所述第三参考帧之间的运动;以及

至少部分地通过基于所述估计的运动,变形所述第三参考帧的一个或多个像素,来生成变形帧。

20. 根据权利要求19所述的方法,还包括:

至少部分地通过确定所述输入帧和所述变形帧之间的差异来确定残差;以及

使用所述残差生成预测残差。

21. 根据权利要求20所述的方法,还包括:

基于所述变形帧和所述预测残差,生成表示所述输入帧的重构帧,所述重构帧包括双向预测帧;以及

使所述重构帧存储在存储器中。

22. 根据权利要求20所述的方法,其中,所述运动是使用第一自动编码器估计的,并且其中所述预测残差是使用第二自动编码器生成的。

23. 根据权利要求18所述的方法,还包括:

获得第四参考帧;

估计另外输入帧和所述第四参考帧之间的运动；

至少部分地通过基于所述估计的运动来变形所述第四参考帧的一个或多个像素，生成另外的变形帧；以及

基于所述另外的变形帧和另外的预测残差，生成表示所述另外输入帧的重构帧，表示所述另外输入帧的所述重构帧包括单向预测帧。

24. 根据权利要求18所述的方法，还包括：

确定表示从所述第一参考帧到所述第三参考帧的像素运动的第一组运动信息；

确定表示从所述第二参考帧到所述第三参考帧的像素运动的第二组运动信息；

至少部分地通过使用所述第一组运动信息对所述第一参考帧执行变形函数来生成第一变形信息；

至少部分地通过使用所述第二组运动信息对所述第二参考帧执行所述变形函数来生成第二变形信息；以及

基于所述第一变形信息和所述第二变形信息生成所述第三参考帧。

25. 根据权利要求24所述的方法，其中，所述第一组运动信息和所述第二组运动信息是基于所述第一参考帧和所述第二参考帧之间的像素运动来确定的。

26. 根据权利要求24所述的方法，其中，所述第一组运动信息包括第一光流图，并且所述第二组运动信息包括第二光流图。

27. 根据权利要求24所述的方法，其中，所述变形函数包括双线性插值函数。

28. 根据权利要求19所述的方法，其中，所述输入帧与所述第三参考帧之间的所述估计的运动包括光流数据。

29. 根据权利要求19所述的方法，其中，所述输入帧与所述第三参考帧之间的所述估计的运动包括动态卷积数据。

30. 根据权利要求19所述的方法，其中，所述输入帧与所述第三参考帧之间的所述估计的运动包括基于块的运动数据。

31. 一种其上存储有指令的非临时性计算机可读介质，当所述指令被一个或多个处理器执行时，使所述一个或多个处理器执行以下操作：

获得第一参考帧和第二参考帧；

至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成第三参考帧；以及

基于所述第三参考帧对输入帧执行单向帧间预测。

32. 根据权利要求31所述的非临时性计算机可读介质，还包括当被所述一个或多个处理器执行时，使得所述一个或多个处理器执行以下操作的指令：

估计所述输入帧和所述第三参考帧之间的运动；以及

至少部分地通过基于所述估计的运动，变形所述第三参考帧的一个或多个像素，来生成变形帧。

33. 根据权利要求32所述的非临时性计算机可读介质，还包括当被所述一个或多个处理器执行时，使得所述一个或多个处理器执行以下操作的指令：

至少部分地通过确定所述输入帧和所述变形帧之间的差异来确定残差；以及

使用所述残差生成预测残差。

34. 根据权利要求33所述的非临时性计算机可读介质,还包括当被所述一个或多个处理器执行时,使得所述一个或多个处理器执行以下操作的指令:

基于所述变形帧和所述预测残差,生成表示所述输入帧的重构帧,所述重构帧包括双向预测帧;以及

使所述重构帧存储在存储器中。

35. 根据权利要求33所述的非临时性计算机可读介质,其中,所述运动是使用第一自动编码器估计的,并且其中所述预测残差是使用第二自动编码器生成的。

36. 根据权利要求31所述的非临时性计算机可读介质,还包括当被所述一个或多个处理器执行时,使得所述一个或多个处理器执行以下操作的指令:

获得第四参考帧;

估计另外输入帧和所述第四参考帧之间的运动;

至少部分地通过基于所述估计的运动来变形所述第四参考帧的一个或多个像素,生成另外的变形帧;以及

基于所述另外的变形帧和另外的预测残差,生成表示所述另外输入帧的重构帧,表示所述另外输入帧的所述重构帧包括单向预测帧。

37. 根据权利要求31所述的非临时性计算机可读介质,还包括当被所述一个或多个处理器执行时,使得所述一个或多个处理器执行以下操作的指令:

确定表示从所述第一参考帧到所述第三参考帧的像素运动的第一组运动信息;

确定表示从所述第二参考帧到所述第三参考帧的像素运动的第二组运动信息;

至少部分地通过使用所述第一组运动信息对所述第一参考帧执行变形函数来生成第一变形信息;

至少部分地通过使用所述第二组运动信息对所述第二参考帧执行所述变形函数来生成第二变形信息;以及

基于所述第一变形信息和所述第二变形信息生成所述第三参考帧。

38. 一种用于处理视频数据的系统,所述系统包括:

存储器;以及

耦合到所述存储器的至少一个处理器,所述至少一个处理器被配置为:

获得内插参考帧和表示输入帧与所述内插参考帧之间的运动的运动信息;

基于所述运动信息,估计所述输入帧和所述内插参考帧之间的所述运动;

至少部分地通过基于所述估计的运动,对所述内插参考帧的一个或多个像素进行变形来生成变形帧;以及

基于所述变形帧和预测残差来生成表示所述输入帧的重构帧,所述重构帧包括双向预测帧。

39. 根据权利要求38所述的系统,其中,所述至少一个处理器被配置为:

获得第一参考帧和第二参考帧;以及

至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成所述内插参考帧。

40. 根据权利要求38所述的系统,其中,所述至少一个处理器被配置为:

从比特流中获取残差;以及

基于所述获取的残差来生成所述预测残差。

41. 根据权利要求40所述的系统,其中,所述运动信息是从所述比特流中获得的。

42. 根据权利要求38所述的系统,其中,所述至少一个处理器被配置为:

获得第四参考帧;

估计另外输入帧和所述第四参考帧之间的运动;

至少部分地通过基于所述估计的运动,变形所述第四参考帧的一个或多个像素来生成另外的变形帧;以及

基于所述另外的变形帧和另外的预测残差,生成表示所述另外输入帧的重构帧,表示所述另外输入帧的所述重构帧包括单向预测帧。

43. 一种用于处理视频数据的方法,所述方法包括:

获得内插参考帧和表示输入帧与所述内插参考帧之间的运动的运动信息;

基于所述运动信息,估计所述输入帧和所述内插参考帧之间的所述运动;

至少部分地通过基于所述估计的运动,对所述内插参考帧的一个或多个像素进行变形来生成变形帧;以及

基于所述变形帧和预测残差来生成表示所述输入帧的重构帧,所述重构帧包括双向预测帧。

44. 根据权利要求43所述的方法,还包括:

获得第一参考帧和第二参考帧;以及

至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成所述内插参考帧。

45. 根据权利要求43所述的方法,还包括:

从比特流中获取残差;以及

基于所获取的残差来生成所述预测残差。

46. 根据权利要求45所述的方法,其中,所述运动信息是从所述比特流中获得的。

47. 根据权利要求43所述的方法,还包括:

获得第四参考帧;

估计另外输入帧和所述第四参考帧之间的运动;

至少部分地通过基于所述估计的运动,变形所述第四参考帧的一个或多个像素来生成另外的变形帧;以及

基于所述另外的变形帧和另外的预测残差,生成表示所述另外输入帧的重构帧,表示所述另外输入帧的所述重构帧包括单向预测帧。

48. 一种其上存储有指令的非临时性计算机可读介质,当所述指令被一个或多个处理器执行时,使所述一个或多个处理器执行以下操作:

获得内插参考帧和表示输入帧与所述内插参考帧之间的运动的运动信息;

基于所述运动信息,估计所述输入帧和所述内插参考帧之间的所述运动;

至少部分地通过基于所述估计的运动,对所述内插参考帧的一个或多个像素进行变形来生成变形帧;以及

基于所述变形帧和预测残差来生成表示所述输入帧的重构帧,所述重构帧包括双向预测帧。

49. 根据权利要求48所述的非临时性计算机可读介质,还包括当被所述一个或多个处理器执行时,使得所述一个或多个处理器执行以下操作的指令:

获得第一参考帧和第二参考帧;以及

至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成所述内插参考帧。

50. 根据权利要求48所述的非临时性计算机可读介质,还包括当被所述一个或多个处理器执行时,使得所述一个或多个处理器执行以下操作的指令:

从比特流中获取残差;以及

基于所获取的残差来生成所述预测残差。

51. 根据权利要求50所述的非临时性计算机可读介质,其中,所述运动信息是从所述比特流中获得的。

52. 根据权利要求48所述的非临时性计算机可读介质,还包括当被所述一个或多个处理器执行时,使得所述一个或多个处理器执行以下操作的指令:

获得第四参考帧;

估计另外输入帧和所述第四参考帧之间的运动;

至少部分地通过基于所述估计的运动,变形所述第四参考帧的一个或多个像素来生成另外的变形帧;以及

基于所述另外的变形帧和另外的预测残差,生成表示所述另外输入帧的重构帧,表示所述另外输入帧的所述重构帧包括单向预测帧。

使用P帧译码系统的学习B帧译码

技术领域

[0001] 概括地说,本公开内容涉及视频译码(例如,对视频数据进行编码和/或解码)。例如,本公开内容的各方面涉及使用单向译码系统(例如,单向预测帧(P帧)编码器-解码器或CODEC)的学习双向预测帧(B帧)译码。

背景技术

[0002] 许多设备和系统允许处理和输出视频数据以供消费。数字视频数据包括大量数据,以满足视频质量、性能和功能方面日益增长的需求。例如,视频数据的消费者通常想要高保真度、分辨率、帧速率等等的高质量视频。满足这些需求通常需要大量的视频数据,这给处理和存储视频数据的通信网络和设备带来了巨大负担。视频译码技术可以用于对视频数据进行压缩。视频译码的一个示例目标是将视频数据压缩成使用较低比特率的形式,同时避免或最小化视频质量的降低。随着不断发展的视频服务变得可用,以及对大量视频数据的需求不断增加,需要具有更好性能和效率的译码技术。

发明内容

[0003] 本文描述的系统和技术包括使用单向(例如,P帧)译码系统的学习的双向(例如,B帧)译码,该译码系统可以利用通过在参考帧之间插值生成的参考帧。根据至少一个例子,提供了一种用于处理视频数据的方法。该方法可以包括:获得第一参考帧和第二参考帧;至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成第三参考帧;并基于所述第三参考帧对输入帧执行单向帧间预测。

[0004] 在另一个例子中,提供了一种用于处理视频数据的系统。该系统可以包括具有插值引擎的双向译码引擎和单向译码引擎。在一些情况下,该系统可以包括至少一个存储器和耦合到所述存储器的一个或多个处理器(例如,在电路中实现)。所述插值引擎被配置为:获得第一参考帧和第二参考帧;至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成第三参考帧。所述单向译码引擎被配置为基于所述第三参考帧对输入帧执行单向帧间预测。

[0005] 在另一个例子中,提供了一种用于处理视频数据的非临时性计算机可读介质。该非临时性计算机可读介质可以包括存储在其上的指令,当所述指令被一个或多个处理器执行时,使所述一个或多个处理器执行以下操作:获得第一参考帧和第二参考帧;至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成第三参考帧;并基于所述第三参考帧对输入帧执行单向帧间预测。

[0006] 根据另一个例子,提供了一种用于处理视频数据的装置。该装置可以包括:用于获得第一参考帧和第二参考帧的单元;用于至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成第三参考帧的单元;以及用于基于所述第三参考帧对输入帧执行单向帧间预测的单元。

[0007] 在一些方面,上面所描述的方法、装置和非临时性计算机可读介质可以包括:估计

所述输入帧和所述第三参考帧之间的运动;并至少部分地通过基于所述估计的运动,变形所述第三参考帧的一个或多个像素,来生成变形帧。

[0008] 在一些方面,上面所描述的方法、装置和非临时性计算机可读介质可以包括:至少部分地通过确定所述输入帧和所述变形帧之间的差异来确定残差;并使用所述残差生成预测残差。

[0009] 在一些方面,上面所描述的方法、装置和非临时性计算机可读介质可以包括:基于所述变形帧和所述预测残差,生成表示所述输入帧的重构帧,所述重构帧包括双向预测帧;并使所述重构帧存储在存储器中。

[0010] 在一些方面,所述运动是使用第一自动编码器估计的,并且所述预测残差是使用第二自动编码器生成的。

[0011] 在一些方面,上面所描述的方法、装置和非临时性计算机可读介质可以包括:获得第四参考帧;估计另外输入帧和所述第四参考帧之间的运动;至少部分地通过基于所述估计的运动来变形所述第四参考帧的一个或多个像素,生成另外的变形帧;并基于所述另外的变形帧和另外的预测残差,生成表示所述另外输入帧的重构帧,表示所述另外输入帧的所述重构帧包括单向预测帧。

[0012] 在一些方面,上面所描述的方法、装置和非临时性计算机可读介质可以包括:确定表示从所述第一参考帧到所述第三参考帧的像素运动的第一组运动信息;确定表示从所述第二参考帧到所述第三参考帧的像素运动的第二组运动信息;至少部分地通过使用所述第一组运动信息对所述第一参考帧执行变形函数来生成第一变形信息;至少部分地通过使用所述第二组运动信息对所述第二参考帧执行所述变形函数来生成第二变形信息;并基于所述第一变形信息和所述第二变形信息生成所述第三参考帧。

[0013] 在一些方面,基于所述第一参考帧和所述第二参考帧之间的像素运动来确定所述第一组运动信息和所述第二组运动信息。

[0014] 在一些方面,所述第一组运动信息包括第一光流图,并且所述第二组运动信息包括第二光流图。

[0015] 在一些方面,所述变形函数包括双线性插值函数。

[0016] 在一些方面,所述输入帧与所述第三参考帧之间的所述估计的运动包括光流数据。

[0017] 在一些方面,所述输入帧与所述第三参考帧之间的所述估计的运动包括动态卷积数据。

[0018] 在一些方面,所述输入帧与所述第三参考帧之间的所述估计的运动包括基于块的运动数据。

[0019] 根据至少一个例子,提供了一种用于处理视频数据的方法。该方法可以包括:获得内插参考帧和表示输入帧与所述内插参考帧之间运动的运动信息;基于所述运动信息,估计所述输入帧和所述内插参考帧之间的运动;至少部分地通过基于所述估计的运动,对所述内插参考帧的一个或多个像素进行变形来生成变形帧;并基于所述变形帧和预测残差来生成表示所述输入帧的重构帧,所述重构帧包括双向预测帧。

[0020] 在另一个例子中,提供了一种用于处理视频数据的系统。该系统可以包括至少一个存储器和耦合到所述存储器的一个或多个处理器(例如,在电路中实现)。所述至少一个

处理器被配置为：获得内插参考帧和表示输入帧与所述内插参考帧之间运动的运动信息；基于所述运动信息，估计所述输入帧和所述内插参考帧之间的运动；至少部分地通过基于所述估计的运动，对所述内插参考帧的一个或多个像素进行变形来生成变形帧；并基于所述变形帧和预测残差来生成表示所述输入帧的重构帧，所述重构帧包括双向预测帧。

[0021] 在另一个例子中，提供了一种用于处理视频数据的非临时性计算机可读介质。该非临时性计算机可读介质可以包括存储的指令，当所述指令被一个或多个处理器执行时，使所述一个或多个处理器执行以下操作：获得内插参考帧和表示输入帧与所述内插参考帧之间运动的运动信息；基于所述运动信息，估计所述输入帧和所述内插参考帧之间的运动；至少部分地通过基于所述估计的运动，对所述内插参考帧的一个或多个像素进行变形来生成变形帧；并基于所述变形帧和预测残差来生成表示所述输入帧的重构帧，所述重构帧包括双向预测帧。

[0022] 根据另一个例子，提供了一种用于处理视频数据的装置。该装置可以包括：用于获得内插参考帧和表示输入帧与所述内插参考帧之间运动的运动信息的单元；用于基于所述运动信息，估计所述输入帧和所述内插参考帧之间的运动的单元；用于至少部分地通过基于所述估计的运动，对所述内插参考帧的一个或多个像素进行变形来生成变形帧的单元；以及用于基于所述变形帧和预测残差来生成表示所述输入帧的重构帧的单元，所述重构帧包括双向预测帧。

[0023] 在一些方面，上面所描述的方法、装置和非临时性计算机可读介质可以包括：获得第一参考帧和第二参考帧；并至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成所述内插参考帧。

[0024] 在一些方面，上面所描述的方法、装置和非临时性计算机可读介质可以包括：从比特流中获取残差；并基于所获得的残差来生成预测残差。

[0025] 在一些方面，所述运动信息是从所述比特流中获得的。

[0026] 在一些方面，上面所描述的方法、装置和非临时性计算机可读介质可以包括：获得第四参考帧；估计另外输入帧和所述第四参考帧之间的运动；至少部分地通过基于所估计的运动，变形所述第四参考帧的一个或多个像素来生成另外的变形帧；并基于所述另外的变形帧和另外的预测残差，生成表示所述另外输入帧的重构帧，表示所述另外输入帧的所述重构帧包括单向预测帧。

[0027] 在一些方面，装置可以包括移动设备、可穿戴设备、扩展现实设备（例如，虚拟现实（VR）设备、增强现实（AR）设备或混合现实（MR）设备）、个人计算机、膝上型计算机、服务器计算机、电视、视频游戏控制台或其它设备，也可以是这些设备的一部分。在一些方面，该装置包括移动设备（例如，移动电话或所谓的“智能电话”）。在一些方面，该装置还包括用于捕获一个或多个图像或视频帧的至少一个相机。例如，该装置可以包括用于捕获一个或多个图像和/或包括视频帧的一个或多个视频的一个相机（例如，RGB相机）或多个相机。在一些方面，该装置包括用于显示一个或多个图像、视频、通知或其它可显示数据的显示器。在一些方面，该装置包括发射器，其被配置为通过传输介质将重构的视频帧传输到至少一个设备。在一些方面，处理器包括神经处理单元（NPU）、中央处理单元（CPU）、图形处理单元（GPU）或其它处理设备或组件。

[0028] 本概括内容既不是旨在标识所要求保护主题的关键或必要特征，也不是旨在单独

用于确定所要求保护主题的范围。通过参考本专利的整个说明书的适当部分、任何或所有附图以及每个权利要求,应当理解本发明的主题。

[0029] 通过参考以下的说明书、权利要求书和附图,前述内容以及其它特征和实施例将变得更加显而易见。

附图说明

[0030] 下面结合附图来详细地描述本申请的说明性实施例:

[0031] 图1根据本公开内容的一些例子,示出了可以实现本文所描述的各种技术的示例图像处理系统;

[0032] 图2A示出了根据本公开内容的一些例子的全连接神经网络的示例;

[0033] 图2B示出了根据本公开内容的一些例子的局部连接的神经网络的示例;

[0034] 图2C示出了根据本公开内容的一些例子的卷积神经网络的示例;

[0035] 图2D示出了根据本公开内容的一些例子的、被设计为从图像中识别特征的深度卷积网络(DCN)的详细示例;

[0036] 图3是示出根据本公开内容的一些例子的另一种示例性DCN的框图;

[0037] 图4是根据本公开内容的一些例子,示出被配置为执行单向运动压缩和补偿以生成预测的单向预测帧(P帧)译码系统的例子的图;

[0038] 图5A是根据本公开内容的一些例子,示出被配置为执行双向运动压缩和补偿以生成预测的双向预测帧(B帧)译码系统的例子的图;

[0039] 图5B是根据本公开内容的一些例子,示出被配置为使用帧内插来生成预测的B帧译码系统的例子的图;

[0040] 图6是根据本公开内容的一些例子,示出被配置为使用P帧译码系统和帧插值引擎来执行学习的B帧译码的B帧译码系统的例子的图;

[0041] 图7A是根据本公开内容的一些例子,示出可以用于执行单向译码和/或双向译码的P帧译码系统的例子的图;

[0042] 图7B是根据本公开内容的一些例子,示出可以用于使用P帧译码系统执行双向译码的B帧译码系统的例子的图;

[0043] 图7C是根据本公开内容的一些例子,示出可以由B帧译码系统使用的帧插值引擎的例子的图;

[0044] 图8是示出根据本公开内容的一些例子的运动估计技术的例子的图;

[0045] 图9是示出根据本公开内容的一些例子的I帧、B帧、I帧(IBI)译码技术的例子的图;

[0046] 图10是示出根据本公开内容的一些例子的I帧、B帧、P帧(IBP)译码技术的例子的图;

[0047] 图11是根据本公开内容的一些例子,示出用于双向帧预测的示例顺序方案的图;

[0048] 图12是根据本公开内容的一些例子,示出用于双向帧预测的示例分层方案的图;

[0049] 图13A和图13B是根据本公开内容的一些例子,示出本文所描述的B帧译码系统的性能结果的图表;

[0050] 图14A-图14F是根据本公开内容的一些例子,示出本文所描述的B帧译码系统的进

一步性能结果的图表；

[0051] 图15是根据本公开内容的一些例子，示出用于处理视频数据的处理的示例流程图；

[0052] 图16是根据本公开内容的一些例子，示出用于处理视频数据的处理的另一个例子的流程图；以及

[0053] 图17示出了可以用于实现本文描述的各个方面的示例计算系统。

具体实施方式

[0054] 下面提供本公开内容的某些方面和实施例。对于本领域普通技术人员显而易见的是，这些方面和实施例中的一些可以独立地应用，并且它们中的一些可以组合地应用。在下面的描述中，为了便于解释起见，阐述了特定的细节以便提供对本申请的实施例的透彻理解。但是，显而易见的是，可以在不具有这些特定细节的情况下实践各个实施例。附图和描述并不旨在是限制性的。

[0055] 随后的描述仅提供了示例性实施例，并且无意于限制本公开内容的保护范围、适用性或配置。相反，对示例性实施例的随后描述，将向本领域普通技术人员提供用于实现示例性实施例的使能描述。应当理解的是，在不脱离所附权利要求书阐述的本申请的精神和保护范围的情况下，可以对元件的功能和布置进行各种改变。

[0056] 数字视频数据可以包括大量数据，特别是随着对高质量视频数据的需求持续增长。例如，视频数据的消费者通常希望视频质量越来越高，具有高保真度、分辨率、帧速率等。然而，满足这些需求通常需要大量的视频数据，这给处理和存储视频数据的通信网络和设备带来了巨大负担。

[0057] 可以使用各种技术来对视频数据进行译码。在一些情况下，可以根据特定的视频译码标准和/或方案来执行视频译码。示例视频译码标准包括高效视频编码(HEVC)、高级视频编码(AVC)、运动图像专家组(MPEG)编码、通用视频编码(VVC)等。视频译码技术的一个目标是将视频数据压缩成使用较低比特率的形式，同时避免或最小化视频质量的下降。随着对视频服务需求的增长和新视频服务的出现，需要具有更高效率和性能的译码技术。

[0058] 视频译码可以使用诸如帧内预测或帧间预测之类的预测方法，它们利用存在于视频帧或其它图像或帧序列中的冗余。帧内预测是使用单个视频帧内的数据执行的，并且基于帧的空间特征。使用帧内预测编码的帧称为I帧。帧的帧间预测是基于该帧相对于其它帧的时间特性来执行的。例如，可以通过识别包括相对于视频帧的变化的其它视频帧的区域和包括相对于视频帧的冗余的区域(例如，基本保持不变的背景区域)，来执行视频帧的帧间预测。可以去除冗余，从而导致视频帧的残差。可以进一步对残差进行编码(例如，使用熵译码)，并且可以将结果包括在存储、传输或以其它方式输出的比特流中。

[0059] 帧间预测的示例包括单向预测(uni-prediction)和双向预测(bi-prediction)。单向预测包括在执行帧间预测时使用单个参考帧。使用单向预测译码的帧称为P帧。双向预测涉及在执行帧的帧间预测时使用两个参考帧。使用双向预测编码的帧称为B帧。

[0060] 在一些情况下，可以使用机器学习系统来执行视频编码(压缩)和解码(解压缩)。一般来说，机器学习(ML)是人工智能(AI)的一个子集。ML系统包括计算机系统可以用来通过依赖模式和推理来执行各种任务的算法和统计模型。ML系统的一个例子是神经网络(也

称为人工神经网络),它可以包括一组互连的人工神经元(例如,神经元模型)。神经网络可以用于各种应用和/或设备,例如图像分析和/或计算机视觉应用、互联网协议(IP)相机、物联网(IoT)设备、自主车辆、服务机器人等等。

[0061] 神经网络中的各个节点可以通过获取输入数据并对数据执行操作来模拟生物神经元。对输入数据执行的操作的结果,有选择地传递给其它神经元。权重值与网络中的每个向量和节点相关联,并且这些值限制了输入数据与输出数据的关联方式。例如,可以将每个节点的输入数据乘以相应的权重值,然后将乘积相加。可以通过可选的偏差来调整乘积之和,并且可以将激活函数应用于该结果,产生节点的输出信号或“输出激活”(有时称为激活图或特征图)。最初可以由通过网络来训练数据的迭代流,来确定这些权重值。例如,可以在网络学习如何通过其典型输入数据特征识别特定类别的训练阶段,建立权重值。在一个例子中,可以通过调整与神经元相关联的参数值(例如,激活参数和/或权重、偏差等)、添加和/或移除神经元或甚至神经元的层、添加或删除神经元之间的边缘等,来训练网络以学习特定任务。

[0062] 存在不同类型的神经网络,例如自动编码器、卷积神经网络(CNN)、循环神经网络(RNN)、多层感知器(MLP)神经网络等。卷积神经网络可以包括人工神经元的集合,每个人工神经元都有一个感受野(例如,输入空间的空间局部区域),并且共同平铺输入空间。卷积神经网络有许多应用。例如,CNN可以广泛地用于模式识别和分类领域。RNN的工作原理是保存层的输出,并将该输出反馈给输入,以帮助预测层的结果。在MLP神经网络中,可以将数据馈送到输入层,并且一个或多个隐藏层为数据提供抽象级别。然后可以基于抽象数据对输出层进行预测。MLP可能特别适用于为输入分配类别或标签的分类预测问题。

[0063] 在分层神经网络架构中(当存在多个隐藏层时,称为深度神经网络),第一层人工神经元的输出成为第二层人工神经元的输入,第二层人工神经元的输出成为第三层人工神经元的输入,依此类推。可以训练卷积神经网络以识别特征层次结构。卷积神经网络架构中的计算可以分布在一组处理节点上,可以将这些节点配置在一个或多个计算链中。这些多层架构可以一次训练一层,并且可以使用反向传播进行微调。

[0064] 在一些例子中,基于机器学习的P帧和B帧译码系统可以用于分别执行单向预测和双向预测。在一些情况下,此类系统可以包括神经网络架构(例如,一个或多个深度神经网络,例如一个或多个自动编码器)。基于机器学习的P帧译码系统的示例可以对当前帧和参考帧执行运动压缩和运动补偿,以确定当前帧和参考帧之间的运动预测。该运动预测可以用于修改参考帧的像素(例如,通过根据包括在运动预测中的运动向量,移动参考帧的像素),从而产生对当前帧的预测。P帧译码系统的残差部分可以生成表示预测帧和当前帧之间的差异的预测残差。预测的残差可以与预测相结合,以生成重构的当前帧。

[0065] 可以使用各种类型的基于机器学习的B帧译码系统来执行双向预测。基于机器学习的B帧译码系统的一个示例类似于上面所描述的P帧系统,但使用两个参考帧而不是单个参考帧来生成重构的当前帧。基于机器学习的B帧译码系统的另一个例子包括帧插值引擎,其基于两个参考帧之间的插值来生成预测。然后,残差部分确定与来自插值引擎的预测相结合的预测残差。

[0066] 本文描述了用于使用单向(例如,P帧)译码系统来执行学习的双向(例如,B帧)译码的系统、方法(也称为过程)、装置和计算机可读介质(统称为“系统和技术”)。如本文所使

用的,术语译码可以指代编码(例如,压缩)、解码(例如,解压缩)、或编码和解码两者。例如,提供了一种双向译码系统(也称为B帧译码系统),其包括用于基于至少两个参考帧来生成内插参考帧的帧插值引擎。在一些例子中,帧插值引擎可以使用与用于生成内插参考帧的参考帧相关联的运动信息(例如,光流或其它基于运动的技术)来生成内插参考帧。

[0067] 双向译码系统可以使用单向译码系统(也称为P帧译码系统)来生成双向预测帧(B帧)。例如,单向译码系统可以使用内插参考帧来生成B帧。B帧可以包括通过参考来自先前帧和未来帧(例如,在显示顺序和/或解码顺序中在B帧之前和之后的帧)的数据而生成的一种帧。除了用于B帧译码外,单向译码系统还可以用于在对特定帧执行单向预测模式时,生成单向预测帧(P帧)。例如,单向译码系统可以使用实际参考帧来生成P帧。P帧可以包括通过参考来自单个参考帧的数据而生成的一种帧,该参考帧可以包括先前帧(例如,在译码顺序中在B帧之前的帧)或未来帧(例如,在译码顺序上在B帧之后的帧)。如本文所使用的,术语帧、图片和图像可以互换地使用。

[0068] 在一些例子中,单向译码系统可以包括视频编码器和视频解码器,它们可以使用机器学习来处理视频数据并生成对输入帧的预测。在一些方面,视频编码器可以包括运动预测系统(例如,运动自动编码器)的编码器神经网络,并且视频解码器可以包括运动预测系统的解码器神经网络。例如,视频编码器(例如,运动预测系统的编码器神经网络)可以使用机器学习来生成表示输入帧和参考帧(例如,在执行双向预测时的内插参考帧或实际参考帧,例如在执行单向预测时不是基于插值生成的重构的P或B帧或I帧)之间的预测运动的潜在数据(例如,比特流)。视频解码器(例如,运动预测系统的解码器神经网络)可以使用机器学习来处理潜在数据,并预测或重建输入帧和参考帧之间的运动。单向译码系统可以使用预测的运动,通过变形参考帧的相应部分来生成输入帧的预测。例如,预测的运动可以包括运动向量(例如,包括输入帧的每个像素或像素块的运动向量的光流图)。变形引擎可以根据运动向量,移动参考帧的像素以生成对输入帧的预测(称为变形帧)。

[0069] 在一些例子中,单向译码系统的视频编码器和视频解码器也可以使用机器学习来生成输入帧的残差。在一些方面,单向译码系统可以包括残差预测系统(例如,残差自动编码器),其包括编码器神经网络和解码器神经网络,这些网络可以为输入帧生成残差。例如,视频编码器(例如,残差预测系统的编码器神经网络)可以接收残差作为输入,该残差包括输入帧和参考帧(例如,当执行双向预测时的内插参考帧,或执行单向预测时的实际参考帧)之间的差异。使用残差作为输入,视频编码器可以使用机器学习来生成表示残差的潜在数据(例如,比特流)。视频解码器(例如,残差预测系统的解码器神经网络)可以使用机器学习来处理潜在数据,并预测或重构残差。单向译码系统可以通过将预测的残差添加到为输入帧生成的预测(例如,由变形引擎生成),来生成表示输入帧的重构帧。

[0070] 在一些例子中,系统可以将表示运动的潜在数据和表示残差的潜在数据(例如,在编码的视频比特流中),传输到另一个设备以进行解码。例如,另一个设备可以包括视频解码器(例如,如上所述的运动预测系统和残差预测系统),该视频解码器可以使用机器学习,基于表示运动的潜在数据和表示残差的潜在数据来重构输入帧,类似于上面描述的(例如,通过处理运动潜在数据来预测输入帧和参考帧之间的运动,通过根据预测的运动来变形参考帧来生成输入帧的预测,处理残差潜在数据来预测残差,并通过将预测的残差与为输入帧生成的预测相加来生成重构帧)。

[0071] 将关于附图描述系统和技术的其它方面。图1示出了图像处理系统100的示例实现,在一些情况下,该图像处理系统100可以用于实现本文所描述的系统和技术。图像处理系统100可以包括中央处理单元(CPU) 102或多核CPU,其被配置为执行本文描述的功能中的一个或多个。参数或变量(例如,神经信号和突触权重)、与计算设备相关联的系统参数(例如,具有权重的神经网络)、延迟、频点信息、任务信息、图像数据等信息可以存储在神经处理单元(NPU) 108相关联的存储器块中,存储在与CPU 102相关联的存储器块中,存储在与图形处理单元(GPU) 104相关联的存储器块中,存储在与数字信号处理器(DSP) 106相关联的存储器块中,存储在存储器块118中,和/或可以分布在多个块中。在CPU 102处执行的指令可以从与CPU 102相关联的程序存储器和/或从存储器块118加载。

[0072] 图像处理系统100还可以包括用于执行特定功能的其它处理块,例如GPU 104;DSP 106;连接块110(其可以包括第五代(5G)连接、第四代长期演进(4G LTE)连接、Wi-Fi连接、USB连接、蓝牙连接等);和/或可以例如检测图像特征的多媒体处理器112。在一些例子中,可以在CPU 102、DSP 106和/或GPU 104中实现NPU 108。在一些情况下,图像处理系统100还可以包括一个或多个传感器114、一个或多个图像信号处理器(ISP) 116,和/或存储设备120。

[0073] 在一些例子中,图像处理系统100可以为一个或多个处理器实现ARM指令集架构。在本公开内容的一个方面,加载到CPU 102中的指令可以包括:用于在查找表(LUT)中搜索存储的与输入值和过滤器权重的乘积相对应的乘法结果的代码。加载到CPU 102中的指令还可以包括:用于当检测到乘积的查找表命中时,在乘积的乘法运算期间禁用乘法器的代码。此外,加载到CPU 102中的指令可以包括:用于在检测到乘积的查找表未命中时,存储输入值和过滤器权重的计算乘积的代码。

[0074] 图像处理系统100可以是计算设备或多个计算设备的一部分。在一些例子中,图像处理系统100可以是电子设备(或多个电子设备)的一部分,例如相机系统(例如,数码相机、IP相机、摄像机、安全相机等)、电话系统(如,智能手机、蜂窝电话、会议系统等)、桌面型计算机、XR设备(例如,头戴式显示器等)、智能可穿戴设备(例如,智能手表、智能眼镜等)、膝上型计算机或笔记本计算机、平板计算机、机顶盒、电视、显示设备、片上系统(SoC)、数字媒体播放器、游戏机、视频流设备、服务器、无人机、计算机汽车、物联网(IoT)设备、或者任何其它适当的电子设备。

[0075] 在一些实施方式中,CPU 102、GPU 104、DSP 106、NPU 108、连接块110、多媒体处理器112、一个或多个传感器114、ISP 116、存储器块118和/或存储设备120可以是同一计算设备的一部分。例如,在一些情况下,CPU 102、GPU 104、DSP 106、NPU 108、连接块110、多媒体处理器112、一个或多个传感器114、ISP 116、存储器块118和/或存储设备120可以集成到智能手机、膝上型计算机、平板计算机、智能可穿戴设备、视频游戏系统、服务器和/或任何其它计算设备中。在其它实施方式中,CPU 102、GPU 104、DSP 106、NPU 108、连接块110、多媒体处理器112、一个或多个传感器114、ISP 116、存储器块118和/或存储设备120可以是两个或更多个单独的计算设备的一部分。

[0076] 图像处理系统100和/或其组件可以被配置为使用根据本文所讨论的本公开内容的各方面的技术,来执行视频压缩和/或解压缩(也称为视频编码和/或解码,统称为视频编码)。通过使用深度学习架构和本文描述的技术来执行视频压缩和/或解压缩,本公开内容

的各方面可以提高设备上的视频压缩和/或解压缩的效率和/或减少相关联的资源需求和/或使用。例如,使用本文描述的视频译码技术的设备可以更高效地压缩视频数据,可以减少以压缩视频数据传输到目的地设备的数据量,并且目的地设备可以高效地接收和解压缩经压缩的视频数据。在一些例子中,本文描述的深度学习架构和技术可以减少在译码设备或组件(例如,编码器和解码器)之间交换的数据量,以对视频内容进行译码。为视频译码传输的减少的数据量可以减少延迟,提高性能,并减少计算资源(例如,带宽、内存、存储、功率、计算、硬件等)的成本或负担。

[0077] 如上所述,神经网络是机器学习系统的示例,并且可以包括输入层、一个或多个隐藏层以及输出层。从输入层的输入节点提供数据,由一个或多个隐藏层的隐藏节点执行处理,并通过输出层的输出节点来产生输出。深度学习网络通常包括多个隐藏层。神经网络的每一层都可以包括特征图或激活图,其中这些图可以包括人工神经元(或节点)。特征图可以包括过滤器、内核等。节点可以包括用于指示一个或多个层的节点的重要性的一个或多个权重。在一些情况下,深度学习网络可以具有一系列许多隐藏层,早期层用于确定输入的简单和低级特征,而后面的层则建立更复杂和抽象特征的层次结构。

[0078] 深度学习架构可以学习特征的层次结构。例如,如果呈现视觉数据,第一层可以进行学习,以识别输入流中的特征(例如,边缘)。在另一个例子中,如果呈现听觉数据,则第一层可以进行学习,以识别特定频率中的频谱功率。第二层以第一层的输出作为输入,可以学习识别特征,例如视觉数据的形状或听觉数据的声音组合。例如,更高层可以进行学习,以表示视觉数据中的复杂形状或听觉数据中的单词。更高的层可以进行学习,以识别常见的视觉对象和/或口语短语。

[0079] 神经网络可以被设计成具有多种连接模式。在前馈网络中,将信息从较低层传递到较高层,给定层中的每个神经元都与较高层中的神经元进行通信。如上所述,可以在前馈网络的连续层中建立分层表示。神经网络也可以具有循环或反馈(也称为自上而下)连接。在循环连接中,给定层中神经元的输出可以传递到同一层中的另一个神经元。循环架构可以有助于识别跨越多个输入数据块的模式,这些输入数据块按顺序传递给神经网络。从给定层中的神经元到较低层中的神经元的连接称为反馈(或自顶向下)连接。当对高级概念的识别可以有助于区分输入的特定低级特征时,具有许多反馈连接的网络会有所帮助。

[0080] 神经网络的层之间的连接可以是全连接的或局部连接的。图2A示出了全连接神经网络202的示例。在全连接神经网络202中,第一隐藏层中的神经元可以将其输出传送到第二隐藏层中的每个神经元,使得第二层中的每个神经元将接收来自第一层中每个神经元的输入。图2B示出了局部连接的神经网络204的示例。在局部连接的神经网络204中,第一隐藏层中的神经元可以连接到第二隐藏层中的有限数量的神经元。更一般地,局部连接神经网络204的局部连接层可以被配置为使得层中的每个神经元将具有相同或相似的连接模式,但连接强度可以具有不同的值(例如,210、212、214和216)。局部连接的模式可以在较高层中产生空间上不同的感受野,这是因为给定区域中的较高层神经元可以接收通过训练调整到网络总输入的受限部分的属性的输入。

[0081] 局部连接的神经网络的一个例子是卷积神经网络。图2C示出了卷积神经网络206的示例。卷积神经网络206可以被配置为使得与第二层中的每个神经元的输入相关联的连接强度是共享的(例如,208)。卷积神经网络可能非常适合输入的空间位置有意义的问题。

根据本公开内容的各方面,卷积神经网络206可以用于执行视频压缩和/或解压缩的一个或多个方面。

[0082] 一种类型的卷积神经网络是深度卷积网络(DCN)。图2D示出了DCN 200的示例,该DCN 200被设计为识别来自图像捕获设备230(例如,相机或图像传感器)输入的图像226的特征。在一些例子中,可以对当前示例的DCN 200进行训练,以识别图像226中的视觉特征(例如,图像226中的一个或多个对象或标志)。

[0083] 在一些例子中,可以用监督学习来训练DCN 200。在训练期间,可以向DCN 200呈现图像(例如,图像226),然后可以计算前向传递以产生输出222。DCN 200可以包括特征提取部分和分类部分。在接收到图像226时,卷积层232可以将卷积核(没有示出)应用于图像226,以生成第一组特征图218。举一个例子,卷积层232的卷积核可以是5x5核,其生成28x28的特征图。在本例子中,因为在第一组特征图218中生成了四个不同的特征图,所以在卷积层232将四个不同的卷积核应用于图像226。卷积核也可以称为滤波器或卷积滤波器。

[0084] 第一组特征图218可以由最大池化层(没有示出)进行二次采样以生成第二组特征图220。最大池化层减小了第一组特征图218的大小。也就是说,第二组特征图220的大小(例如,14x14)小于第一组特征图218的大小(例如,28x28)。减小后的大小为后续层提供了类似的信息,同时减少了存储器消耗。可以通过一个或多个后续卷积层(没有示出)进一步卷积第二组特征图220,以生成一组或多个后续特征图集合(没有示出)。

[0085] 在图2D的例子中,对第二组特征图220进行卷积以生成第一特征向量224。此外,对第一特征向量224进一步卷积以生成第二特征向量228。第二特征向量228的每个特征可以包括与图像226的可能特征相对应的标记,例如“符号”、“60”和“100”。softmax函数(没有示出)可以将第二特征向量228中的标记转换为概率。这样,DCN 200的输出222是图像226包括一个或多个特征的概率。

[0086] 在本例子中,输出222中“sign”和“60”的概率高于输出222中其它标记(例如,“30”、“40”、“50”、“70”、“80”、“90”和“100”)的概率。在训练之前,DCN 200产生的输出222可能是不正确的。因此,可以计算输出222和目标输出之间的误差。目标输出是图像226的基本事实(例如,“sign”和“60”)。然后,可以调整DCN 200的权重,使得DCN 200的输出222更接近于目标输出。

[0087] 为了调整权重,学习算法可以计算权重的梯度向量。梯度可以指示在调整权重时误差将增加或减少的量。在顶层,梯度可以直接对应于连接倒数第二层中激活的神经元和输出层中的神经元的权重值。在较低层中,梯度可以取决于权重的值和较高层的计算误差梯度。然后,可以调整权重以减少误差。这种调整权重的方式可以称为“反向传播”,因为它涉及通过神经网络的“反向传递”。

[0088] 在实践中,可以在少量示例上计算权重的误差梯度,使得计算的梯度接近真实的误差梯度。这种近似方法可以称为随机梯度下降。可以重复随机梯度下降,直到整个系统的可实现错误率停止下降或直到错误率达到目标水平为止。在学习之后,可以向DCN呈现新图像,并且通过网络的前向传递可以产生可以被认为是DCN的推断或预测的输出222。

[0089] 深度置信网络(DBN)是包含多层隐藏节点的概率模型。DBN可以用于提取训练数据集的分层表示。可以通过堆叠受限玻尔兹曼机(RBM)层来获得DBN。RBM是一种人工神经网络,其可以学习一组输入的概率分布。由于RBM可以在没有关于每个输入应归入的类别信息

的情况下学习概率分布,因此RBM通常用于无监督学习。使用混合的无监督和监督范式,可以以无监督方式训练DBN的底部RBM并用作特征提取器,而可以以监督方式训练顶部RBM(在前一层输入和目标类别的联合分布上),并且其可以用作分类器。

[0090] 深度卷积网络(DCN)是卷积网络的网络,其配置有另外的池化层和归一化层。DCN可以在许多任务上实现高性能。可以使用监督学习来训练DCN,其中输入和输出目标对于许多示例都是已知的,并用于通过使用梯度下降方法来修改网络的权重。

[0091] DCN可以是前馈网络。此外,如上所述,在第一层中的神经元之间共享从DCN的第一层中的神经元到下一个更高层中的一组神经元的连接。DCN的前馈和共享连接可以用于快速处理。例如,DCN的计算负担可能远小于包含循环或反馈连接的类似大小的神经网络的计算负担。

[0092] 可以认为卷积网络的每一层的处理是空间不变的模板或基础投影。如果首先将输入分解到多个通道,例如彩色图像的红色、绿色和蓝色通道,那么可以认为在该输入上训练的卷积网络是三维的,两个空间维度沿着图像的轴,并且第三维捕获颜色信息。可以将卷积连接的输出视为在后续层中形成特征图,其中该特征图的每个元素(例如,220)接收来自前一层中的一系列神经元(例如,特征图218)和来自多个通道中的每一个通道的输入。可以用非线性进一步处理(例如,校正, $\max(0, x)$) 特征图中的值。来自相邻神经元的值可以进一步合并,这对应于下采样,并且可以提供另外的局部不变性和降维。

[0093] 图3是示出深度卷积网络350的例子的框图。深度卷积网络350可以包括基于连通性和权重共享的多个不同类型的层。如图3中所示,深度卷积网络350包括卷积块354A、354B。卷积块354A、354B中的每一个可以配置有卷积层(CONV) 356、归一化层(LNorm) 358和最大池化层(MAX POOL) 360。

[0094] 卷积层356可以包括能够应用于输入数据352以生成特征图的一个或多个卷积滤波器。尽管仅示出了两个卷积块354A、354B,但本公开内容不限于此,而是可以根据设计偏好,在深度卷积网络350中包括任意数量的卷积块(例如,块354A、354B)。归一化层358可以对卷积滤波器的输出进行归一化。例如,归一化层358可以提供白化或横向抑制。最大池化层360可以在空间上提供下采样聚合,以用于局部不变性和降维。

[0095] 例如,深度卷积网络的并行滤波器组可以加载到图像处理系统100的CPU 102或GPU 104上以实现高性能和低功耗。在一些例子中,并行滤波器组可以加载在图像处理系统100的DSP 106或ISP 116上。深度卷积网络350可以访问可能存在于图像处理系统100上的其它处理块。

[0096] 深度卷积网络350可以包括一个或多个全连接层,例如层362A(标记为“FC1”)和层362B(标记为“FC2”)。深度卷积网络350可以包括逻辑回归(LR)层364。在深度卷积网络350的每一层356、358、360、362、364之间是要更新的权重(没有示出)。每一层(例如,356、358、360、362、364)的输出可以用作深度卷积网络350中的后续层(例如,356、358、360、362、364)的输入,以从在第一个卷积块354A处提供的输入数据352(例如,图像、音频、视频、传感器数据和/或其它输入数据)学习分层特征表示。深度卷积网络350的输出是针对输入数据352的分类得分366。分类得分366可以是一组概率,其中每个概率是输入数据包括来自一组特征中的特征的概率。

[0097] 另一种类型的神经网络是自动编码器。可以对自动编码器进行训练(例如,使用训

练数据和一个或多个损失函数),以接收输入并在其输出处生成该输入的版本(例如,实质上将其输入复制到其输出)。可以对自动编码器进行训练,以无监督的方式学习高效的数据编码。例如,给定一个对象的图像,自动编码器可以首先将图像编码为较低维的潜在表示,然后将潜在表示解码回该对象的图像。自动编码器可以学习(通过训练)压缩输入数据,同时最小化重构误差。

[0098] 如上所述,数字视频数据可以包括大量的数据,这会给处理和存储视频数据的通信网络和设备带来很大的负担。例如,录制未压缩的视频内容通常会导致文件大小随着录制视频内容的分辨率增加而大大增加。在一个说明性示例中,以1080p/24录制的未压缩的每通道16位视频(例如,分辨率为1920像素宽和1080像素高,每秒捕获24帧)可能每帧占用12.4兆字节,或每秒占用297.6兆字节。以每秒24帧的4K分辨率录制的未压缩的每通道16位视频,可能每帧占用49.8兆字节,或每秒占用1195.2兆字节。

[0099] 网络带宽是大视频文件可能成为问题的另一个约束。例如,视频内容通常通过无线网络(例如,通过LTE、高级LTE、新无线电(NR)、Wi-Fi™、Bluetooth™或其它无线网络)传送,并且可以构成消费者互联网流量的很大一部分。因此,希望减少用于在这些网络中传送视频内容的带宽量。

[0100] 因为未压缩的视频内容可能导致相当大的存储器可能涉及用于物理存储和相当大的带宽用于传输的大型文件,所以可以使用视频译码技术来压缩和解压缩这样的视频内容,如本文进一步描述的。为了减小视频内容的大小——以及因此存储视频内容所涉及的存储量和传送视频内容所涉及的带宽量——可以根据特定的视频译码标准和/或方案(例如,HEVC、AVC、MPEG、VVC等)来执行各种视频译码技术。视频译码可以使用诸如帧间预测或帧内预测之类的预测方法,这些方法利用了视频图像或序列中存在的冗余。视频译码技术的一个目标是将视频数据压缩成使用较低比特率的形式,同时避免或最小化视频质量的下降。随着对视频服务需求的增长和新视频服务的出现,需要具有更高译码效率、性能和码率控制的译码技术。

[0101] 编码设备可以根据视频译码标准对视频数据进行编码以生成经编码的视频比特流。在一些例子中,经编码的视频比特流(或“视频比特流”或“比特流”)可以包括一系列一个或多个编码视频序列。编码设备可以通过将每个图片分割成多个切片来生成图片的编码表示。切片独立于其它切片,因此在不依赖于来自同一图片内的其它切片的数据的情况下,对切片中的信息进行译码。一个切片包括一个或多个切片段,这些切片段包括独立切片段、以及一个或多个依赖于先前切片段的从属切片段(如果存在的话)。在HEVC中,将切片划分为亮度样本和色度样本的译码树块(CTB)。亮度样本的CTB和色度样本的一个或多个CTB以及样本的语法称为译码树单元(CTU)。CTU也可以称为“树块”或“最大译码单元”(LCU)。CTU是HEVC编码的基本处理单元。一个CTU可以拆分为多个不同大小的译码单元(CU)。CU包含称为译码块(CB)的亮度和色度样本阵列。

[0102] 可以进一步将亮度和色度CB拆分成预测块(PB)。PB是亮度分量或色度分量的样本块,它使用相同的运动参数进行帧间预测或块内复制(IBC)预测(当可用或启用时)。亮度PB和一个或多个色度PB与相关联的语法一起形成预测单元(PU)。对于帧间预测,在每个PU的比特流中用信号通知一组运动参数(例如,一个或多个运动向量、参考索引等),并且这些参数用于亮度PB和一个或多个色度PB的帧间预测。运动参数也可以称为运动信息。也可以将

CB划分为一个或多个变换块(TB)。TB表示颜色分量的样本的正方形块,在其上应用残差变换(例如,在某些情况下相同的二维变换)以对预测残差信号进行译码。变换单元(TU)表示亮度和色度样本的TB以及相应的语法元素。下面更详细地描述变换译码。

[0103] 根据HEVC标准,可以使用TU来执行变换。可以基于给定CU内的PU的大小来确定TU的大小。TU可以与PU大小相同或更小。在一些例子中,可以使用称为残差二叉树(RQT)的二叉树结构,将对应于CU的残差样本细分成更小的单元。RQT的叶子节点可以对应TU。可以将与TU相关联的像素差值进行变换以产生变换系数。然后,可以由编码设备对变换系数进行量化。

[0104] 一旦将视频数据的图片分割成CU,编码设备就使用预测模式来预测每个PU。然后,从原始视频数据中减去预测单元或预测块以获得残差(如下所述)。对于每个CU,可以使用语法数据,在比特流内用信号通知预测模式。预测模式可以包括帧内预测(或图片内预测)或帧间预测(或图片间预测)。帧内预测利用图片内空间相邻样本之间的相关性。例如,使用帧内预测,从同一图片中的相邻图像数据预测每个PU,例如使用DC预测来找到PU的平均值,使用平面预测以将平面表面拟合到PU,使用方向预测从相邻数据进行推断,或任何其它适当类型的预测。帧间预测使用图片之间的时间相关性来导出图像样本块的运动补偿预测。例如,使用帧间预测,使用运动补偿预测从一个或多个参考图片中的图像数据(按输出顺序在当前图片之前或之后)预测每个PU。例如,可以在CU级别,做出是使用图片间预测还是图片内预测来对图片区域进行译码的决定。

[0105] 在一些例子中,向图片的一个或多个切片分配切片类型。切片类型可以包括I切片、P切片和B切片。I切片(帧内,可独立解码)是仅通过帧内预测进行编码的图片切片,因此可独立解码,这是因为I切片仅需要帧内的数据来预测切片的任何预测单元或预测块。P切片(单向预测帧)是可以用帧内预测和用单向帧间预测进行译码的图片的切片。P切片内的每个预测单元或预测块都使用帧内预测或帧间预测进行译码。当应用帧间预测时,通过一个参考图片来预测预测单元或预测块,并且因此参考样本来自一帧的一个参考区域。B切片(双向预测帧)是可以使用帧内预测和帧间预测(例如,双向预测或单向预测)进行译码的图片切片。可以从两个参考图片,双向预测B切片的预测单元或预测块。每个图片可以贡献一个参考区域,并且可以对两个参考区域的样本集进行加权(例如,具有相等的权重或具有不同的权重)以产生双向预测块的预测信号。如上面所解释的,对一个图片的切片进行独立地译码。在一些情况下,可以将一个图片只译码为一个切片。

[0106] 在使用帧内和/或帧间预测执行预测之后,编码设备可以执行变换和量化。例如,在预测之后,编码设备可以计算对应于PU的残差值。残差值可以包括正在译码的当前像素块(PU)和用于预测当前块的预测块(例如,当前块的预测版本)之间的像素差值。例如,在生成预测块(例如,使用帧间预测或帧内预测)之后,编码设备可以通过从当前块中减去由预测单元产生的预测块来生成残差块。残差块包括量化当前块的像素值和预测块的像素值之间的差的一组像素差值。在一些例子中,可以以二维块格式(例如,二维矩阵或像素值阵列)来表示残差块。在这样的例子中,残差块是像素值的二维表示。

[0107] 在执行预测之后可能剩余的任何残差数据使用块变换进行变换,该块变换可以基于离散余弦变换、离散正弦变换、整数变换、小波变换、其它适当的变换函数或者其任何组合。在一些情况下,一个或多个块变换(例如,大小32x 32、16x 16、8x 8、4x 4或其它适当的

大小)可以应用于每个CU中的残差数据。在一些实施例中,TU可以用于由编码设备实现的变换和量化过程。具有一个或多个PU的给定CU还可以包括一个或多个TU。如下文进一步详细描述,可以使用块变换将残差值变换为变换系数,然后可以使用TU来量化和扫描,以产生用于熵译码的经序列化的变换系数。

[0108] 编码设备可以执行变换系数的量化。量化通过量化变换系数来提供进一步的压缩,以减少用于表示这些系数的数据量。例如,量化可以减少与一些或所有系数相关联的位深度。在一个例子中,可以在量化期间,将具有n位值的系数向下舍入为m位值,其中n大于m。

[0109] 一旦执行了量化,编码的视频比特流就包括量化的变换系数、预测信息(例如,预测模式、运动向量、块向量等)、分区信息和任何其它适当的数据(例如,其它语法数据)。然后,编码设备可以对编码的视频比特流的不同元素进行熵编码。在一些例子中,编码设备可以利用预定义的扫描顺序来扫描量化的变换系数,以产生可以进行熵编码的序列化向量。在一些例子中,编码设备可以执行自适应扫描。在扫描经量化的变换系数以形成向量(例如,一维向量)之后,编码设备可以对该向量进行熵编码。例如,编码设备可以使用上下文自适应可变长度译码、上下文自适应二进制算术译码、基于语法的上下文自适应二进制算术译码、概率区间分割熵译码或者另一种适当的熵编码技术。

[0110] 编码设备可以存储编码的视频比特流,和/或可以通过通信链路将编码的视频比特流数据发送到接收设备,接收设备可以包括解码设备。解码设备可以通过对构成编码的视频数据的一个或多个编码视频序列的元素进行熵解码(例如,使用熵解码器)和提取,来解码经编码的视频比特流数据。然后,解码设备可以对编码的视频比特流数据执行重新缩放和逆变换。然后,将残差数据传递到解码设备的预测阶段。解码设备然后使用帧内预测、帧间预测、IBC和/或其它类型的预测,来预测像素块(例如,PU)。在一些例子中,将预测添加到逆变换的输出(残差数据)。解码设备可以将解码的视频输出到视频目的地设备,该视频目的地设备可以包括用于将解码的视频数据显示给内容消费者的显示器或其它输出设备。

[0111] 由各种视频译码标准定义的视频译码系统和技术(例如,上面描述的HEVC视频译码技术)能够保留原始视频内容中的大部分信息,并且可以基于信号处理和信息论概念进行先验定义。然而,虽然各种压缩算法可能普遍适用(例如,适用于任何类型的视频内容),但这些视频译码技术可能缺乏在基于网络的环境或其它类型的环境中有益的某些特性。例如,视频译码系统应该能够实现低延迟操作。然而,根据现有视频译码标准操作的视频译码系统和相关技术,可以以几帧为单位(有时称为图片组或GoP)对视频进行编码和传输,并且可能需要等待一批视频帧(例如,GoP)的积累以进行传输。与使用逐帧译码和传输所能实现的延迟相比,这种基于批处理的视频译码导致更高的延迟。

[0112] 在一些情况下,可以使用基于机器学习的P帧译码系统(例如,其包括一个或多个神经网络)来执行单向预测。图4是示出基于机器学习的P帧译码系统400的例子的图。P帧译码系统400的组件包括运动压缩引擎402、运动补偿引擎404和残差压缩引擎406。P帧译码系统400的组件可以包括一个或多个神经网络,这些神经网络被配置为使用运动压缩引擎402、运动补偿引擎404和残差压缩引擎406来执行单向帧间预测(uni-prediction)。

[0113] 例如,运动压缩引擎402可以获得输入帧401和参考帧403。运动压缩引擎402可以执行基于机器学习的运动压缩,以确定输入帧401和参考帧之间的运动预测403。运动压缩引擎402可以将预测的运动信息提供给运动补偿引擎404。运动补偿引擎404还可以接收参

考帧403作为输入。运动补偿引擎404可以使用运动预测来修改参考帧的像素,并生成输入帧403的预测。例如,运动补偿引擎404可以根据包含在运动预测中的运动向量来移动参考帧的像素。残差压缩引擎406可以生成表示预测和输入帧401之间的差异的预测残差。然后,可以将预测残差与预测进行组合,以生成重构的输入帧。可以输出重构的输入帧以用于存储、显示等。

[0114] 在一些情况下,可以使用基于机器学习的B帧译码系统来执行双向预测。图5A是示出B帧译码系统500的例子的图。B帧译码系统500的组件包括运动压缩引擎502、运动补偿引擎504和残差压缩引擎506。B帧译码系统500可以包括一个或多个神经网络,这些神经网络被配置为使用运动压缩引擎502、运动补偿引擎504和残差压缩引擎506来执行双向帧间预测(双向预测)。双向预测是基于两个参考帧(在图5A中表示为参考帧0(Ref₀)和参考帧1(Ref₁))而不是单个参考帧来执行。

[0115] 例如,可以将输入帧501以及参考帧Ref₀和Ref₁作为输入提供给运动压缩引擎502。运动压缩引擎502可以执行基于机器学习的运动压缩,以确定输入帧501和第一参考帧Ref₀之间的运动的第一预测以及输入帧501和第二参考帧Ref₁之间的运动的第二预测。运动补偿引擎504可以接收参考帧Ref₀和Ref₁以及来自运动压缩引擎502的预测运动作为输入。运动补偿引擎504可以使用两个运动预测来修改参考帧Ref₀和Ref₁的像素并生成输入帧501的预测。例如,运动补偿引擎504可以根据第一运动预测来移动第一参考帧Ref₀的像素,并且可以根据第二运动预测来移动第二参考帧Ref₁的像素。运动补偿引擎504可以合并这两个预测以生成用于输入帧501的合并预测。

[0116] 残差压缩引擎506可以生成表示合并预测和输入帧501之间的差异的预测残差。然后,可以将预测残差与合并预测进行组合以生成重构的输入帧,可以输出该重构的输入帧以用于存储或显示。

[0117] 图5B是示出B帧译码系统510的另一示例的图。B帧译码系统510的组件包括帧插值引擎512和残差压缩引擎516。B帧译码系统510的组件可以包括一个或多个神经网络,这些神经网络被配置为使用帧插值引擎512和残差压缩引擎516执行双向预测。

[0118] 类似于图5A的B帧译码系统500,B帧译码系统510可以基于两个参考帧(表示为参考帧0(Ref₀)和参考帧1(Ref₁))执行双向预测。然而,插值引擎512可以通过在参考帧Ref₀和参考帧Ref₁之间执行插值来生成输入帧511的预测。然后将预测输出到残差压缩引擎516。残差压缩引擎516可以生成表示预测和输入帧511之间的差异的预测残差。B帧译码系统510可以将预测的残差与合并的预测进行组合以生成重构的输入帧。B帧译码系统510可以输出重构的输入帧以用于存储、显示等。

[0119] 图5A和图5B的上述B帧译码系统在某些情况下可能不是最理想的。例如,图5A的B帧译码系统500基于两个参考帧Ref₀和Ref₁以及输入帧511,执行双向运动压缩以生成两个运动预测。B帧译码系统500基于两个运动预测修改参考帧Ref₀和Ref₁,然后合并修改后的帧以生成针对输入帧511的预测。这样的过程涉及训练和推理期间的许多计算(在训练机器学习系统并用于处理输入视频数据之后)。此外,(图5A或图5B的)单独B帧译码系统和P帧译码系统(图4)需要单独的神经网络(具有不同的参数,例如权重、偏差等)。训练图5A和图5B的机器学习系统500和510还需要大量的训练数据(例如,每个预测需要两个参考帧)。此外,图5B的B-帧译码系统510确定了仅在线性运动假设下准确的插值,并且该预测可能没有与输

入帧对准。在一些情况下,B帧译码系统510可以使用非线性帧内插器;但是,在这种情况下仍可能发生错位。B帧译码系统510仅依靠残差来补偿未对准。

[0120] 如上所述,本文描述了用于使用P帧译码系统和内插参考帧来执行学习的B帧译码的系统和技术。本文描述的系统和技术可以使用从参考帧内插(例如,使用诸如光流信息的运动信息进行内插)的参考帧,来高效和智能地预测双向预测帧(B帧)。这些系统和技术可以实现具有较低资源需求(例如,较低的计算和带宽需求)的高效、高质量视频译码。例如,当神经网络用于P帧译码系统时,P帧译码系统需要一组神经网络参数(例如,权重、偏差等)来执行双向预测(B帧)并执行单向预测(P帧)。此外,由于仅使用单个参考帧(内插参考帧)来执行双向预测,因此每个B帧的双向预测仅需要单个运动信息集。在一些情况下,与训练图5B的B帧译码系统500所需的训练数据相比,训练机器学习系统(例如,神经网络)可能还需要更少的训练数据。

[0121] B帧译码系统600还提供改进的率失真性能,例如与图5A的B帧译码系统500和图5B的B帧译码系统510相比。例如,对于给定的输入帧,当来自过去(例如,Ref₀)和未来(例如,Ref₁)的参考帧都可用时,例如在线性运动假设下,B帧译码系统600可以确定通过对两个参考帧进行线性插值,粗略预测输入帧611。该预测消耗零位,这是因为两个参考帧已经可用并且不需要运动场或其它信息来执行插值。图5A的B帧译码系统500使用双向运动(例如,流动或变形),其不利用这种信息并且发送关于两个参考帧的运动信息(例如,光流)(例如,好像解码器/接收器不知道物体运动)。如上所述,关于图5B的B帧译码系统510,帧插值引擎512的插值结果仅在线性运动假设下是准确的。在这样的系统中,预测可能没有与输入帧对准。即使在采用非线性帧内插器时,仍可能发生未对准。在这些情况下,译码系统仅依靠残差来补偿未对准。结果,与图6的B帧译码系统600相比,这种译码系统的译码效率会严重降低,图6的B帧译码系统600可以使用廉价的边信息(例如,参考帧之间的插值)来纠正未对准,然后可以使用残差来确定重构的输入帧。

[0122] 图6是示出被配置为执行学习的B帧译码的B帧译码系统600的示例的图。B帧译码系统600包括用于处理视频数据的各种组件。如图所示,B帧译码系统600的组件包括帧插值引擎622和P帧译码系统624。使用帧插值引擎622和P帧译码系统624,B帧译码系统600可以执行输入帧611的双向帧间预测(bi-prediction)以生成输入帧611的重构版本(或解码版本),其称为重构的输入帧615。帧插值引擎622可以基于对参考帧0(Ref₀)和参考帧1(Ref₁)进行处理,来生成内插参考帧613。帧插值引擎622可以利用任何类型的插值技术。在一些例子中,帧插值引擎622可以使用与参考帧Ref₀和Ref₁相关联的运动信息来生成内插参考帧613。例如,如参考图7C所描述的,帧插值引擎622可以基于为参考帧Ref₀和参考帧Ref₁确定的双向运动信息(例如,前向光流和后向光流)来确定内插参考帧615。在一些情况下,可以存储内插参考帧615,例如用作一个或多个其它输入帧的参考帧。

[0123] B帧译码系统600可以使用P帧译码系统624来生成重构的输入帧615(作为B帧)。例如,如图6中所示,P帧译码系统624可以使用内插参考帧613作为输入来生成重构的输入帧615。P帧译码系统624可以使用任何类型的P帧译码技术。在一些例子中,P帧译码系统624可以包括至少一个基于机器学习的视频编码器和至少一个基于机器学习的视频解码器,以处理内插参考帧613和输入帧611来生成重构的输入帧615。在一些实施方式中,如下文关于图7A和图7B所描述的,P帧译码系统624可以包括具有编码器神经网络和解码器神经网络的运

动神经网络系统。该运动神经网络系统可以包括任何类型的神经网络结构,例如自动编码器、卷积神经网络(CNN)和/或其它神经网络结构。运动神经网络系统的编码器神经网络可以生成表示针对输入帧611的运动数据的潜在数据。运动神经网络系统的解码器神经网络可以从潜在数据重构运动,以生成针对输入帧611的预测运动。P帧译码系统624可以使用预测的运动,来变形内插参考帧613以生成输入帧611的预测。

[0124] P帧译码系统624还可以包括残差神经网络系统,该残差神经网络系统包括编码器神经网络和解码器神经网络。残差神经网络系统可以包括任何类型的神经网络结构,例如自动编码器、CNN和/或其它神经网络结构。P帧译码系统624可以例如通过确定变形的内插参考帧613(使用预测的运动变形)和输入帧611之间的差异,为输入帧611生成残差。残差神经网络系统的编码器神经网络可以对残差进行处理,以生成表示残差的残差潜在数据。残差神经网络系统的解码器神经网络可以从残差潜在数据重构残差,以生成针对输入帧611的预测残差。P帧译码系统624可以通过将预测残差添加到变形的内插参考帧613,来生成重构的输入帧615。

[0125] 在一些例子中,P帧译码系统624可以用于执行双向预测(以生成B帧),并且还可以用于执行单向预测(以生成P帧)。例如,P帧译码系统624可以用于在确定对输入帧执行单向预测译码模式时生成P帧,并且可以用于在确定对输入帧执行双向预测译码模式时生成B帧。参考图9-图12描述了关于为某些帧选择不同译码模式(例如,帧内预测、单向预测或双向预测)的细节。当为输入帧选择单向预测模式时,P帧译码系统624可以处理实际参考帧(例如,参考帧 Ref_0 或参考帧 Ref_1),以生成重构帧作为P-帧。下面参考图7A描述了关于使用P帧译码系统624进行单向预测的细节。当为输入帧选择双向预测模式时,P帧译码系统624可以处理内插参考帧(例如,内插参考帧613)以生成重构帧作为B帧。下面参考图7B描述了关于使用P帧译码系统624进行双向预测的细节。

[0126] 虽然将B帧译码系统600示出为包括某些组件,但本领域普通技术人员应当理解,B帧译码系统600可以包括比图6中所示的组件更多的组件。在一些情况下,B帧译码系统600的其它组件可以包括软件、硬件、或软件和硬件的一种或多种组合。例如,在一些情况下,B帧译码系统600可以包括一个或多个显示设备、一个或多个其它处理引擎、一个或多个图像传感器(例如,相机)、一个或多个其它传感器(例如,一个或多个惯性测量单元(IMU)、雷达、光探测和测距(LIDAR)传感器、音频传感器等)、和/或用于无线通信的一个或多个无线收发器(或单独的无线接收器和发射器),该无线通信例如为蜂窝网络通信(例如,使用长期演进(LTE)协议、5G/新无线电(NR)协议或其它蜂窝协议)、无线局域网(WLAN)通信(例如,使用802.11Wi-Fi协议)、蓝牙或其它短程通信、其任意组合和/或其它通信。本领域普通技术人员应当理解,B帧译码系统600可以包括图6中没有示出的一个或多个其它软件和/或硬件组件。

[0127] 在一些实施方式中,B帧译码系统600的其它组件可以包括电子电路或其它电子硬件和/或可以使用电子电路或其它电子硬件来实现,其可以包括一个或多个可编程电子电路(例如,数字信号处理器(DSP))、微处理器、微控制器、图形处理单元(GPU)、中央处理单元(CPU)、它们的任何组合和/或其它适当的电子电路),和/或可以包括计算机软件、固件或其任意组合和/或使用计算机软件、固件或其任意组合来执行本文所描述的各种操作。软件和/或固件可以包括存储在计算机可读存储介质上,并且可由实现B帧译码系统600的电子

设备的一个或多个处理器执行的一个或多个指令。

[0128] B帧译码系统600可以包括计算设备或者是计算设备的一部分。例如,该计算设备可以包括移动或固定电话手机(例如,智能手机、蜂窝电话等)、扩展现实设备(例如,虚拟现实(VR)设备、增强现实(AR)设备和/或混合现实设备)、图像捕获设备(例如,相机)、车辆、显示设备、服务器计算机(例如,与车辆计算系统通信)、台式计算机、膝上型计算机或笔记本电脑、平板计算机、机顶盒、电视机、数字媒体播放器、视频流设备或者任何其它适当的计算设备,或者是这些设备的一部分。

[0129] 图7A是说明可以用于执行P帧的单向译码和/或B帧的双向译码的P帧译码系统700的示例的图。P帧译码系统700是图6的P帧译码系统624的说明性示例。P帧译码系统700包括运动预测系统702、变形引擎704和残差预测系统706。运动预测系统702和残差预测系统706可以包括任何类型的机器学习系统(例如,使用一个或多个神经网络)。在一个说明性示例中,运动预测系统702包括运动自动编码器,并且残差预测系统706包括残差自动编码器。虽然将图7A的P帧译码系统700示出为包括某些组件,但本领域普通技术人员应当理解,P帧译码系统700可以包括比图7A中所示的组件更少或更多的组件。

[0130] 对于给定时间 t ,P帧译码系统700接收输入帧(表示为帧 X_t)和参考帧 $\hat{X}_{ref}$ 。参考帧 $\hat{X}_{ref}$ 是在时间 t 之前生成的先前重构的帧(如帽子符号“ $\hat{\cdot}$ ”所示)。例如,输入帧 X_t 是时间 t 的当前帧,并且参考帧 $\hat{X}_{ref}$ 可以是在时间 $t-n$ 重构或生成的过去帧(例如,在时间 t 在输入帧 X_t 之前的时间或顺序上),其中 n 是大于或等于1的整数值。在一个说明性示例中,参考帧 $\hat{X}_{ref}$ 可以是在时间上或顺序上紧接在输入帧 X_t 之前的帧,在这种情况下参考帧 $\hat{X}_{ref}$ 可以表示为 $\hat{X}_{t-1}$ 。在一些情况下,从译码系统的解码图像缓冲器(DPB)接收参考帧 $\hat{X}_{ref}$ 。

[0131] 在一些情况下,输入帧 X_t 可以是P帧,参考帧 $\hat{X}_{ref}$ 可以是I帧、P帧或B帧。例如,参考帧 $\hat{X}_{ref}$ 可以由I帧译码系统重构或生成(例如,I帧译码系统可以是包括P帧译码系统700的设备的一部分,或者是与包括P帧译码系统700的设备不同的设备)、由P帧译码系统700重构或生成(或者不同于包括P帧译码系统700的设备的设备的P帧译码系统)、或者由B帧译码系统来重构或生成(例如,图6的B帧译码系统600、图7B的B帧译码系统720、或者与包括P帧译码系统700的设备不同的设备的B帧译码系统)。在一些情况下,可以根据图9或图10中所示的一组图片(GoP)结构和/或根据顺序方案(例如,图11中所示的顺序方案)或分层方案(例如,图12中所示的分层方案)、或基于任何其它方案,来选择参考帧 $\hat{X}_{ref}$ 。

[0132] 运动预测系统702接收参考帧 $\hat{X}_{ref}$ 和输入帧 X_t 。运动预测系统702可以确定参考帧 $\hat{X}_{ref}$ 的像素和输入帧 X_t 的像素之间的运动(例如,由向量表示,诸如光流运动向量)。然后,运动预测系统702可以将运动进行编码,并且在一些情况下将运动解码为输入帧 X_t 的预测运动 $\hat{f}_t$ 。例如,运动预测系统702的编码器网络705可以确定当前帧 X_t 和参考帧 $\hat{X}_{ref}$ 之间的运动。编码器网络705可以将运动编码为潜在表示(表示为潜在 z_m)。在一些例子中,编码器网络705可以将运动映射到可以用作潜在 z_m 的潜在码。在一些情况下,编码器网络705可以通过对潜在码执行熵译码,将潜在 z_m 转换为比特流。在一些例子中,编码器网络705可以量化潜在 z_m (例如,在对潜在码执行熵译码之前)。量化的潜在 z_m 可以包括残差 z_m 的量化表示。在一

些情况下,潜在 z_m 可以包括表示量化码的神经网络数据(例如,神经网络节点的激活图或特征图)。编码器网络705可以存储潜在 z_m ,将潜在 z_m 发送到运动预测系统702的解码器网络703,和/或可以将潜在 z_m 发送到可以解码潜在 z_m 的另一个设备或系统。在接收到潜在 z_m 时,解码器网络703可以对潜在 z_m 进行解码(例如,逆熵译码、去量化和/或重构)以生成参考帧 $\hat{X}_{ref}$ 的像素和输入帧 X_t 的像素之间的预测运动 $\hat{f}_t$ 。编码器网络705和解码器网络703可以使用训练数据(例如,训练图像或帧)和一个或多个损失函数进行训练和优化,如下所述。

[0133] 在一些例子中,预测的运动 $\hat{f}_t$ 可以包括光流数据(例如,包括一个或多个位移向量的光流图)、动态卷积数据(例如,用于数据卷积的矩阵或核)或基于块的运动数据(例如,每个块的运动向量),如下面关于图8所描述的。例如,在一些情况下,预测的运动 $\hat{f}_t$ 可以包括光流图。在一个说明性示例中,光流图可以包括输入帧 X_t 的每个像素的运动向量(例如,第一像素的第一运动向量、第二像素的第二运动向量,等等)。运动向量表示当前帧 X_t 中的像素相对于参考帧 $\hat{X}_{ref}$ 中的相应像素的运动。

[0134] P帧译码系统的变形引擎704可以获得运动预测系统702输出的运动预测 $\hat{f}_t$ 。例如,变形引擎704可以从存储中检索运动预测 $\hat{f}_t$,或者可以从运动预测系统702接收运动预测 $\hat{f}_t$ 。变形引擎704可以使用运动预测 $\hat{f}_t$ 来变形(例如,通过执行运动补偿)参考帧 $\hat{X}_{ref}$ 的像素,导致生成变形帧 $\tilde{X}_t$ 。例如,参考帧 $\hat{X}_{ref}$ 的像素可以移动到新的位置或者根据重构的运动估计数据 $\hat{f}_t$ 进行修改。在使用光流来确定参考帧 $\hat{X}_{ref}$ 的像素和输入帧 X_t 的像素之间的运动的一个说明性示例中,变形引擎704可以将参考帧 $\hat{X}_{ref}$ 的像素移动光流图的运动向量所指示的量,产生变形的帧 $\tilde{X}_t$ 。如上所述,其它基于运动的技术也可以用于确定运动预测 $\hat{f}_t$,以变形参考帧 $\hat{X}_{ref}$ (例如,动态卷积数据、基于块的运动数据等)。

[0135] 如上所述,为了生成变形的帧 $\tilde{X}_t$,P帧译码系统700可以通过预测 X_t 和 $\hat{X}_{ref}$ 之间的运动 $\hat{f}_t$ (例如,光流或其它运动)并且使用运动预测 $\hat{f}_t$ 变形参考帧 $\hat{X}_{ref}$,来执行运动补偿。然而,基于运动预测 $\hat{f}_t$ 的帧预测 $\tilde{X}_t$ 可能不够准确,无法将输入帧 X_t 表示为重构的帧 $\hat{X}_t$ 。例如,输入帧 X_t 描绘的场景中可能存在一个或多个遮挡区域、过度照明、缺乏照明、或导致帧预测 $\tilde{X}_t$ 不够准确而无法用作重构帧 $\hat{X}_t$ 的其它影响。

[0136] 残差预测系统706可以用于校正 $\tilde{X}_t$ 的预测并生成表示帧 X_t 的重构帧 $\hat{X}_t$ 。如图7A中所示,可以通过从输入帧 X_t 中减去(使用减法运算708确定)预测帧 $\tilde{X}_t$ 来确定残差 r_t 。例如,一旦变形引擎704确定了预测帧 $\tilde{X}_t$,P帧译码系统700可以通过确定预测帧 $\tilde{X}_t$ 和输入帧 X_t 之间的差异(使用减法运算708),来确定残差 r_t 。残差预测系统706的编码器网络707可以将残差 r_t 编码成表示残差 r_t 的潜在 z_r 。在一些例子中,编码器网络707可以将残差 r_t 映射到可以用作潜在 z_r 的潜在码。在一些情况下,编码器网络707可以通过对潜在码执行熵译码,将潜在 z_r 转换为比特流。在一些例子中,编码器网络707还可以量化潜在 z_r (例如,在执行熵译码之前)。量化的潜在 z_r 可以包括残差 r_t 的量化表示。在一些情况下,潜在 z_r 可以包括表示量化码的神经网络数据(例如,神经网络节点的激活图或特征图)。编码器网络707可以存储潜在 z_r ,将潜在 z_r 发送到残差预测系统706的解码器网络709,和/或可以将潜在 z_r 发送到能够解码潜在 z_r 的另一个设备或系统。在接收到潜在 z_r 时,解码器网络709可以对潜在 z_r 进行解码(例如,逆

熵译码、去量化和/或重构)以生成预测残差 $\hat{r}_t$ 。编码器网络707和解码器网络709可以使用训练数据(例如,训练图像或帧)和一个或多个损失函数进行训练和优化,如下所述。

[0137] 预测的残差 $\hat{r}_t$ 可以与来自变形引擎704的帧预测 $\hat{X}_t$ 一起使用,以生成表示时间t处的输入帧 X_t 的重构帧 $\hat{X}_t$ 。例如,P帧译码系统700可以将预测残差 $\hat{r}_t$ 添加(使用加法运算710)到帧预测 $\hat{X}_t$ 以生成重构帧 $\hat{X}_t$ 。在一些情况下,残差预测系统706的解码器网络709可以将预测的残差 $\hat{r}_t$ 添加到帧预测 $\hat{X}_t$ 。重构帧 $\hat{X}_t$ 也可以称为解码帧。可以输出重构的当前帧 $\hat{X}_t$ 用于存储(例如,存储在解码的图片缓冲区(DPB)或其它存储设备中)、传输、显示、用于进一步处理(例如,作为进一步的帧间预测中的参考帧,用于后处理等)、和/或用于任何其它用途。

[0138] 如上所述,在一些情况下,P帧译码系统700可以将表示运动的潜在数据(潜在 z_m)和表示比特流中的残差(潜在 z_r)的潜在数据发送到另一个设备以进行解码。例如,该另一个设备可以包括能够从潜在 z_m 中解码预测运动 $\hat{f}_t$,从潜在 z_r 中解码预测残差 $\hat{r}_t$,并且可以基于预测运动 $\hat{f}_t$ 和预测残差 $\hat{r}_t$ 生成重构帧 $\hat{X}_t$ (例如,通过将预测残差 $\hat{r}_t$ 添加到使用预测运动 $\hat{f}_t$ 生成的帧预测 $\hat{X}_t$,如上所述)的视频解码器(例如,运动预测系统702和残差预测系统706,如上所述,或者被配置为解码潜在 z_m 和 z_r 的另一个解码器)。

[0139] 在一些例子中,可以使用训练数据和一个或多个损失函数来训练和优化运动预测系统702和/或残差预测系统706。在一些情况下,可以以端到端方式训练运动预测系统702和/或残差预测系统706(其中,在相同的训练过程中训练所有的神经网络组件)。在一些方面,训练数据可以包括训练图像或帧。在一些情况下,损失函数Loss可以用于基于运动预测系统702和/或残差预测系统706处理训练图像或帧的训练。在一个例子中,可以将Loss定义为 $Loss = D + \beta R$,其中D是帧(例如,输入帧 X_t)与其重构帧(例如, $D(\text{frame } X_t, \text{frame } \hat{X}_t)$)之间的失真, β 是控制比特率(每像素比特)的超参数,R是用于将残差(例如,残差 r_t)转换为压缩比特流(例如,潜在 z_r)的比特数。在一些例子中,可以基于峰值信噪比(PSNR)、结构相似性指数测量(SSIM)、多尺度SSIM(MS-SSIM)等来计算失真D。使用训练数据和损失函数,可以调整运动预测系统702和/或残差预测系统706的参数(例如,权重、偏差等),直到P帧译码系统700达到期望的译码结果为止。

[0140] 图7B是说明可以用于使用图7A的P-帧译码系统700执行双向译码的B-帧译码系统720的示例的图。B帧译码系统720还包括帧插值引擎722。虽然将B帧译码系统720示出为包括某些组件,但本领域普通技术人员应当理解,B帧译码系统720可以包括比图7B中所示的更少或更多的组件。

[0141] 插值引擎722可以获得两个参考帧,分别标记为 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 。如图所示,P帧译码系统700在时间t获得输入帧 X_t ,类似于上面关于图7A描述的那些。输入帧 X_t 可以是在时间上或顺序上在参考帧 $\hat{X}_{ref_0}$ 之后并在参考帧 $\hat{X}_{ref_1}$ 之前的帧。在一些例子中,输入帧 X_t 可以是时间t的当前帧,参考帧 $\hat{X}_{ref_0}$ 可以是在时间t-n的过去帧(例如,在时间上或顺序上在时间t的输入帧 X_t 之前),并且参考帧 $\hat{X}_{ref_1}$ 可以是在时间t-n的未来帧(例如,在时间上或顺序上在输入帧 X_t 之后),其中n和m可以表示相同或不同的值。在一个说明性示例中,参考帧 $\hat{X}_{ref_0}$ 可以是在时间t-1的紧接输入帧 X_t 之前的帧,并且参考帧 $\hat{X}_{ref_1}$ 可以是在时间t+1的紧接在输入帧 X_t 之后的帧。

[0142] 在一些情况下,输入帧 X_t 可以是B帧,并且参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 中的每一者可以是I帧、P帧、或B帧。例如,参考帧 $\hat{X}_{ref_0}$ 可以是I帧、P帧或B帧,参考帧 $\hat{X}_{ref_1}$ 可以是I帧、P帧或B帧。在一些情况下,参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 可以是相同类型的帧,例如B帧、P帧或I帧。在其它情况下,参考帧 $\hat{X}_{ref_0}$ 和参考帧 $\hat{X}_{ref_1}$ 可以是不同类型的帧。例如,参考帧 $\hat{X}_{ref_0}$ 可以是I帧,参考帧 $\hat{X}_{ref_1}$ 可以是P帧。在一些情况下,可以根据图9或图10中所示的一组图片(GoP)结构和/或根据顺序方案(例如,图11中所示的顺序方案)或分层方案(例如,图12中所示的分层方案)、或基于任何其它方案,来选择参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 。

[0143] 插值引擎722可以对两个参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 进行插值操作以生成内插参考帧 $\hat{X}_{ref_t}$ 。例如,在一些情况下,执行插值以生成可以用作输入帧 X_t 的参考的整个参考帧(例如,不仅是帧的部分,例如像素块)。可以对内插参考帧 $\hat{X}_{ref_t}$ 进行存储和/或输出到P-帧译码系统700。帧插值引擎722可以使用参考帧 $\hat{X}_{ref_0}$ 和参考帧 $\hat{X}_{ref_1}$,执行任何类型的插值以生成内插参考帧 $\hat{X}_{ref_t}$ 。可以由帧插值引擎722执行的插值技术的示例可以包括例如但不限于:基于光流的帧插值、移动梯度方法、基于相位的方法、基于内核的方法和/或任何其它帧插值技术。在一个说明性示例中,为了生成内插参考帧 $\hat{X}_{ref_t}$,帧插值引擎722可以对于参考帧 $\hat{X}_{ref_0}$ 和参考帧 $\hat{X}_{ref_1}$,计算运动估计信息(例如,光流信息,例如光流图,也称为运动向量图)。帧插值引擎722可以对运动估计信息(例如,流图)进行插值以确定与内插参考帧 $\hat{X}_{ref_t}$ 相关联的时间 t 的运动估计(例如,光流)。帧插值引擎725可使用内插运动估计信息(例如,内插光流)和参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 来产生内插参考帧 $\hat{X}_{ref_t}$ (例如,至少部分通过变形参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$)。

[0144] 在一些例子中,帧插值引擎722可以逐个像素地确定运动估计信息(例如,光流信息)。例如,帧插值引擎722可以将参考帧 $\hat{X}_{ref_0}$ (或参考帧 $\hat{X}_{ref_0}$ 中的一个像素子集)中的每个像素映射到参考帧 $\hat{X}_{ref_1}$ (或参考帧 $\hat{X}_{ref_1}$ 中的相应像素子集)中对应像素的位置,反之亦然(例如,在两个方向上)。在一个例子中,每个像素的运动估计信息(例如,光流信息)可以包括用于指示像素在参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 之间的移动的向量(例如,像素从参考帧 $\hat{X}_{ref_0}$ 移动到参考帧 $\hat{X}_{ref_1}$ 和/或像素从参考帧 $\hat{X}_{ref_1}$ 移动到参考帧 $\hat{X}_{ref_0}$)。在一些例子中,可以基于参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 之间的光流向量来生成光流图。该光流图可以包括帧中每个像素的光流向量,其中每个向量指示像素在帧之间的移动。

[0145] 在一个说明性示例中,像素的光流向量可以是位移向量(例如,其指示水平和垂直位移,例如x位移和y位移),其显示像素从第一帧到第二帧的移动。在一些情况下,帧插值引擎722可以计算参考帧 $\hat{X}_{ref_0}$ 的二维(2D)流图(例如,指示水平和垂直位移)和参考帧 $\hat{X}_{ref_1}$ 的2D流图。在一些方面,帧插值引擎722可以合并2D流图以生成用于内插参考帧 $\hat{X}_{ref_t}$ 的三维(3D)流图。该3D流图可以包括分别指示水平和垂直位移的水平尺寸和垂直尺寸,以及包括标识水平位移和垂直位移应用到的帧的帧索引的帧尺寸。例如,如果帧尺寸标识参考帧 $\hat{X}_{ref_0}$,则水平和垂直尺寸可以指示参考帧 $\hat{X}_{ref_0}$ 中像素的水平和垂直位移(例如, $\hat{X}_{ref_0}$ 中的像素应当根据光流信息而移动多少像素)。

[0146] 帧插值引擎722可以使用插值后的光流(例如,3D流图)将参考帧 $\hat{X}_{ref_0}$ 和参考帧 $\hat{X}_{ref_1}$

变形到输入帧 X_t 的特定时间步长(例如,时间 t),并且可以融合变形的帧以生成内插参考帧 $\hat{X}_{ref_t}$ 。下面关于图7C来描述帧插值引擎的这种示例实现的进一步细节。

[0147] 使用内插参考帧 $\hat{X}_{ref_t}$ 和输入帧 X_t 作为输入,P帧译码系统700可以执行与上面关于图7A描述的那些相同或相似的操作,来生成重构帧 $\hat{X}_t$ 。例如,运动预测系统702的编码器网络705可以确定输入帧 X_t 和内插参考帧 $\hat{X}_{ref_t}$ 之间的运动,并且可以将运动编码到潜在 z 中。运动预测系统702的解码器网络703可以对潜在 z 进行解码,以生成预测的运动 $\hat{f}_t$ 。变形引擎704可以使用预测的运动 $\hat{f}_t$ 来修改内插参考帧 $\hat{X}_{ref_t}$ 的像素(例如,通过执行运动补偿),以生成帧预测 $\hat{X}_t$,如上面参照图7A所描述的。残差预测系统706的编码器网络707可以获得残差 r_t (确定为帧预测 $\hat{X}_t$ 与输入帧 X_t 之间的差),并且可以将残差 r_t 编码到潜在 z_r 中。残差预测系统706的解码器网络709可以对潜在 z_r 进行解码,以生成预测残差 $\hat{r}_t$ 。然后,B帧译码系统720可以将帧预测 $\hat{X}_t$ 添加到预测残差 $\hat{r}_t$ 以生成重构帧 $\hat{X}_t$ 。

[0148] 图7C是说明可以由B帧译码系统使用的示例性帧插值引擎725的系统流程的图。帧插值引擎725是图7B的帧插值引擎722的说明性示例。在图7C的例子中,帧插值引擎725包括光流估计层730、光流预测层732、2D变形层734、细化层736和3D变形层738。

[0149] 如图7C中所示,参考帧 $\hat{X}_{ref_0}$ 和参考帧 $\hat{X}_{ref_1}$ 输入到光流估计层730。在一些例子中,光流估计层730可以包括一个或多个神经网络或其它机器学习系统。例如,可以使用一个或多个光流估计神经网络来实现光流估计层730。光流估计层730可以使用参考帧 $\hat{X}_{ref_0}$ 和参考帧 $\hat{X}_{ref_1}$ 来估计针对参考帧 $\hat{X}_{ref_0}$ 和参考帧 $\hat{X}_{ref_1}$ 的双向运动信息740(例如,前向光流和后向光流)。双向运动信息740可以包括从参考帧 $\hat{X}_{ref_0}$ 到参考帧 $\hat{X}_{ref_1}$ 的光流或光流图(表示为 $f_{0 \rightarrow 1}$),以及从参考帧 $\hat{X}_{ref_1}$ 到参考帧 $\hat{X}_{ref_0}$ 的光流或光流图(标记为 $f_{1 \rightarrow 0}$)。

[0150] 例如,双向运动信息740可以包括光流图 $f_{0 \rightarrow 1}$,其将参考帧 $\hat{X}_{ref_0}$ (或参考帧 $\hat{X}_{ref_0}$ 的一个像素子集)中的每个像素的位置映射到参考帧 $\hat{X}_{ref_1}$ (或参考帧 $\hat{X}_{ref_1}$ 中的相应像素子集)中相应像素的位置。双向运动信息740还可以包括光流图 $f_{1 \rightarrow 0}$,其将参考帧 $\hat{X}_{ref_1}$ (或参考帧 $\hat{X}_{ref_1}$ 的一个像素子集)中的每个像素的位置映射到参考帧 $\hat{X}_{ref_0}$ (或参考帧 $\hat{X}_{ref_0}$ 中的相应像素子集)中相应像素的位置。在一些情况下,每个像素的运动信息740可以包括指示像素在参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 之间的移动的向量。在一些例子中,可以基于参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 的像素之间的光流向量,来生成光流图 $f_{0 \rightarrow 1}$ 和 $f_{1 \rightarrow 0}$ 。光流图可以包括针对帧中每个像素的光流向量,其中每个向量指示像素在帧之间的移动。

[0151] 光流估计层730可以向光流预测层732提供双向运动信息740(例如, $f_{0 \rightarrow 1}$ 和 $f_{1 \rightarrow 0}$)。光流预测层732可以获得指示时间间隔0到1中的时间 t 的时间信息742(例如, $t \in (0, 1)$),其中0表示参考帧 $\hat{X}_{ref_0}$ 的时间步长,1表示参考帧 $\hat{X}_{ref_1}$ 的时间步长,并且时间 t 对应于要内插的帧(例如,内插帧 $\hat{X}_{ref_t}$)的时间步长,其也对应于输入帧(例如,图7B的输入帧 X_t)的时间步长。光流预测层732可以使用双向运动信息740和时间信息742来预测2D光流图744。在一些例子中,光流估计层730可以针对参考帧 $\hat{X}_{ref_0}$ 和参考帧 $\hat{X}_{ref_1}$,对双向运动信息740(例如, $f_{0 \rightarrow 1}$ 和 $f_{1 \rightarrow 0}$)进行插值,以确定时间 t 的光流。在一些情况下,时间 t 的光流可以是 $f_{0 \rightarrow 1}$ 和 $f_{1 \rightarrow 0}$ 之间的

线性插值。

[0152] 在一些例子中,2D光流图744可以包括从时间t到参考帧 $\hat{X}_{ref_0}$ (例如,到参考帧 $\hat{X}_{ref_0}$ 的时间步长0)的光流图(例如, $\hat{f}_{t \rightarrow 0}^{2D}$)、以及从时间t到参考帧 $\hat{X}_{ref_1}$ (例如,到参考帧 $\hat{X}_{ref_1}$ 的时间步长1)的光流图(例如, $\hat{f}_{t \rightarrow 1}^{2D}$)。例如,2D光流图744可以包括具有从时间t到参考帧 $\hat{X}_{ref_0}$ 的时间序列(例如,图7C所示示例中的时间序列0)的运动信息的反向2D流图 $\hat{f}_{t \rightarrow 0}^{2D}$ 、以及具有从时间t到参考帧 $\hat{X}_{ref_1}$ 的时间序列(例如,图7C所示示例中的时间序列1)的运动信息的前向2D流图 $\hat{f}_{t \rightarrow 1}^{2D}$ 。

[0153] 在一个说明性示例中,光流预测层732可以根据式1和式2来预测2D光流图744,如下所示:

$$[0154] \quad \hat{f}_{t \rightarrow 0}^{2D} = -1(1-t)tf_{0 \rightarrow 1} + t^2f_{1 \rightarrow 0} \quad \text{式(1)}$$

$$[0155] \quad \hat{f}_{t \rightarrow 1}^{2D} = 1(1-t)^2f_{0 \rightarrow 1} - t(1-t)f_{1 \rightarrow 0} \quad \text{式(2)}$$

[0156] 在上面的式1和式2中,t表示在参考帧 $\hat{X}_{ref_0}$ 的时间步长和参考帧 $\hat{X}_{ref_1}$ 的时间步长(例如,0到1区间的时间步长)之间的内插帧 $\hat{X}_{ref_t}$ (对应于输入帧 X_t)的时间步长, $f_{0 \rightarrow 1}$ 表示从参考帧 $\hat{X}_{ref_0}$ 到参考帧 $\hat{X}_{ref_1}$ 的光流, $f_{1 \rightarrow 0}$ 表示从参考帧 $\hat{X}_{ref_1}$ 到参考帧 $\hat{X}_{ref_0}$ 的光流, $\hat{f}_{t \rightarrow 0}^{2D}$ 表示标识从t到与参考帧 $\hat{X}_{ref_0}$ 相关联的时间步长(例如,图7C所示示例中的时间步长0)的反向运动的2D光流图, $\hat{f}_{t \rightarrow 1}^{2D}$ 表示标识从t到与参考帧 $\hat{X}_{ref_1}$ 相关联的时间步长(例如,图7C所示示例中的时间步长1)的前向运动的2D光流图。

[0157] 2D变形层734可以获得2D光流图744、参考帧 $\hat{X}_{ref_0}$ 和参考帧 $\hat{X}_{ref_1}$,并且可以生成变形帧746。变形帧746可以包括对应于参考帧 $\hat{X}_{ref_0}$ 的变形帧(W_{x_0})。2D变形层734可以通过将2D光流图 $\hat{f}_{t \rightarrow 0}^{2D}$ 应用于参考帧 $\hat{X}_{ref_0}$ 来生成变形帧 W_{x_0} 。变形帧746还可以包括对应于参考帧 $\hat{X}_{ref_1}$ 的变形帧(W_{x_1})。2D变形层734可以通过将2D光流图 $\hat{f}_{t \rightarrow 1}^{2D}$ 应用于参考帧 $\hat{X}_{ref_1}$ 来生成变形帧 W_{x_1} 。2D变形层734可以实现一个或多个变形函数,以将参考帧 $\hat{X}_{ref_0}$ 变形到时间步t,并且将参考帧 $\hat{X}_{ref_1}$ 变形到时间步t。

[0158] 在一些例子中,为了生成变形帧 W_{x_0} ,2D变形层734可以使用为参考帧 $\hat{X}_{ref_0}$ 计算的2D光流图 $\hat{f}_{t \rightarrow 0}^{2D}$,以将参考帧 $\hat{X}_{ref_0}$ 中的像素移动到对应于和/或相对于时间t的各个位置。类似地,为了生成变形帧 W_{x_1} ,2D变形层734可以使用为参考帧 $\hat{X}_{ref_1}$ 计算的2D光流图 $\hat{f}_{t \rightarrow 1}^{2D}$,将参考帧 $\hat{X}_{ref_1}$ 中的像素移动到对应于和/或相对于时间t的位置。在一些情况下,变形帧746可以包括用于参考帧 $\hat{X}_{ref_0}$ 的2D帧(其考虑了在时间t的像素(和/或像素位置)的垂直和/或水平位移)和用于参考帧 $\hat{X}_{ref_1}$ 的2D帧(其考虑了在时间t的像素(和/或像素位置)的垂直和/或水平位移)。

[0159] 细化层736可以获得2D光流图744、变形帧746以及参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 作为输入,并且可以使用输入数据来执行细化操作。在一些例子中,细化层736可以是帧插值引擎725的可选组件。在一些情况下,细化层736可以包括一个或多个神经网络或其它机器学习系统。在一个说明性示例中,细化层736可以包括一个或多个插值卷积神经网络(CNN)。在一些例

子中,由细化层736执行的细化操作可以包括:细化和/或校正2D光流图744中的值。例如,在一些情况下,对应于参考帧 $\hat{X}_{ref_0}$ 的时间步长0和对应于参考帧 $\hat{X}_{ref_1}$ 的时间步长1之间的一些或全部移动可能不是线性的。细化层736可以检测和校正这种非线性运动。细化层736可以生成解释这种非线性的细化2D光流。另外地或替代地,细化层736可以细化或校正任何变形帧746(和/或参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 中的任何一个)中的伪影,细化或校正任何变形帧746(和/或参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 中的任何一个)的一部分(例如,边缘),进行细化或校正以解决任何变形帧746(和/或参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 中的任何一个)中的遮挡,和/或执行图像数据和/或与变形帧746和/或参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 相关联的光流信息的任何其它细化。

[0160] 在一些例子中,细化操作还可以包括:将2D光流图744合并到用于内插帧 $\hat{X}_{ref_t}$ 的3D光流图748中。例如,如先前所描述的,2D光流图744可以包括用于参考帧 $\hat{X}_{ref_0}$ 的2D光流图(例如, $\hat{f}_{t \rightarrow 0}^{2D}$)和用于参考帧 $\hat{X}_{ref_1}$ 的2D光流图(例如, $\hat{f}_{t \rightarrow 1}^{2D}$)。细化层736可以合并2D光流图744(例如,2D光流图 $\hat{f}_{t \rightarrow 0}^{2D}$ 和2D光流图 $\hat{f}_{t \rightarrow 1}^{2D}$)以生成3D光流图748。3D光流图748可以将特定帧(例如,参考帧 $\hat{X}_{ref_0}$ 和参考帧 $\hat{X}_{ref_1}$)映射到对应的光流信息。例如,3D光流图748可以包括两个空间维度(例如,X和Y轴或者垂直和水平维度)和包括帧信息的第三维度(例如,帧维度)。第三维度中的帧信息可以识别运动信息(例如,垂直和/或水平位移)在两个空间维度上适用的帧(例如,参考帧 $\hat{X}_{ref_0}$ 或参考帧 $\hat{X}_{ref_1}$)。

[0161] 例如,如果第三维度(例如,帧维度)标识参考帧 $\hat{X}_{ref_0}$,则空间维度中的对应值(例如,在X和Y轴或垂直和水平维度上)可以指示参考帧 $\hat{X}_{ref_0}$ 中像素的水平和/或垂直位移(例如,参考帧 $\hat{X}_{ref_0}$ 中的像素应基于空间维度中的光流信息来移动多少)。类似地,如果第三维度(例如,帧维度)标识参考帧 $\hat{X}_{ref_1}$,则空间维度中的对应值(例如,在X和Y轴或垂直和水平维度上)可以指示参考帧 $\hat{X}_{ref_1}$ 中像素的水平和/或垂直位移(例如,参考帧 $\hat{X}_{ref_1}$ 中的像素应基于空间维度中的光流信息来移动多少)。

[0162] 在一些情况下,细化层736可以使用2D光流图744将2D光流图744合并到3D光流图748中。在一些例子中,细化层736可以使用通过细化2D光流图744生成的经细化将2D光流,将2D光流图744合并到3D光流图748中。在一些情况下,细化层736可以使用参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 、变形帧746和/或2D光流图744(或细化的2D光流)作为输入,来生成3D光流图748。

[0163] 3D变形层738可以使用3D光流图748和参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ (和/或变形帧746)来生成与时间步长t相对应(与输入帧 X_t 的时间步长相对应)的内插参考帧 $\hat{X}_{ref_t}$ 。例如,3D变形层738可以使用3D光流图748对参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ (或变形帧746)进行变形,并融合变形的帧(例如,变形参考帧 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 或变形帧746),以生成内插参考帧 $\hat{X}_{ref_t}$ 。在一些情况下,3D变形层738可以实现一个或多个变形函数。在一个说明性示例中,可以使用下式来生成内插帧:

[0164] $\hat{I}_t = \alpha_0 \odot g(I_0, F_{t \rightarrow 0}) + (1 - \alpha_0) \odot g(I_1, F_{t \rightarrow 1})$, 式(3)

[0165] 其中 $g(\cdot, \cdot)$ 是后向变形函数(在一些情况下可以使用双线性插值来实现),参数 α_0 控制两个输入图像 I_0 和 I_1 的贡献,并取决于时间一致性和遮挡推理,而 $\odot$ 项表示元素乘

法(其导致基于 ∞_0 参数的输入图像的内容感知权重)。

[0166] 在一些情况下,可以使用一个或多个损失函数来训练和/或优化光流估计层730、光流预测层732、2D变形层734、细化层736和/或3D变形层738。在一些情况下,光流估计层730、光流预测层732、2D变形层734、细化层736和3D变形层738可以使用一个或多个损失函数进行端到端训练。在一个说明性示例中,可以使用上面所描述的损失($Loss=D+BR$)。

[0167] 虽然关于使用光流插值来描述了图7C,但应当注意,光流插值在本文中用作说明性示例以用于解释目的。本领域普通技术人员应当认识到,图7B的帧插值引擎722和/或图7C的帧插值引擎725可以实现任何其它帧插值技术。

[0168] 图8是说明可以执行以确定参考帧(例如,从参考帧 $\hat{X}_{ref_0}$ 到参考帧 $\hat{X}_{ref_1}$ 或反之亦然)之间的运动信息的不同类型的运动估计的图。图8中的运动估计可以用于确定内插参考帧(例如,如关于图7B和图7C所描述的)和/或对参考帧进行变形(例如,通过图7A和图7B的变形引擎704)。在图8中,项x表示可以从其估计运动的参考帧,项f表示运动估计,并且项y表示可以如下计算的变形帧: $y=f(x)$ 。

[0169] 一种类型的运动估计是基于块的运动估计技术802。可以在逐块的基础上执行基于块的运动估计。例如,对于帧y中的每个块,运动估计f定义了帧x中相应块的位置。在一个说明性示例中,运动估计f可以包括指示帧y中的块相对于帧x中的对应块的位移(例如,水平和垂直位移)的运动向量。通过确定块之间的相似性(例如,像素值的相似性),可以确定来自帧x的块对应于帧y中的块。

[0170] 可以执行的另一种类型的运动估计是光流运动估计技术804。可以在逐个像素的基础上执行光流运动估计。例如,对于帧y中的每个像素,运动估计f定义了帧x中相应像素的位置。每个像素的运动估计f可以包括指示像素在帧之间的移动的向量。在一些情况下,可以基于帧之间的光流向量的计算,来生成光流图(也称为运动向量图)。光流图可以包括针对帧中每个像素的光流向量,其中每个向量指示像素在帧之间的移动。在一个说明性示例中,像素的光流向量可以是位移向量(例如,其指示水平和垂直位移,比如x位移和y位移),显示像素从第一帧到第二帧的移动。

[0171] 在某些情况下,光流图可以包括帧中少于所有像素的向量。例如,可以在帧之间计算密集光流,以生成针对帧中每个像素的光流向量,该向量可以包含在密集光流图中。在一些例子中,每个光流图可以包括2D向量场,每个向量是位移向量,其显示了点从第一帧到第二帧的移动。

[0172] 如上所述,可以在帧序列中的帧之间计算光流向量或光流图。两个帧可以包括作为连续捕获的帧的两个直接相邻的帧,也可以包括在帧序列中相隔一定距离(例如,在彼此的两个帧内、在彼此的三个帧内、或者任何其它适当的距离)的两个帧。在一个说明性示例中,帧x中的像素 $I(x,y,t)$ 可以在帧y中移动某个距离或位移($\Delta x, \Delta y$)。

[0173] 可以执行的另一种类型的运动估计是动态卷积(也称为自适应卷积)运动估计技术806。可以逐个像素地执行动态卷积运动估计。例如,对于帧y中的每个像素,运动估计f定义了帧x中像素邻域的线性组合。如图8中所示,对于帧y中某个像素位置处的像素,围绕帧x中相同像素位置处的像素周围的像素邻域定义一个内核或矩阵。对帧y中的位置(2,7)(对应于(行号,列号))的像素执行图8中的示例性动态卷积运动估计,其中帧y中的左上像素位置在位置(0,0)。围绕帧x中位置(2,7)处的像素来定义内核。例如,如图所示,将大小为5x5

的内核应用于帧 x 中位置(2,7)处的像素周围的像素邻域。

[0174] 内核或矩阵中的每个条目可以包括一个数字(例如,图8的示例中的一组 5×5 数字)。内核中的数字集可以描述如何组合来自在帧 x 中的该像素位置处的像素周围的像素,其中该像素位置对应于在帧 y 中正在映射和/或重建的像素的相同像素位置。

[0175] 图9和图10是示出图片组(GoP)结构(对应于GoP中的帧的顺序)的不同示例的图。图9示出了I帧、B帧、I帧(IBC)译码技术的示例。例如,IBC技术可以用于传统译码系统(例如,根据诸如HEVC或VVC之类的一种或多种视频译码标准进行操作)中的随机存取模式。如图所示,对于第一GoP(标记为GoP1),译码系统(例如,图6的译码系统600)可以将第一帧和最后帧译码(例如,编码/压缩和/或解码/解压缩)为I帧,并且可以将中间帧压缩为B帧。如标记为“Ref”的行中的箭头所示,两个I帧用作B帧压缩的参考帧。标记有“顺序”的行表示首先对I帧进行译码,然后对中间的B帧进行译码。在第二GoP(标记为GoP2)中,可以使用来自GoP1的参考帧对B帧进行译码,因此不需要将GoP2中的第一帧译码为I帧。可以使用GoP1中的最后帧(I帧)并使用最后帧(也是I帧)作为参考帧,对GoP2中的B帧进行译码。图6的B帧译码系统600或图7B的B帧译码系统720可以用于对B帧进行译码。例如,图9的GoP1中的第一个I帧可以是图7B中的参考帧 $\hat{X}_{ref_0}$,图9的GoP1中的最后一个I帧可以是图7B中的参考帧 $\hat{X}_{ref_1}$ 。图9的GoP1中的B帧可以是图7B中的重构帧 $\hat{X}_t$,在这种情况下,可以使用GoP1中的两个I帧来生成内插帧($\hat{X}_{ref_t}$)。

[0176] 图10是示出I帧、B帧、P帧(IBP)译码技术的示例的图。例如,因为本文所描述的B帧译码系统(例如,图7B的B帧译码系统720)使用P帧译码系统(例如,图7A和图7B的P帧译码系统700),当P帧译码系统用于单向预测和双向预测时,如果参数(例如,权重、偏差等)是常见的,则可以使用相同的组件生成P帧,而不增加额外的复杂性。在图10的例子中,每个GoP的最后一帧不必是I帧(如图9的IBC GoP结构中的情况),而是可以译码为P帧。与图9的译码顺序相比,图10中的译码顺序也不同。例如,首先将图1中的GoP1的第一帧译码(例如,编码/压缩和/或解码/解压缩)为I帧。使用I帧作为参考帧,可以将GoP1中的最后一个帧译码为P帧(例如,通过图7A的P帧译码系统700)。然后,GoP1中的I帧和P帧可以用于将中间帧译码为B帧(例如,使用图6的B帧译码系统600或图7B的B帧译码系统720)。对于GoP2,GoP1的最后一帧可以用作(例如,由图7A的P-帧译码系统700用作)参考帧,以用于将GoP2的最后一帧译码为P-帧。GoP1的最后一帧和GoP2的最后一帧可以用作(例如,由图6的B帧译码系统600或图7B的B帧译码系统720用作)参考帧,以用于将GoP2的中间帧译码为B帧。

[0177] 图11和图12是说明GoP内的帧译码顺序的图。例如,一旦译码系统(例如,I帧译码系统和/或P帧译码系统)确定了用于GoP的两个参考帧(例如,图11和图12的GoP中的第一个帧和最后一个帧),B帧译码系统(例如,由图6的B帧译码系统600或图7B的B帧译码系统720)可以开始将GoP的中间帧译码为B-帧。图11和图12的示例译码顺序定义了对那些B帧进行译码(例如,编码/压缩和/或解码/解压缩)的顺序的示例。在一个说明性示例中,图11和图12的GoP的第一个帧和最后一个帧可以包括I帧,类似于图9中所示的。在另一个说明性示例中,图11和图12的GoP的第一帧可以包括I帧,并且最后帧可以包括P帧,类似于图10所示的。

[0178] 图11示出了用于双向预测的顺序方案的例子。在图11的顺序方案中,从左到右顺序地译码B帧(例如,由图6的B帧译码系统600或图7B的B帧译码系统720进行译码)。如图所

示,在将GoP的第一帧和最后一帧用作参考帧的情况下,首先使用双向预测对序号为2的B帧进行译码。接下来在将前一个B帧(序号为2)和GoP的最后一帧用作参考帧的情况下,使用双向预测对序号为3的B帧进行译码。在序号为3的B帧之后,对序号为4的B帧进行译码,然后是序号为5的B帧,以此类推。

[0179] 图12示出了用于双向帧预测的分层方案的例子。在图12的分层方案中,一旦两个初始参考帧(GoP的第一帧和最后一帧)可用,译码系统(例如,图6的B帧译码系统600或图7B的B帧译码系统720)就可以在双向预测中使用这两个初始参考帧,将第五帧(序号为2的中间帧)译码为第一B帧。然后,译码系统可以使用图12中“Ref”行指示的参考帧,对GoP中的第三帧或第七帧(都具有序号3)进行译码。例如,可以使用GoP的第一帧(其可以是I帧)和GoP的第五帧(其可以是B帧)作为参考帧,对第三帧(序号为3)进行译码,并且可以使用GoP的最后一帧(其可以是I帧或P帧)和GoP的第五帧(B帧)作为参考帧,对第七帧(序号为3)进行译码。然后,可以使用由图12中的“Ref”行指示的参考帧,以任何顺序对标有序号4的帧进行译码。

[0180] 图13A和图13B是示出图6的B帧译码系统600(例如,使用图7B的B帧译码系统720实现)相对于不使用内插参考帧的基线译码系统的性能的图表。在图13A和图13B的图表中说明的所有三个译码系统是使用相同训练条件(例如,使用相同的训练数据集、相同的损失函数和相同的训练迭代次数等)训练的基于神经网络的译码系统。这些图表通过绘制速率(表示为每像素位数)作为峰值信噪比(PSNR)的函数来表示性能。线1302示出了当使用图12的分层译码顺序时,图7B的B帧译码系统720的性能。线1304示出了当使用图11的顺序译码顺序时,B帧译码系统720的性能。线1306示出了基线译码系统的性能。这些图的左上角(较低的速率,较高的PSNR)表示更好的性能。如图13A和图13B中所示,线1302(使用分层译码的B帧译码系统720)表示最高性能,其次是线1304(使用顺序译码的B帧译码系统720),其次是线1306(基线译码系统)。

[0181] 图14A至图14F是依据每帧速率(如标记为速率与帧索引的图表所示)、每帧的PSNR(如标记为PSNR与帧索引的图表所示)、以及每帧的多尺度结构相似性(MS-SSIM)(如标记为MS-SSIM与帧索引的图表所示),示出图6的B帧译码系统600(例如,使用图7B的B帧译码系统720实现)的性能的图表。这些图表的x轴包括表示帧顺序的帧索引,并且y轴包括各种性能指标(例如,速率、PSNR和MS-SSIM)。在图14A至图14F的例子中,每个GoP中的帧数是12,并且这些图示出了用于生成结果的视频序列中所有GoP的平均值。

[0182] 使用帧内预测,将每个GoP序列中的第一帧(具有帧索引0)生成I帧(例如,如图9和图10的GoP结构所示)。图14A中的点1402、图14B中的点1404、图14C中的点1406、图14D中的点1408、图14E中的点1410和图14E中的点1412表示I帧的度量(速率、PSNR或MS-SSIM)。例如,图14A中的点1402显示了用于该I帧的压缩的速率(比特数)。使用单向预测,将每个GoP序列中的最后一帧(具有帧索引11)生成P帧(例如,由图7A的P帧译码系统700生成)。图14A中的点1414、图14B中的点1416、图14C中的点1418、图14D中的点1420、图14E中的点1422和图14E中的点1424表示P帧的度量(速率、PSNR或MS-SSIM)。例如,如图14A中所示,与用于I帧的比特相比,更少的比特用于P帧。

[0183] 将第一帧和最后一帧之间的帧译码为B帧(例如,由图7B的B帧译码系统720执行)。显示了B帧的两组结果,一组使用图11的顺序译码顺序,并且另一组使用图12的分层译码顺

序。参照图14A作为示例,线1426表示使用顺序译码顺序的结果,其中首先将帧索引为1的帧译码为B帧,接着是帧索引为2的帧,依此类推。图15A中的线1428表示使用分层译码顺序的结果,其中首先将帧索引为5的帧译码为B帧,接着是帧索引为2的帧,接着是帧索引为8的帧,依此类推。如图14A中所示,基于参考帧到B帧的距离对于后续译码的B帧较小,较少的比特用于后续译码的B帧,如图12中所示。例如,使用I帧(GoP的第一帧)和P帧(GoP的最后一帧)作为参考帧,对帧索引为5的帧进行双向预测,而使用帧索引为2的I帧和B帧作为参考帧,对帧索引为1的帧进行双向预测。图14B-F中光滑的线条表示使用图12的分层译码顺序的相应结果,并且图14B-F中的锯齿线表示使用图11的顺序译码顺序的相应结果。

[0184] 图15是说明用于使用本文描述的技术来执行帧的双向预测的处理1500的示例的流程图。在框1502处,处理1500包括:获得第一参考帧和第二参考帧。在一个说明性示例中,第一参考帧可以包括图6的参考帧0(Ref₀),第二参考帧可以包括图6的参考帧1(Ref₁)。在另一个说明性示例中,第一参考帧可以包括图7B和/或图7C的参考帧0($\hat{X}_{ref0}$),第二参考帧可以包括图7B和/或图7C的参考帧1($\hat{X}_{ref1}$)。

[0185] 在框1504处,处理1500包括:至少部分地通过在第一参考帧和第二参考帧之间执行插值来生成第三参考帧。例如,处理1500可以包括:确定表示从第一参考帧到第三参考帧的像素运动的第一组运动信息。处理1500还可以包括:确定表示从第二参考帧到第三参考帧的像素运动的第二组运动信息。在一些方面,第一组运动信息和第二组运动信息是基于第一参考帧和第二参考帧之间的像素运动来确定的。在一些情况下,第一组运动信息包括第一光流图,而第二组运动信息包括第二光流图。参考图7C的描述作为一个说明性示例,第一运动信息可以包括上面使用式(1)定义的 $\hat{f}_{t \rightarrow 0}^{2D}$,第二运动信息可以包括上面使用式(2)定义的 $\hat{f}_{t \rightarrow 1}^{2D}$ 。处理1500可以包括:至少部分地通过使用第一组运动信息对第一参考帧执行变形函数,来生成第一变形信息。在一些方面,变形函数包括双线性插值函数,例如上面关于图7C所描述的。处理1500还可以包括:至少部分地通过使用第二组运动信息对第二参考帧执行变形函数来生成第二变形信息。处理1500可以包括:基于第一变形信息和第二变形信息来生成第三参考帧。

[0186] 在框1506处,处理1500包括:基于第三参考帧对输入帧执行单向帧间预测。例如,为了执行单向帧间预测,处理1500可以包括:估计输入帧和第三参考帧之间的运动。在一个说明性示例中,预测运动可以包括图7B的预测运动 $\hat{f}_t$ 。在一些情况下,输入帧和第三参考帧之间的估计运动包括光流数据(例如,使用图8中所示的光流运动估计技术804来确定)。在一些方面,输入帧和第三参考帧之间的估计运动包括动态卷积数据(例如,使用图8所示的动态卷积运动估计技术806来确定)。在一些方面,输入帧和第三参考帧之间的估计运动包括基于块的运动数据(例如,使用图8中所示的基于块的运动估计技术802来确定)。处理1500还可以包括:至少部分地通过基于估计的运动,变形第三参考帧的一个或多个像素来生成变形帧。例如,参照图7B作为说明性示例,变形引擎704可以基于预测运动 $\hat{f}_t$ 中包括的运动向量(或其它运动数据)来变形内插参考帧 $\hat{X}_{ref,t}$ 的像素。

[0187] 在一些例子中,为了进一步执行单向帧间预测,处理1500可以包括:至少部分地通过确定输入帧和变形帧之间的差异来确定残差。处理1500还可以包括:使用残差来生成预测残差。参照图7B作为说明性示例,P帧译码系统700可以生成残差 r_t 作为帧t预测 $\hat{X}_t$ 和输入

帧 X_t 之间的差。残差预测系统706可以使用上面所描述的技术来生成预测残差 $\hat{r}_t$ 。

[0188] 在一些方面,使用第一神经网络(例如,第一自动编码器或其它类型的神经网络)来估计运动,并且使用第二神经网络(例如,第二自动编码器或其它类型的神经网络)来生成预测残差。在一个说明性示例中,第一神经网络可以包括图7B的运动预测系统702,第二神经网络可以包括图7B的残差预测系统706。在一些例子中,处理1500可以包括:基于变形帧和预测的残差,生成表示输入帧的重构帧。该重构帧包括双向预测帧。再次参照图7B作为说明性示例,P帧译码系统720可以通过将帧 t 预测 $\tilde{X}_t$ (通过对内插参考帧 $\hat{X}_{ref,t}$ 进行变形来生成)与预测的残差 $\hat{r}_t$ 相加来生成重构帧 $t(\hat{X}_t)$ 。在一些情况下,处理1500还可以包括:使重构帧存储在存储器(例如,解码图片缓冲器(DPB)或其它存储器)中。在一些例子中,存储的重构帧可以用于其它帧的帧间预测。

[0189] 如上所述,本文所描述的P帧译码系统(例如,图7A的P帧译码系统700)可以使用内插参考帧(例如,内插参考帧 $\hat{X}_{ref,t}$)来执行双向预测,也可以执行单向预测。例如,为了执行单向预测,处理1500可以包括:获得第四参考帧(例如,参考帧 $\hat{X}_{ref}$)。处理1500还可以包括:估计另外输入帧和第四参考帧之间的运动(例如,使用运动预测系统702)。在一个说明性示例中,估计的运动可以包括图7A中所示的预测运动 $\hat{f}_t$ 。处理1500可以包括:至少部分地通过基于估计的运动来变形第四参考帧的一个或多个像素,来生成另外的变形帧(例如,使用变形引擎704)。再次参照图7A作为说明性示例,另外的变形帧可以包括帧预测 $\tilde{X}_t$ 。处理1500可以包括:基于另外的变形帧和另外的预测残差,生成表示另外输入帧的重构帧,在这种情况下,表示另外输入帧的重构帧包括单向预测帧。参照图7A作为说明性示例,P帧译码系统720可以通过将帧 t 预测 $\tilde{X}_t$ (通过变形参考帧 $\hat{X}_{ref}$ 生成)添加到预测残差 $\hat{r}_t$ 来生成重构帧 $t(\hat{X}_t)$ 。在一些情况下,处理1500可以包括:使表示另外输入帧的重构帧存储在存储器(例如,解码图片缓冲器(DPB)或其它存储器)中。在一些例子中,存储的表示另外输入帧的重构帧可以用于其它帧的帧间预测。

[0190] 图16是说明用于使用本文描述的技术,来执行帧的双向预测的处理1600的另一示例的流程图。在一些情况下,处理1600的至少一部分可以由运动预测系统702的解码器网络703和残差预测系统706的解码器网络709来执行。在框1602处,处理1600包括:获得内插参考帧和表示输入帧与内插参考帧之间的运动的运动信息。在一个说明性示例中,内插参考帧可以包括图7B的内插参考帧 $\hat{X}_{ref,t}$ 。在一些方面,运动信息是从比特流中获得的。例如,运动信息可以包括图7B中所示的潜在 z_m 。在框1604处,处理1600包括:基于运动信息,估计输入帧和内插参考帧之间的运动。例如,运动预测系统702的解码器网络703可以基于潜在 z_m ,估计输入帧 X_t 和内插参考帧 $\hat{X}_{ref,t}$ 之间的运动。估计的运动可以包括图7B的预测运动 $\hat{f}_t$ 。

[0191] 在框1606处,处理1600包括:至少部分地通过基于估计的运动,对内插参考帧的一个或多个像素进行变形,来生成变形帧。例如,参照图7B作为说明性示例,变形引擎704可以基于包括在预测运动 $\hat{f}_t$ 中的运动向量(或其它运动数据),对内插参考帧 $\hat{X}_{ref,t}$ 的像素进行变形。

[0192] 在框1608处,处理1600包括:基于变形帧和预测残差,生成表示输入帧的重构帧。该重构帧包括双向预测帧。再次参照图7B作为说明性示例,P帧译码系统720可以通过将帧 t

预测 $\hat{X}_t$ (通过变形内插参考帧 $\hat{X}_{ref,t}$ 产生) 与预测残差 $\hat{r}_t$ 相加来产生重构帧 $t(\hat{X}_t)$ 。在一些情况下, 处理1500还可以包括: 使重构帧存储在存储器 (例如, 解码图片缓冲器 (DPB) 或其它存储器) 中。在一些例子中, 存储的重构帧可以用于其它帧的帧间预测。

[0193] 在一些例子中, 处理1600可以包括: 获得第一参考帧和第二参考帧。处理1600还可以包括: 至少部分地通过在第一参考帧和第二参考帧之间执行插值来生成内插参考帧。本文描述的技术 (例如, 关于图7C和/或图15描述的技术) 可以用于生成内插参考帧。

[0194] 在一些例子中, 处理1600可以包括: 从比特流中获得残差。例如, 残差预测系统706的解码器网络709可以从比特流中获得潜在 z_r 。潜在 z_r 表示残差 r_t 。处理1600还可以包括: 基于获得的残差来生成预测残差。例如, 参照图7B作为说明性示例, 解码器网络709可以使用潜在 z_r 作为输入来生成预测残差 $\hat{r}_t$ 。

[0195] 在一些方面, 类似于上面关于处理1500所描述的, 处理1600可以包括: 获得第四参考帧, 估计其它输入帧和第四参考帧之间的运动, 并至少部分地通过基于估计的运动, 对第四参考帧的一个或多个像素进行变形, 来生成另外的变形帧。处理1600可以包括: 基于另外变形帧和另外的预测残差, 生成表示另外输入帧的重构帧, 表示另外输入帧的重构帧包括单向预测帧。

[0196] 在一些例子中, 处理1500和/或处理1600可以包括: 针对一个或多个训练迭代, 使用损失函数 (例如, Loss) 来训练神经网络系统。在一些情况下, 损失函数可以确定一个或多个重构的B帧与一个或多个重构的B帧表示的一个或多个帧之间的损失。在一些例子中, 第一损失函数可以等于 $D+\beta R$, 如先前所描述的。

[0197] 在一些例子中, 本文所描述的处理 (例如, 处理1500、处理1600和/或本文所描述的其它处理) 可以由计算设备或装置 (例如, 具有图17所示的计算设备架构1700的计算设备) 来执行。该计算设备可以包括任何适当的设备, 例如自主车辆计算机、机器人设备、移动设备 (例如, 移动电话)、桌面型计算设备、平板计算设备、可穿戴设备、服务器设备、视频游戏设备、扩展现实设备 (例如, 虚拟现实 (VR) 设备、增强现实 (AR) 设备或混合现实 (MR) 设备)、照相机设备、机顶盒设备、和/或具有资源能力来执行本文所描述的处理的任何其它计算设备, 这些处理包括处理800。在一些例子中, 该计算设备可以包括移动设备、可穿戴设备、XR设备、个人计算机、膝上型计算机、视频服务器、电视、照相机、机顶盒、视频游戏控制台、或其它设备。在一些例子中, 处理1500和/或处理1600可以由具有实现图6的B帧译码系统600和/或图7的B帧译码系统720的计算设备架构1700的计算设备来执行。

[0198] 在一些情况下, 计算设备或装置可以包括各种组件, 例如一个或多个输入设备、一个或多个输出设备、一个或多个处理器、一个或多个微处理器、一个或多个微型计算机、一个或多个发射器、接收器、或者组合的发射器接收器 (例如, 称为收发器)、一个或多个照相机、一个或多个传感器、和/或被配置为执行本文所描述的处理的步骤的其它组件。在一些例子中, 计算设备可以包括显示器、被配置为传送和/或接收数据的网络接口、它们的任何组合、和/或其它组件。网络接口可以被配置为传送和/或接收基于互联网协议 (IP) 的数据或其它类型的数据。

[0199] 可以利用电路来实现计算设备的组件。例如, 这些组件可以包括和/或可以使用电子电路或其它电子硬件来实现, 这些电子硬件可以包括一个或多个可编程电子电路 (例如, 微处理器、图形处理单元 (GPU)、数字信号处理器 (DSP)、中央处理单元 (CPU)、神经处理单元

(NPU)、专用集成电路(ASIC)、现场可编程门阵列(FPGA)和/或其它适当的电子电路),和/或可以包括和/或使用计算机软件、固件或它们的任何组合来实现,以执行本文所描述的各种操作。

[0200] 将处理1500和处理1600示出为逻辑流程图,其操作表示可以以硬件、计算机指令或其组合来实现的一系列操作。在计算机指令的上下文中,这些操作表示存储在一个或多个计算机可读存储介质上的计算机可执行指令,当所述计算机可执行指令被一个或多个处理器执行时,执行所叙述的操作。通常,计算机可执行指令包括执行特定功能或实现特定数据类型例程、程序、对象、组件、数据结构等等。描述这些操作的顺序不应被理解为是限制性的,并且可以以任何顺序和/或并行地组合任意数量的所描述的操作以实现处理。

[0201] 另外,可以在配置有可执行指令的一个或多个计算机系统的控制下执行本文所描述的处理(包括处理1500和处理1600),并且可以将这些处理实现为在一个或多个处理器上共同执行的代码、通过硬件执行的代码或者其组合(例如,可执行指令、一个或多个计算机程序或一个或多个应用程序)。如上所述,可以例如以包括多个可由一个或多个处理器执行的指令的计算机程序的形式,将代码存储在计算机可读或机器可读存储介质上。计算机可读或机器可读存储介质可以是非临时性的。

[0202] 图17示出了可以实现本文所描述的各种技术的示例性计算设备的示例性计算设备架构1700。在一些例子中,该计算设备可以包括移动设备、可穿戴设备、XR设备、个人计算机、膝上型计算机、视频服务器、视频游戏控制台、机器人设备、机顶盒、电视机、相机、服务器或其它设备。例如,计算设备架构1700可以实现图6的B帧译码系统600、图7A的P帧译码系统700和/或图7B的B帧译码系统720。示出了使用诸如总线之类的连接1705彼此电通信的计算设备架构1700的组件。示例性计算设备架构1700包括处理单元(CPU或处理器)1710和计算设备连接1705,该计算设备连接1705将包括计算设备存储器1715(例如,只读存储器(ROM) 1720和随机存取存储器(RAM) 1725)的各种计算设备组件耦合到处理器1710。

[0203] 计算设备架构1700可以包括高速存储器的高速缓存,该高速存储器与处理器1710直接连接、紧密接近或作为处理器1710的一部分而集成。计算设备架构1700可以将数据从存储器1715和/或存储设备1730复制到高速缓存1712以供处理器1710快速访问。用此方式,高速缓存可以提供性能提升,从而避免了处理器1710在等待数据时的延迟。这些和其它模块可以控制或者被配置为控制处理器1710执行各种动作。其它计算设备存储器1715也可以可用。存储器1715可以包括具有不同性能特征的多种不同类型的存储器。处理器1710可以包括任何通用处理器和硬件或存储在存储设备1730中的软件服务(例如,服务1 1732、服务2 1734和服务3 1736),并且被配置为控制处理器1710以及专用处理器,其中将软件指令合并到处理器设计方案中。处理器1710可以是包含多个核或处理器、总线、存储器控制器、高速缓存等等的自包含系统。多核处理器可以是对称的或非对称的。

[0204] 为了使用户能够与计算设备架构1700进行交互,输入设备1745可以表示任意数量的输入装置,例如用于语音的麦克风、用于手势或图形输入的触摸敏感型屏幕、键盘、鼠标、运动输入、语音等等。输出设备1735也可以是本领域普通技术人员已知的众多输出装置中的一种或多种,例如显示器、投影仪、电视、扬声器设备等等。在一些实例中,多模式计算设备可以使用户能够提供多种类型的输入,以与计算设备架构1700进行通信。通信接口1740通常可以控制和管理用户输入以及计算设备输出。对于在任何特定硬件装置上的操作没有

限制,因此在改进的硬件或固件装置得到开发时,本文的基本功能可以轻松地替换为改进的硬件或固件装置。

[0205] 存储设备1730是非易失性存储器,并且可以是能够存储能够由计算机访问的数据的硬盘或其它类型的计算机可读介质,例如磁带、闪存卡、固态存储设备、数字通用磁盘、盒式磁带、随机存取存储器(RAM) 1725、只读存储器(ROM) 1720以及其混合。存储设备1730可以包括用于控制处理器1710的服务1732、服务1734和服务1736。可以预期其它硬件或软件模块。存储设备1730可以连接到计算设备连接1705。在一个方面,执行特定功能的硬件模块可以包括结合必要的硬件组件(例如,处理器1710、连接1705、输出设备1735等等)存储在计算机可读介质中的用于执行该功能的软件组件。

[0206] 术语“设备”并不限于一个或特定数量的物理对象(例如,一个智能手机、一个控制器、一个处理系统等)。如本文所使用的,设备可以包括具有可以实施本公开内容的至少一些部分的一个或多个部分的任何电子设备。虽然说明书和示例使用术语“设备”来描述本公开内容的各个方面,但术语“设备”并不限于特定的配置、类型或对象数量。此外,术语“系统”不限于多个组件或特定实施例。例如,系统可以在一个或多个印刷电路板或其它衬底上实现,并且可以具有可移动或静态组件。尽管描述和示例使用术语“系统”来描述本公开内容的各个方面,但术语“系统”并不限于特定的配置、类型或对象数量。

[0207] 在以上的描述中提供了特定细节,以提供对本文提供的实施例和示例的透彻理解。但是,本领域普通技术人员应当理解,可以在没有这些特定细节的情况下实践这些实施例。为了清楚解释起见,在一些实例中,可以将本文的技术呈现为包括单独的功能块,该单独的功能块包括设备、设备组件、以软件体现的方法或例程、或者硬件和软件的组合。除了图中所示的和/或本文所述的组件以外,还可以使用其它组件。例如,可以将电路、系统、网络、过程和其它组件示出为框图形式的组件,以免在不必要的细节上使实施例晦涩难懂。在其它实例中,可以在没有不必要的细节的情况下,示出公知的电路、过程、算法、结构和技术,以避免使实施例晦涩难懂。

[0208] 在上文将各个实施例描述为过程或方法,将过程或方法描绘为流程图、流程、数据流程图、结构图或框图。虽然流程图可以将操作描述为顺序过程,但是许多操作可以并行或同时地执行。另外,可以重新安排这些操作的顺序。流程在其操作完成时终止,但可以有未包含在附图中的其它步骤。过程可以对应于方法、函数、过程、子例程、子程序等等。当过程对应于函数时,其终止可以对应于该函数返回到调用函数或主函数。

[0209] 可以使用存储在计算机可读介质中或者从计算机可读介质中可用的计算机可执行指令,来实现根据上述示例的过程和方法。例如,这样的指令可以包括引起或配置通用计算机、专用计算机或处理设备执行特定功能或功能组的指令和数据。可以通过网络访问使用的计算机资源部分。计算机可执行指令可以是例如二进制、中间格式指令(例如,汇编语言)、固件、源代码等等。

[0210] 术语“计算机可读介质”包括但不限于便携式或非便携式存储设备、光存储设备、以及能够存储、包含或携带指令和/或数据的各种其它介质。计算机可读介质可以包括在其中能够存储数据的非临时性介质,并且不包括无线地或通过有线连接传播的载波和/或临时性电子信号。非临时性介质的例子可以包括但不限于磁盘或磁带、诸如闪存、存储器或存储设备、磁盘或光盘之类的光学存储介质、闪存、配备有非易失性存储器的USB设备、网络存

储设备、压缩光盘 (CD) 或数字通用光盘 (DVD)、其任何适当的组合等等。计算机可读介质上可以存储有代码和/或机器可执行指令,它们可以表示过程、函数、子程序、程序、例程、子例程、模块、软件包、类、或者指令、数据结构或程序语句的任何组合。可以通过传递和/或接收信息、数据、自变量、参数或存储器内容,将代码段耦合到另一个代码段或硬件电路。可以通过包括存储器共享、消息传递、令牌传递、网络传输等等的任何适当方式,来传递、转发或传输信息、自变量、参数、数据等等。

[0211] 在一些实施例中,计算机可读存储设备、介质和存储器可以包括包含比特流等等的电缆或无线信号。但是,当提及时,非临时性计算机可读存储介质明确地排除诸如能量、载波信号、电磁波和信号本身之类的介质。

[0212] 实现根据这些公开内容的过程和方法的设备可以包括硬件、软件、固件、中间件、微代码、硬件描述语言或者其任意组合,并且可以采用多种形式中的任何一种。当以软件、固件、中间件或微代码来实现时,用于执行必要任务的程序代码或代码段(例如,计算机程序产品)可以存储在计算机可读介质或机器可读介质中。处理器可以执行必要的任务。外形尺寸的典型示例包括膝上型计算机、智能手机、移动电话、平板设备或者其它小型外形的个人计算机、个人数字助理、机架安装设备、独立设备等等。本文所描述的功能还可以体现在外围设备或附加卡中。作为进一步的示例,还可以在单个设备中执行的不同芯片或不同过程之间,在电路板上实现这种功能。

[0213] 指令、用于传送这样的指令的介质、用于执行它们的计算资源、以及用于支持这样的计算资源的其它结构,是用于提供本公开内容中描述的功能的示例性装置。

[0214] 在前面的描述中,参考本申请的特定实施例描述了本申请的各方面,但是本领域普通技术人员应当认识到,本申请并不受此限制。因此,虽然本文已经详细描述了本申请的说明性实施例,但是应当理解,可以以其它方式不同地体现和采用本发明的构思,并且所附权利要求旨在被解释为包括这样的变型,除非受到现有技术的限制。可以单独或联合地使用上述申请的各种特征和方面。此外,在不脱离本说明书的保护范围的情况下,可以在本文所述之外的任何数量的环境 and 应用中使用实施例。因此,说明书和附图应被认为是说明性的而不是限制性的。为了说明起见,以特定顺序描述了方法。应当理解,在替代实施例中,可以以与所描述的顺序不同的顺序来执行所述方法。

[0215] 本领域普通技术人员应当理解,可以将本文中使用的“<”和“>”符号或术语分别替换为小于或等于(“≤”)和大于或等于(“≥”)符号,而不脱离本说明书的范围。

[0216] 在将组件描述为“被配置为”执行某些操作的情况下,可以例如通过以下方式来实现这种配置:设计电子电路或其它硬件以执行该操作,通过对可编程电子电路(例如,微处理器或其它适当的电子电路)进行编程以执行该操作,或其任意组合。

[0217] 短语“耦合到”指代直接或间接地物理连接到另一个组件的任何组件、和/或直接或间接地与另一个组件通信的任何组件(例如,通过有线或无线连接来连接到另一个组件和/或其它适当的通信接口)。

[0218] 引用一个集合中的“至少一个”和/或一个集合中的“一个或多个”的声明语言或其它语言,指示该集合中的一个成员或该组中的多个成员(具有任何组合)满足权利要求。例如,引用“A和B中的至少一个”或者“A或B中的至少一个”的声明语言意味着A、B、或A和B。在

另一个例子中,引用“A、B和C中的至少一个”或“A、B或C中的至少一个”的声明语言意味着A、B、C、或A和B、或A和C、或B和C、或A和B和C。一个集合中的“至少一个”和/或一个集合中的“一个或多个”的语言,并不将该集合限制于该集合中列出的项。例如,引用“A和B中的至少一个”或者“A或B中的至少一个”的声明语言可以意味着A、B、或A和B,并且可以另外地包括A和B的集合中没有列出的项。

[0219] 结合本文所公开实施例描述的各种示例性的逻辑框、模块、电路和算法步骤均可以实现成电子硬件、计算机软件、固件或者其组合。为了清楚地表示硬件和软件之间的这种可交换性,上面对各种示例性的部件、框、模块、电路和步骤均围绕其功能进行了总体描述。至于这种功能是实现成硬件还是实现成软件,取决于特定的应用和对整个系统所施加的设计约束条件。熟练的技术人员可以针对每个特定应用,以变通的方式实现所描述的功能,但是,这种实现决策不应解释为背离本公开内容的保护范围。

[0220] 本文所描述的技术还可以利用电子硬件、计算机软件、固件或者其任何组合来实现。可以利用诸如通用计算机、无线通信设备手持装置或者具有多种用途(其包括在无线通信设备手持装置和其它设备中的应用)的集成电路设备等等之类的各种设备中的任何一种来实现这种技术。描述成模块或组件的任何特征可以在集成逻辑设备中一起实现,也可以分别作为离散但可互操作的逻辑设备来实现。如果以软件实施,则这些技术可以至少部分地由计算机可读数据存储介质来实现,所述计算机可读数据存储介质包括程序代码,所述程序代码包括在被执行时执行上述方法、算法和/或操作中的一个或多个的指令。计算机可读数据存储介质可以形成计算机程序产品的一部分,该计算机程序产品可以包括封装材料。计算机可读介质可以包括存储器或数据存储介质,例如,诸如同步动态随机存取存储器(SDRAM)之类的随机存取存储器(RAM)、只读存储器(ROM)、非易失性随机存取存储器(NVRAM)、电可擦除可编程只读存储器(EEPROM)、闪存、磁或光数据存储介质等等。另外地或替代地,可以至少部分地通过计算机可读通信介质来实现这些技术,该计算机可读通信介质以指令或数据结构的形式(例如,传播的信号或波)携带或传递程序代码,并且程序代码可以由计算机访问、读取和/或执行。

[0221] 程序代码可以由处理器执行,该处理器可以包括一个或多个处理器,例如一个或多个数字信号处理器(DSP)、通用微处理器、专用集成电路(ASIC)、现场可编程逻辑阵列(FPGA)或者其它等效的集成逻辑电路或离散逻辑电路。这种处理器可以被配置为执行本公开内容中描述的任何技术。通用处理器可以是微处理器,但是,该处理器也可以是任何常规处理器、控制器、微控制器或状态机。处理器也可以实现为计算设备的组合,例如,DSP和微处理器的组合、若干微处理器、与DSP核心结合的一个或多个微处理器、或者任何其它这样的配置。因此,本文所使用的术语“处理器”可以指代任何的前述结构、前述结构的任何组合、或者适合于实施本文所描述的技术的任何其它结构或装置。

[0222] 本公开内容的说明性示例包括:

[0223] 方面1:一种用于处理视频数据的系统,所述系统包括:双向译码引擎,其包括插值引擎和单向译码引擎,其中:所述插值引擎被配置为获得第一参考帧和第二参考帧,并至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成第三参考帧;所述单向译码引擎被配置为基于所述第三参考帧,对输入帧执行单向帧间预测。

[0224] 方面2:根据方面1所述的系统,其中,所述单向译码引擎包括运动引擎和变形引

擎,并且其中:所述运动引擎被配置为估计所述输入帧和所述第三参考帧之间的运动;所述变形引擎被配置为至少部分地通过基于所估计的运动来变形所述第三参考帧的一个或多个像素,以生成变形帧。

[0225] 方面3:根据方面2所述的系统,其中,所述单向译码引擎还包括残差引擎,并且其中:所述残差引擎被配置为:至少部分地通过确定所述输入帧和所述变形帧之间的差异来确定残差;并使用所述残差来生成预测残差。

[0226] 方面4:根据方面3所述的系统,还包括存储器和耦合到所述存储器的至少一个处理器,其中所述单向译码引擎被配置为:基于所述变形帧和所述预测残差,生成表示所述输入帧的重构帧,所述重构帧包括双向预测帧;并且其中,所述至少一个处理器被配置为使所述重构帧被存储在所述存储器中。

[0227] 方面5:根据方面4所述的系统,其中,所述至少一个处理器包括神经处理单元(NPU)。

[0228] 方面6:根据方面3至5中的任何一项所述的系统,其中,所述运动引擎包括第一神经网络,并且所述残差引擎包括第二神经网络。

[0229] 方面7:根据方面6所述的系统,其中,所述第一神经网络包括第一自动编码器,并且所述第二神经网络包括第二自动编码器。

[0230] 方面8:根据方面1至7中的任何一项所述的系统,其中,所述单向译码引擎被配置为:获得第四参考帧;估计另外输入帧和所述第四参考帧之间的运动;至少部分地通过基于所述估计的运动来变形所述第四参考帧的一个或多个像素,生成另外的变形帧;并基于所述另外的变形帧和另外的预测残差,生成表示所述另外输入帧的重构帧,表示所述另外输入帧的所述重构帧包括单向预测帧。

[0231] 方面9:根据方面1至8中的任何一项所述的系统,其中,为了至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成所述第三参考帧,所述插值引擎被配置为:确定表示从所述第一参考帧到所述第三参考帧的像素运动的第一组运动信息;确定表示从所述第二参考帧到所述第三参考帧的像素运动的第二组运动信息;至少部分地通过使用所述第一组运动信息对所述第一参考帧执行变形函数来生成第一变形信息;至少部分地通过使用所述第二组运动信息对所述第二参考帧执行所述变形函数来生成第二变形信息;并基于所述第一变形信息和所述第二变形信息生成所述第三参考帧。

[0232] 方面10:根据方面9所述的系统,其中,基于所述第一参考帧和所述第二参考帧之间的像素运动来确定所述第一组运动信息和所述第二组运动信息。

[0233] 方面11:根据方面9或10中的任何一项所述的系统,其中,所述第一组运动信息包括第一光流图,并且所述第二组运动信息包括第二光流图。

[0234] 方面12:根据方面9至11中的任何一项所述的系统,其中,所述变形函数包括双线性插值函数。

[0235] 方面13:根据方面2至12中的任何一项所述的系统,其中,所述输入帧与所述第三参考帧之间的所述估计的运动包括光流数据。

[0236] 方面14:根据方面2至12中的任何一项所述的系统,其中,所述输入帧与所述第三参考帧之间的所述估计的运动包括动态卷积数据。

[0237] 方面15:根据方面2至12中的任何一项所述的系统,其中,所述输入帧与所述第三

参考帧之间的所述估计的运动包括基于块的运动数据。

[0238] 方面16:根据方面1至15中的任何一项所述的系统,其中,所述系统是移动设备、扩展现实设备或电视之一。

[0239] 方面17:根据方面1至16中的任何一项所述的系统,还包括被配置为捕获一个或多个视频帧的显示器和相机中的至少一个。

[0240] 方面18:一种用于处理视频数据的方法,所述方法包括:获得第一参考帧和第二参考帧;至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成第三参考帧;并基于所述第三参考帧对输入帧执行单向帧间预测。

[0241] 方面19:根据方面18所述的方法,还包括:估计所述输入帧和所述第三参考帧之间的运动;并至少部分地通过基于所估计的运动,变形所述第三参考帧的一个或多个像素,来生成变形帧。

[0242] 方面20:根据方面19所述的方法,还包括:至少部分地通过确定所述输入帧和所述变形帧之间的差异来确定残差;并使用所述残差生成预测残差。

[0243] 方面21:根据方面20所述的方法,还包括:基于所述变形帧和所述预测残差,生成表示所述输入帧的重构帧,所述重构帧包括双向预测帧;并使所述重构帧存储在存储器中。

[0244] 方面22:根据方面20或21中的任何一项所述的方法,其中,所述运动是使用第一自动编码器估计的,并且其中所述预测残差是使用第二自动编码器生成的。

[0245] 方面23:根据方面18至22中的任何一项所述的方法,还包括:获得第四参考帧;估计另外输入帧和所述第四参考帧之间的运动;至少部分地通过基于所估计的运动来变形所述第四参考帧的一个或多个像素,生成另外的变形帧;并基于所述另外的变形帧和另外的预测残差,生成表示所述另外输入帧的重构帧,表示所述另外输入帧的所述重构帧包括单向预测帧。

[0246] 方面24:根据方面18至23中的任何一项所述的方法,还包括:确定表示从所述第一参考帧到所述第三参考帧的像素运动的第一组运动信息;确定表示从所述第二参考帧到所述第三参考帧的像素运动的第二组运动信息;至少部分地通过使用所述第一组运动信息对所述第一参考帧执行变形函数来生成第一变形信息;至少部分地通过使用所述第二组运动信息对所述第二参考帧执行所述变形函数来生成第二变形信息;并基于所述第一变形信息和所述第二变形信息生成所述第三参考帧。

[0247] 方面25:根据方面24所述的方法,其中,基于所述第一参考帧和所述第二参考帧之间的像素运动来确定所述第一组运动信息和所述第二组运动信息。

[0248] 方面26:根据方面24或25中的任何一项所述的方法,其中,所述第一组运动信息包括第一光流图,并且所述第二组运动信息包括第二光流图。

[0249] 方面27:根据方面24至26中的任何一项所述的方法,其中,所述变形函数包括双线性插值函数。

[0250] 方面28:根据方面19至27中的任何一项所述的方法,其中,所述输入帧与所述第三参考帧之间的所述估计的运动包括光流数据。

[0251] 方面29:根据方面19至27中的任何一项所述的方法,其中,所述输入帧与所述第三参考帧之间的所述估计的运动包括动态卷积数据。

[0252] 方面30:根据方面19至27中的任何一项所述的方法,其中,所述输入帧与所述第三

参考帧之间的所述估计的运动包括基于块的运动数据。

[0253] 方面31:一种其上存储有指令的非临时性计算机可读介质,当所述指令被一个或多个处理器执行时,使所述一个或多个处理器执行以下操作:获得第一参考帧和第二参考帧;至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成第三参考帧;并基于所述第三参考帧对输入帧执行单向帧间预测。

[0254] 方面32:根据方面31所述的非临时性计算机可读介质,还包括当被所述一个或多个处理器执行时,使得所述一个或多个处理器执行方面19至30中的任何一项所述的操作的指令。

[0255] 方面33:一种用于处理视频数据的装置,包括:用于获得第一参考帧和第二参考帧的单元;用于至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成第三参考帧的单元;用于基于所述第三参考帧对输入帧执行单向帧间预测的单元。

[0256] 方面34:根据方面33所述的装置,还包括用于执行方面19至30中的任何一项所述的操作的单元。

[0257] 方面35:一种用于处理视频数据的系统,所述系统包括:存储器;以及耦合到所述存储器的至少一个处理器,所述至少一个处理器被配置为:获得内插参考帧和表示输入帧与所述内插参考帧之间运动的运动信息;基于所述运动信息,估计所述输入帧和所述内插参考帧之间的运动;至少部分地通过基于所述估计的运动,对所述内插参考帧的一个或多个像素进行变形来生成变形帧;并基于所述变形帧和预测残差来生成表示所述输入帧的重构帧,所述重构帧包括双向预测帧。

[0258] 方面36:根据方面35所述的系统,其中,所述至少一个处理器被配置为:获得第一参考帧和第二参考帧;并至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成所述内插参考帧。

[0259] 方面37:根据方面35所述的系统,其中,所述至少一个处理器被配置为:从比特流中获取残差;并基于所获得的残差来生成预测残差。

[0260] 方面38:根据方面37所述的系统,其中,所述运动信息是从所述比特流中获得的。

[0261] 方面39:根据方面35所述的系统,其中,所述至少一个处理器被配置为:获得第四参考帧;估计另外输入帧和所述第四参考帧之间的运动;至少部分地通过基于所估计的运动,变形所述第四参考帧的一个或多个像素来生成另外的变形帧;并基于所述另外的变形帧和另外的预测残差,生成表示所述另外输入帧的重构帧,表示所述另外输入帧的所述重构帧包括单向预测帧。

[0262] 方面40:一种用于处理视频数据的方法,所述方法包括:获得内插参考帧和表示输入帧与所述内插参考帧之间运动的运动信息;基于所述运动信息,估计所述输入帧和所述内插参考帧之间的运动;至少部分地通过基于所估计的运动,对所述内插参考帧的一个或多个像素进行变形来生成变形帧;并基于所述变形帧和预测残差来生成表示所述输入帧的重构帧,所述重构帧包括双向预测帧。

[0263] 方面41:根据方面40所述的方法,还包括:获得第一参考帧和第二参考帧;并至少部分地通过在所述第一参考帧和所述第二参考帧之间执行插值来生成所述内插参考帧。

[0264] 方面42:根据方面40所述的方法,还包括:从比特流中获取残差;并基于所获得的残差来生成预测残差。

[0265] 方面43:根据方面42所述的方法,其中,所述运动信息是从所述比特流中获得的。

[0266] 方面44:根据方面40所述的方法,还包括:获得第四参考帧;估计另外输入帧和所述第四参考帧之间的运动;至少部分地通过基于所估计的运动,变形所述第四参考帧的一个或多个像素来生成另外的变形帧;并基于所述另外的变形帧和另外的预测残差,生成表示所述另外输入帧的重构帧,表示所述另外输入帧的所述重构帧包括单向预测帧。

[0267] 方面45:一种其上存储有指令的非临时性计算机可读介质,当所述指令被一个或多个处理器执行时,使所述一个或多个处理器执行以下操作:获得内插参考帧和表示输入帧与所述内插参考帧之间运动的运动信息;基于所述运动信息,估计所述输入帧和所述内插参考帧之间的运动;至少部分地通过基于所述估计的运动,对所述内插参考帧的一个或多个像素进行变形来生成变形帧;并基于所述变形帧和预测残差来生成表示所述输入帧的重构帧,所述重构帧包括双向预测帧。

[0268] 方面46:根据方面45所述的非临时性计算机可读介质,还包括当被所述一个或多个处理器执行时,使得所述一个或多个处理器执行根据方面41至44中的任何一项所述的操作的指令。

[0269] 方面47:一种用于处理视频数据的装置,包括:用于获得内插参考帧和表示输入帧与所述内插参考帧之间运动的运动信息的单元;用于基于所述运动信息,估计所述输入帧和所述内插参考帧之间的运动的单元;用于至少部分地通过基于所述估计的运动,对所述内插参考帧的一个或多个像素进行变形来生成变形帧的单元;用于基于所述变形帧和预测残差来生成表示所述输入帧的重构帧的单元,所述重构帧包括双向预测帧。

[0270] 方面48:根据方面47所述的装置,还包括用于执行根据方面41至44中的任何一项所述的操作的单元。

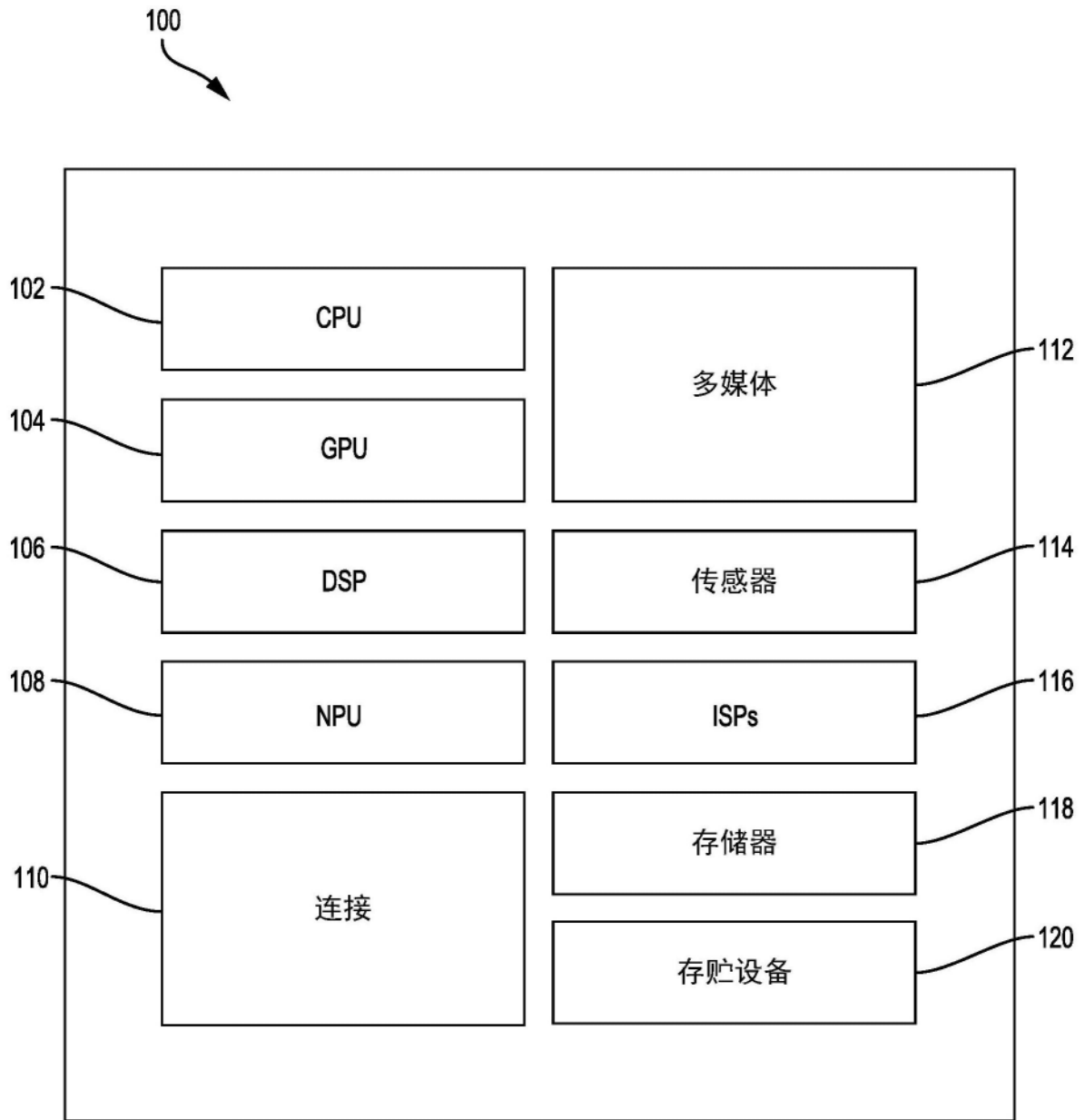


图1

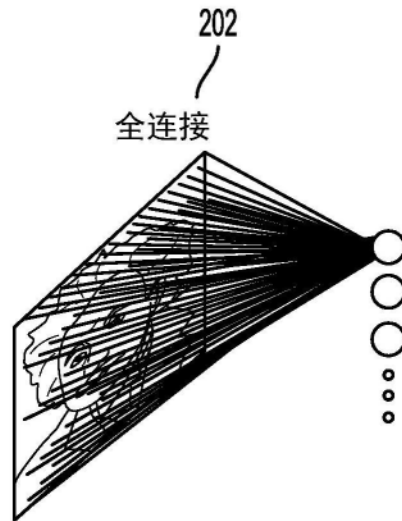


图2A

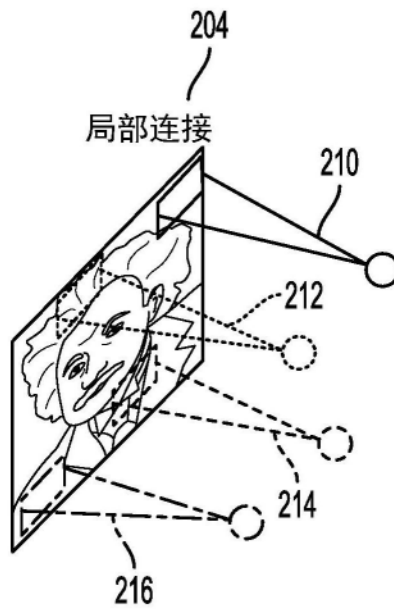


图2B

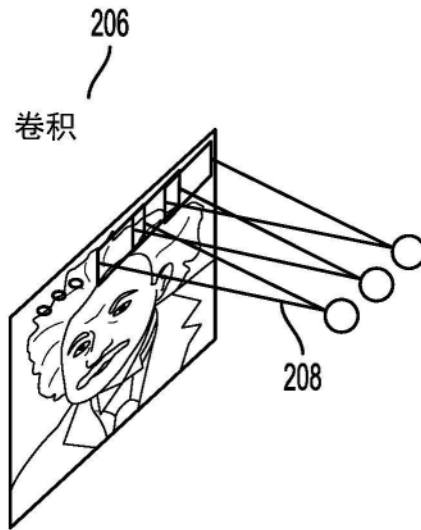


图2C

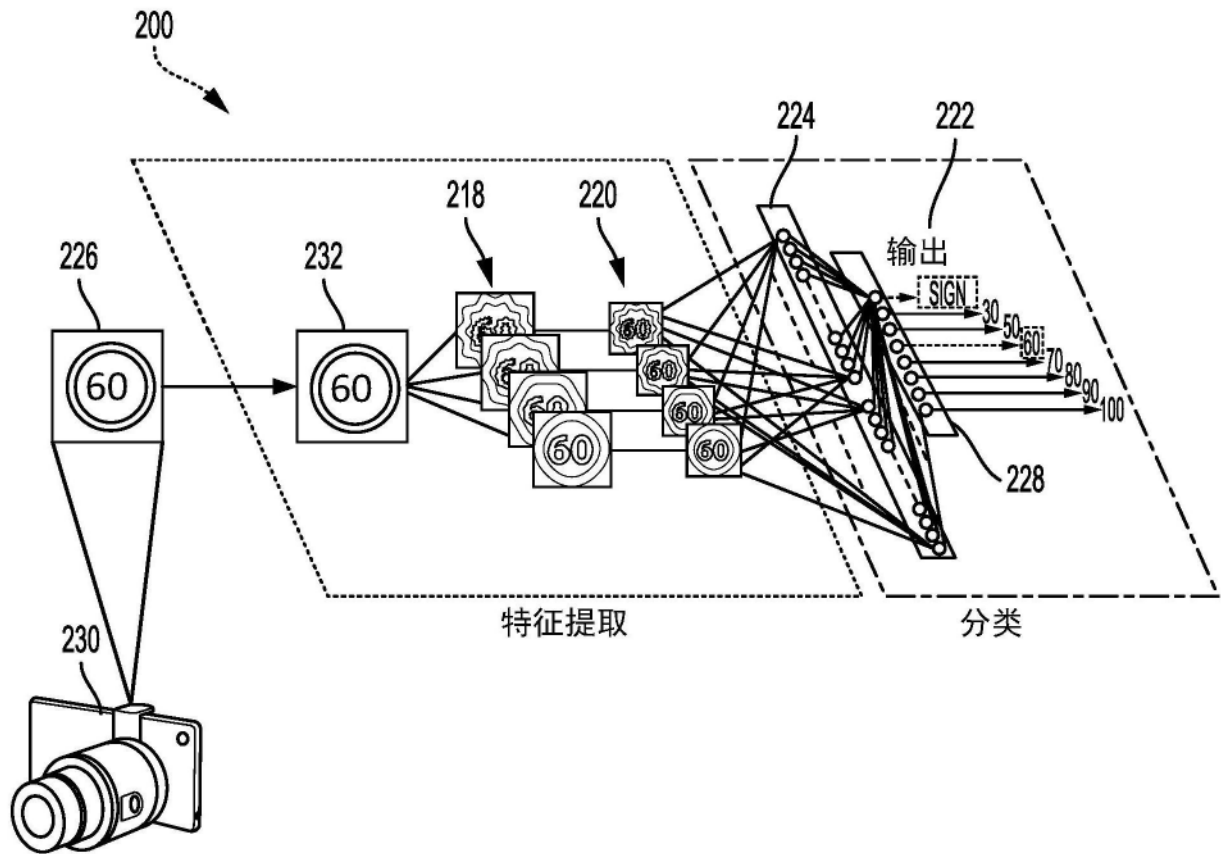


图2D

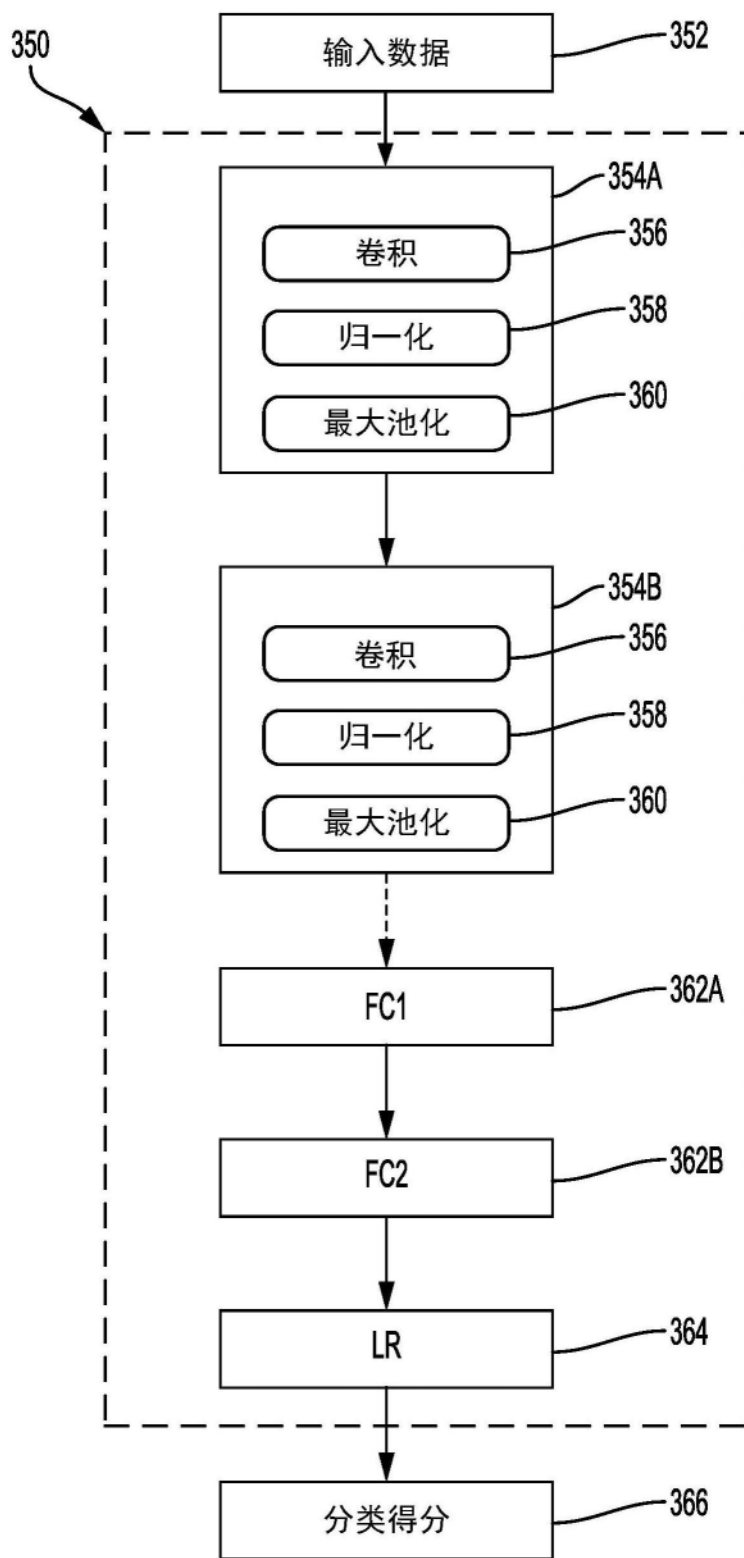


图3

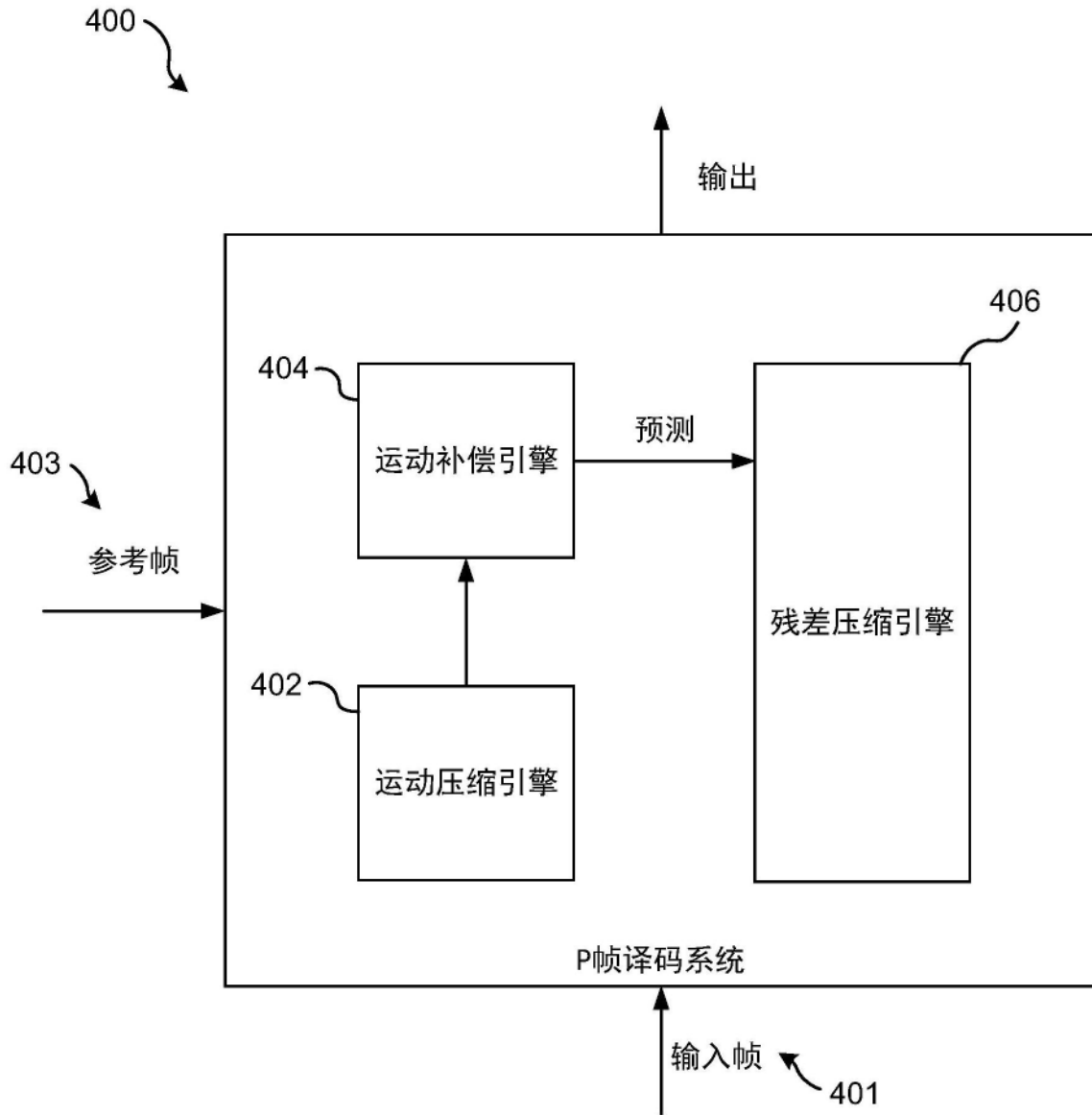


图4

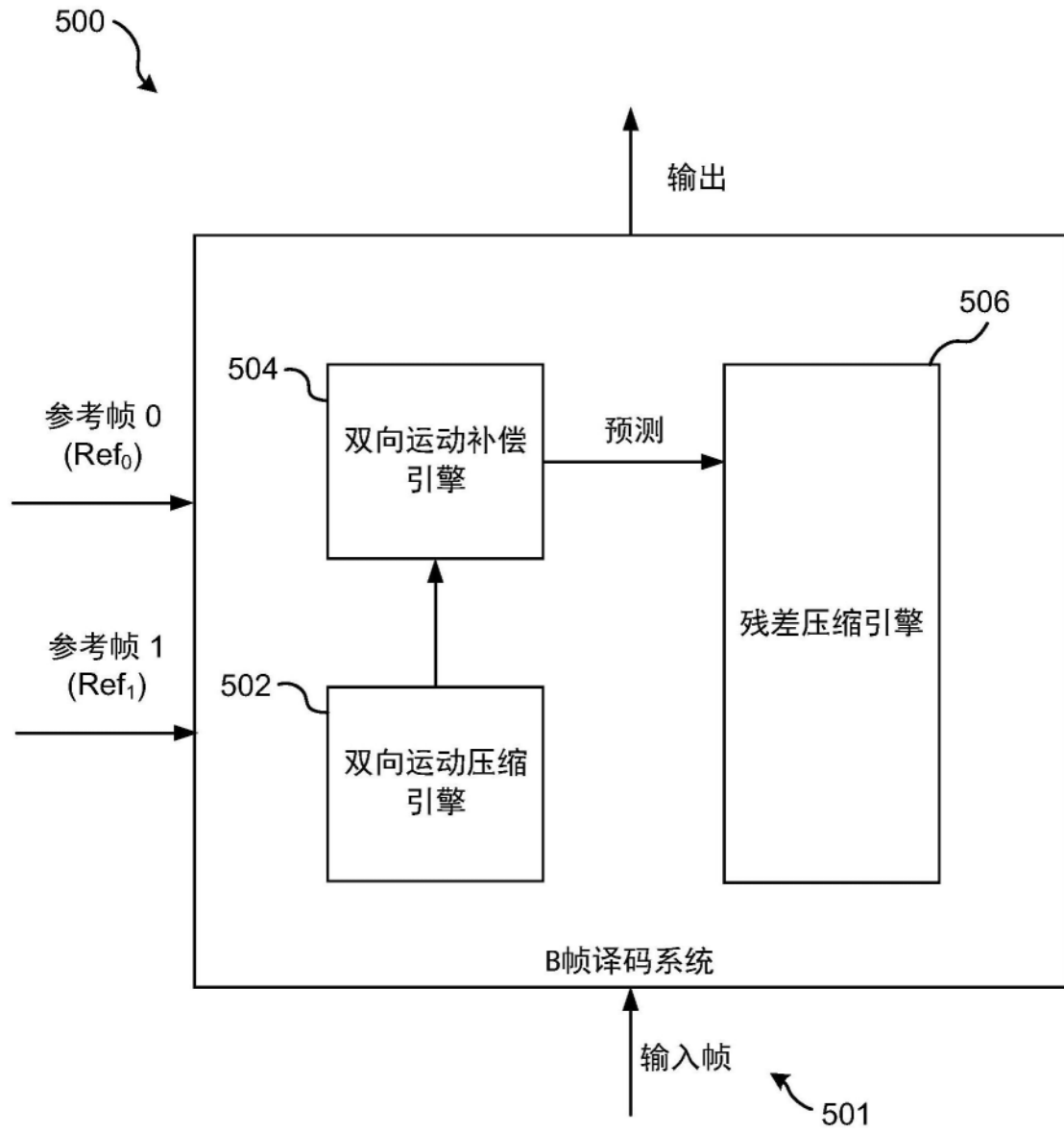


图5A

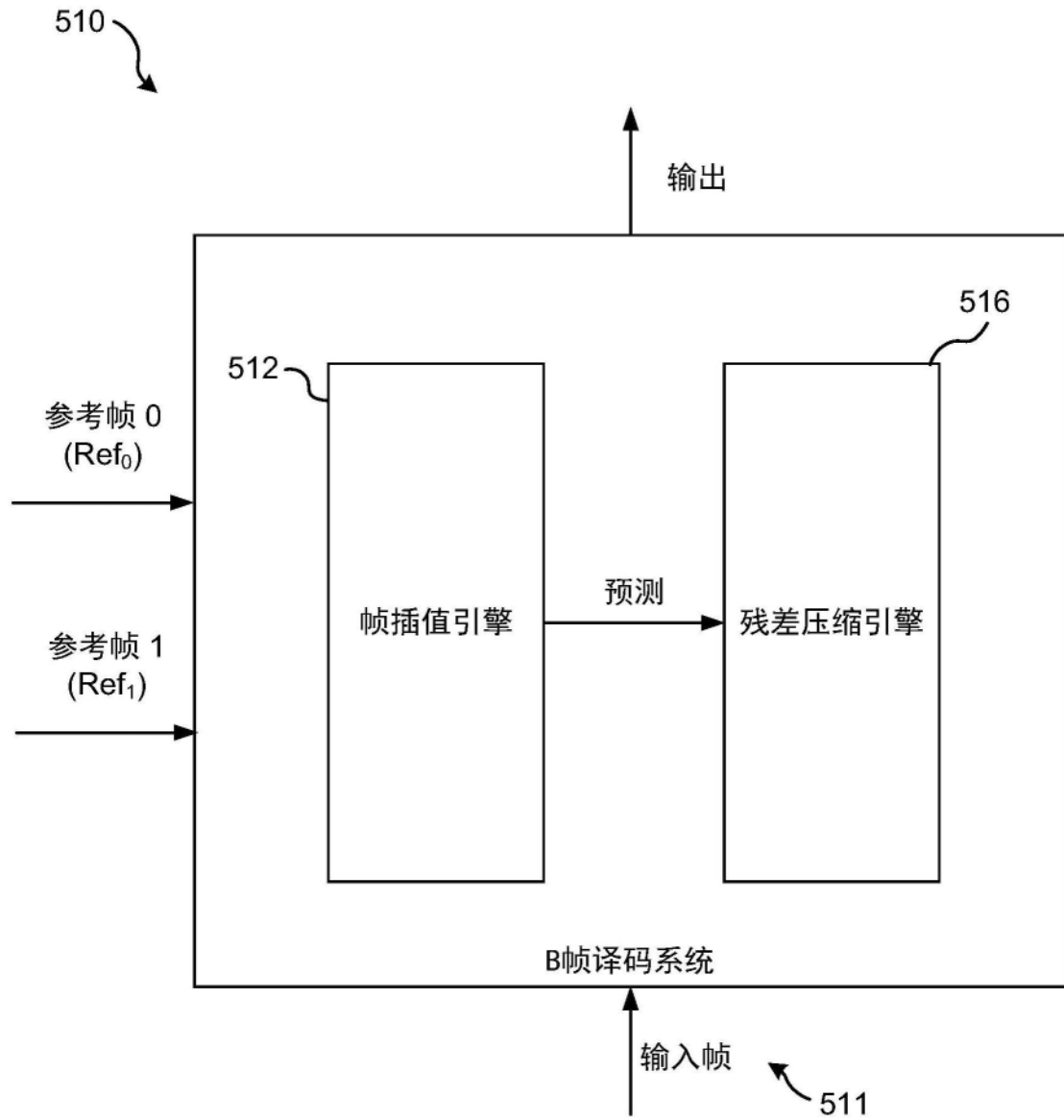


图5B

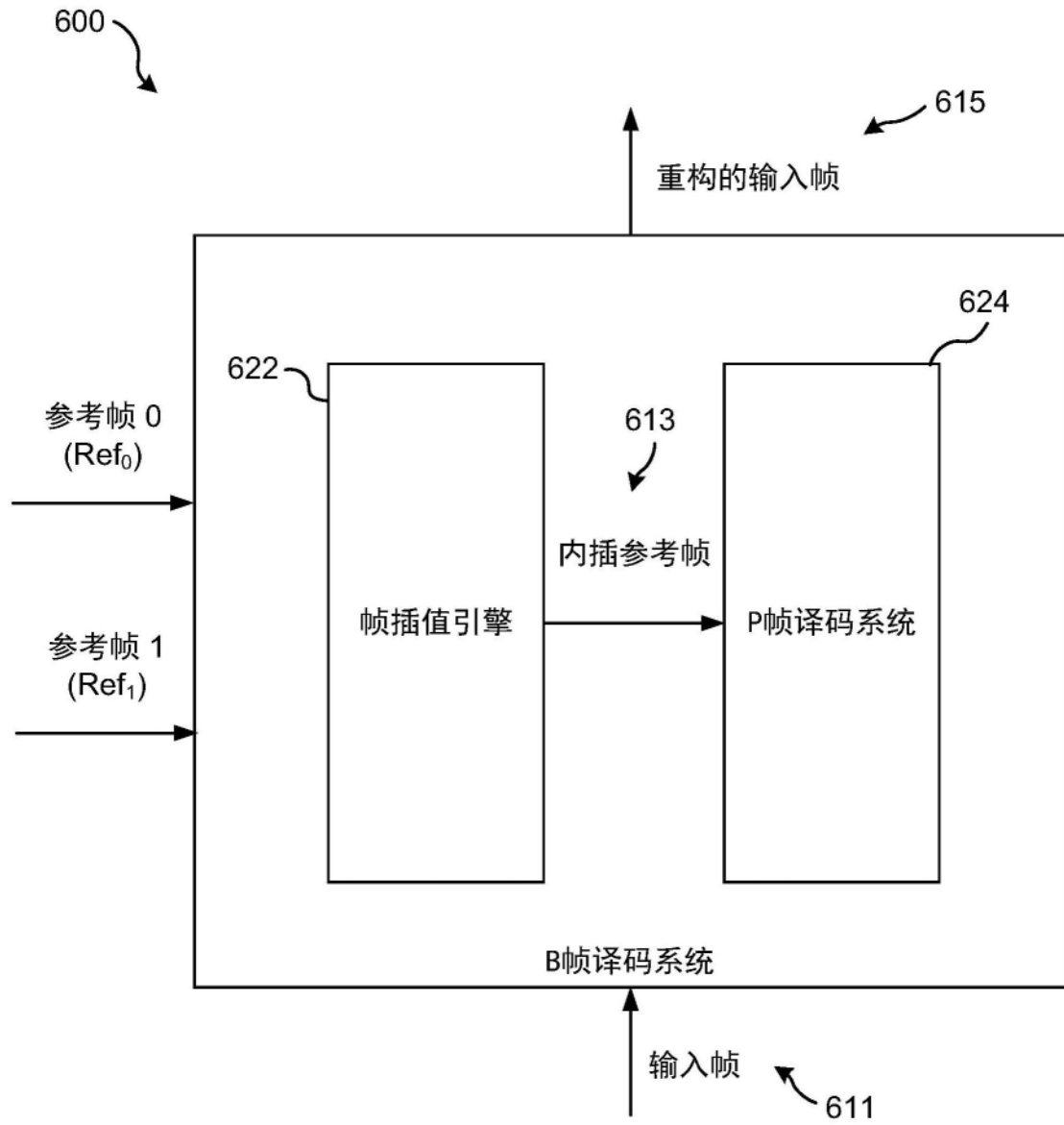


图6

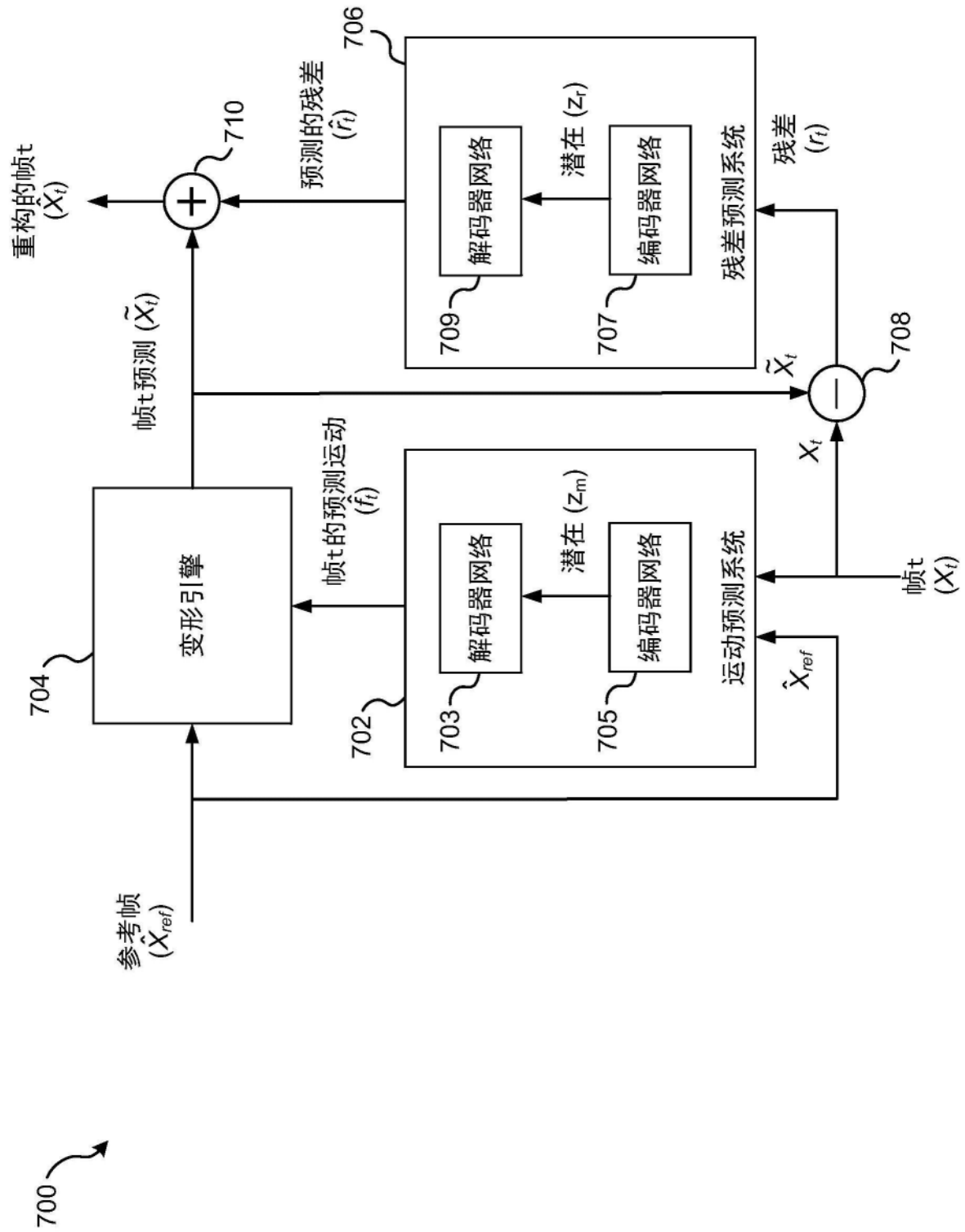


图7A

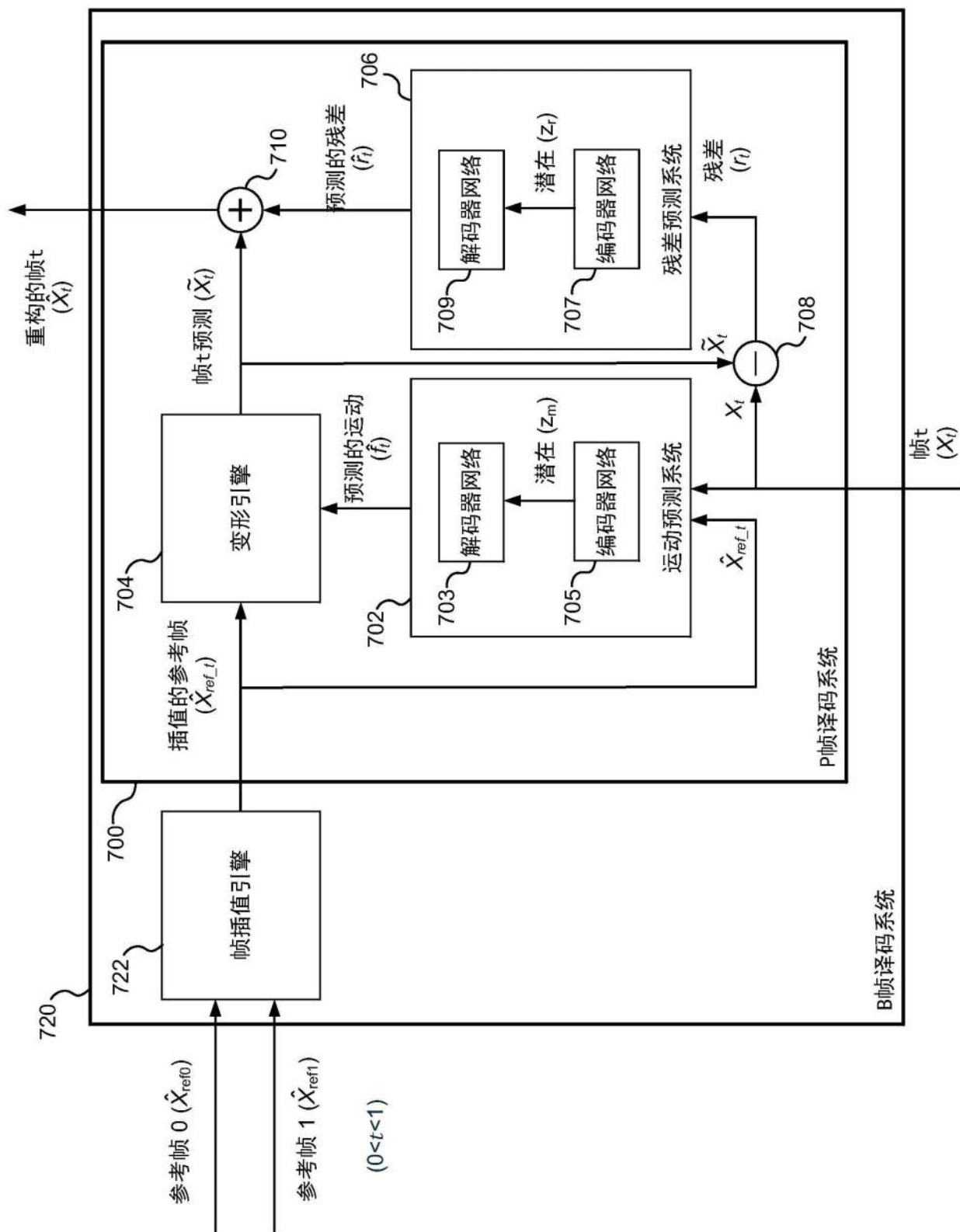


图7B

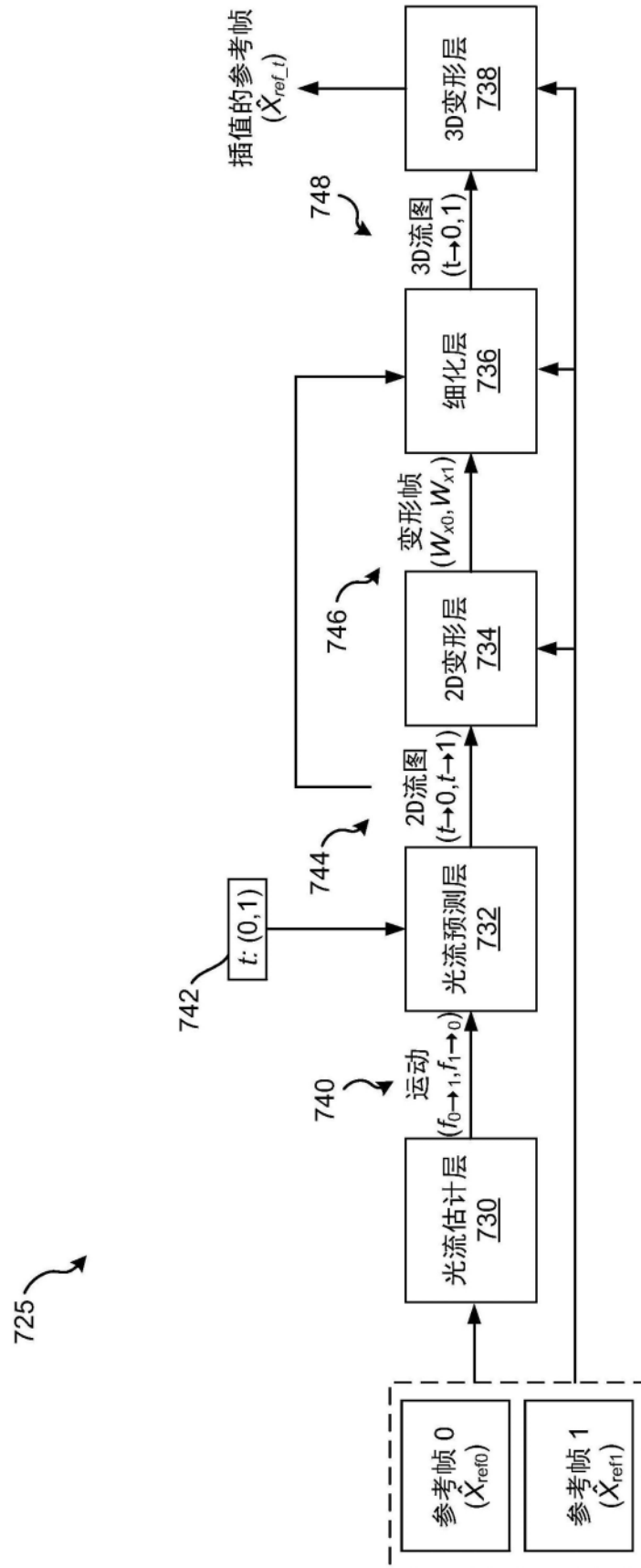


图7C

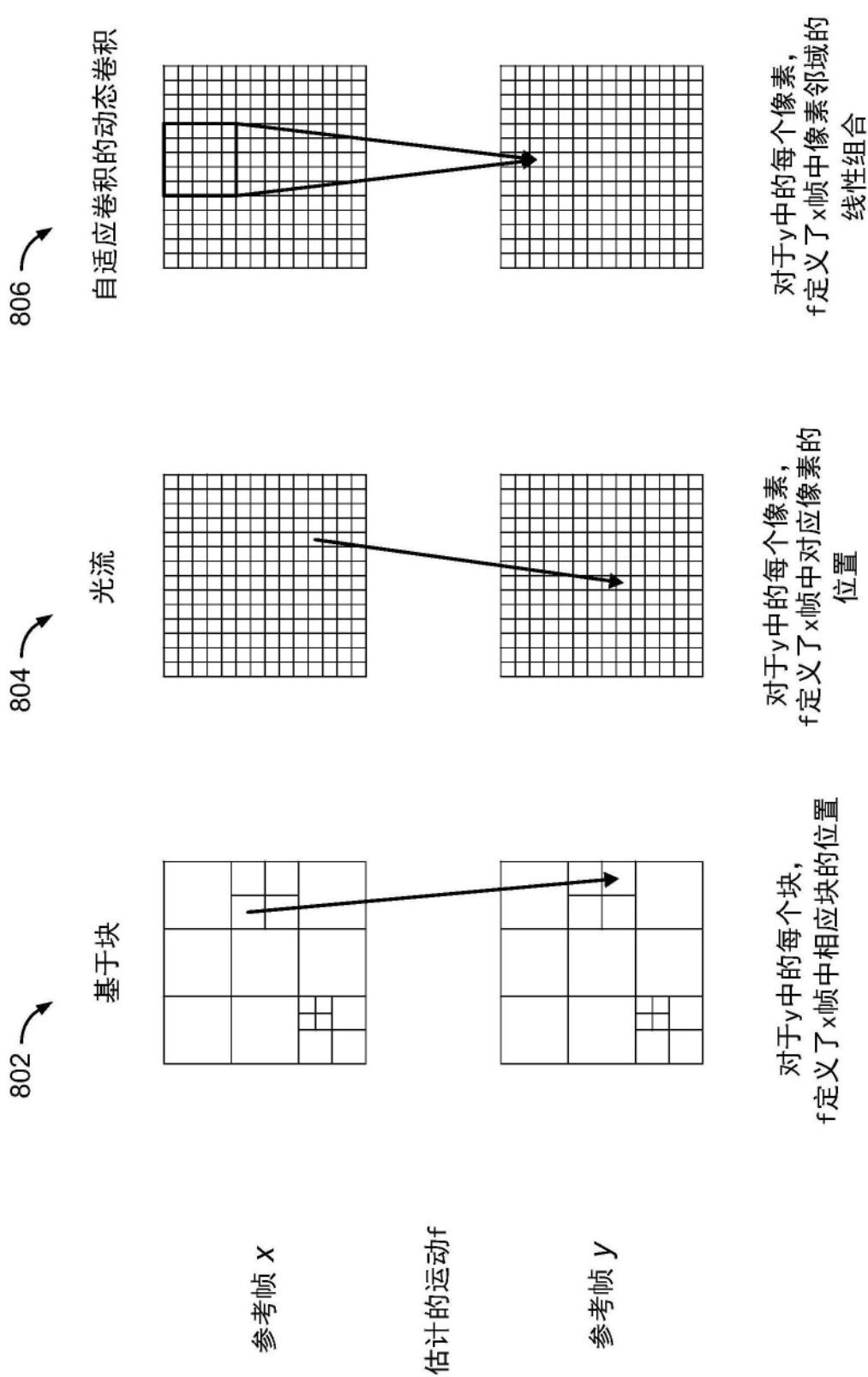


图8

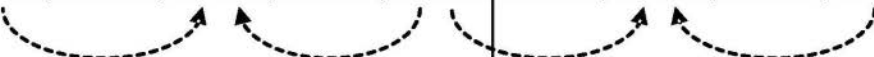
	GoP 1					GoP 2				...
类型	I	...	B	...	I	...	B	...	I	...
顺序	第一		第二		第一		第二		第一	...
参考										...

图9

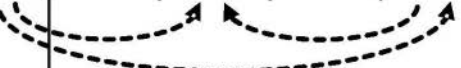
	GoP 1					GoP 2				...
类型	I	...	B	...	P	...	B	...	P	...
顺序	第一		第三		第二		第二		第一	...
参考										...

图10

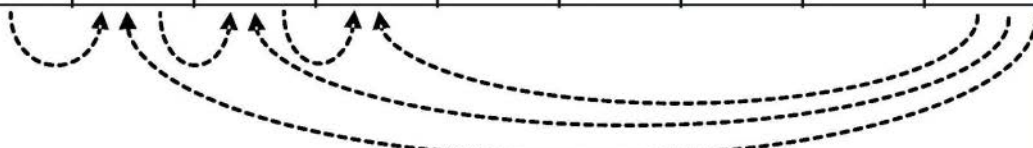
	GoP								
类型	Ref	B	B	B	...	...	...	...	Ref
顺序	1	2	3	4					1
参考									

图11

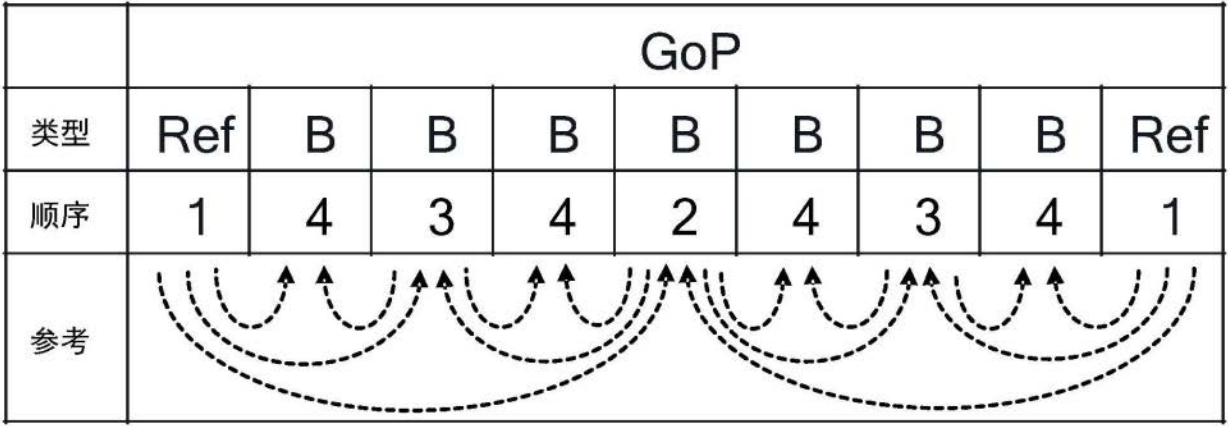


图12

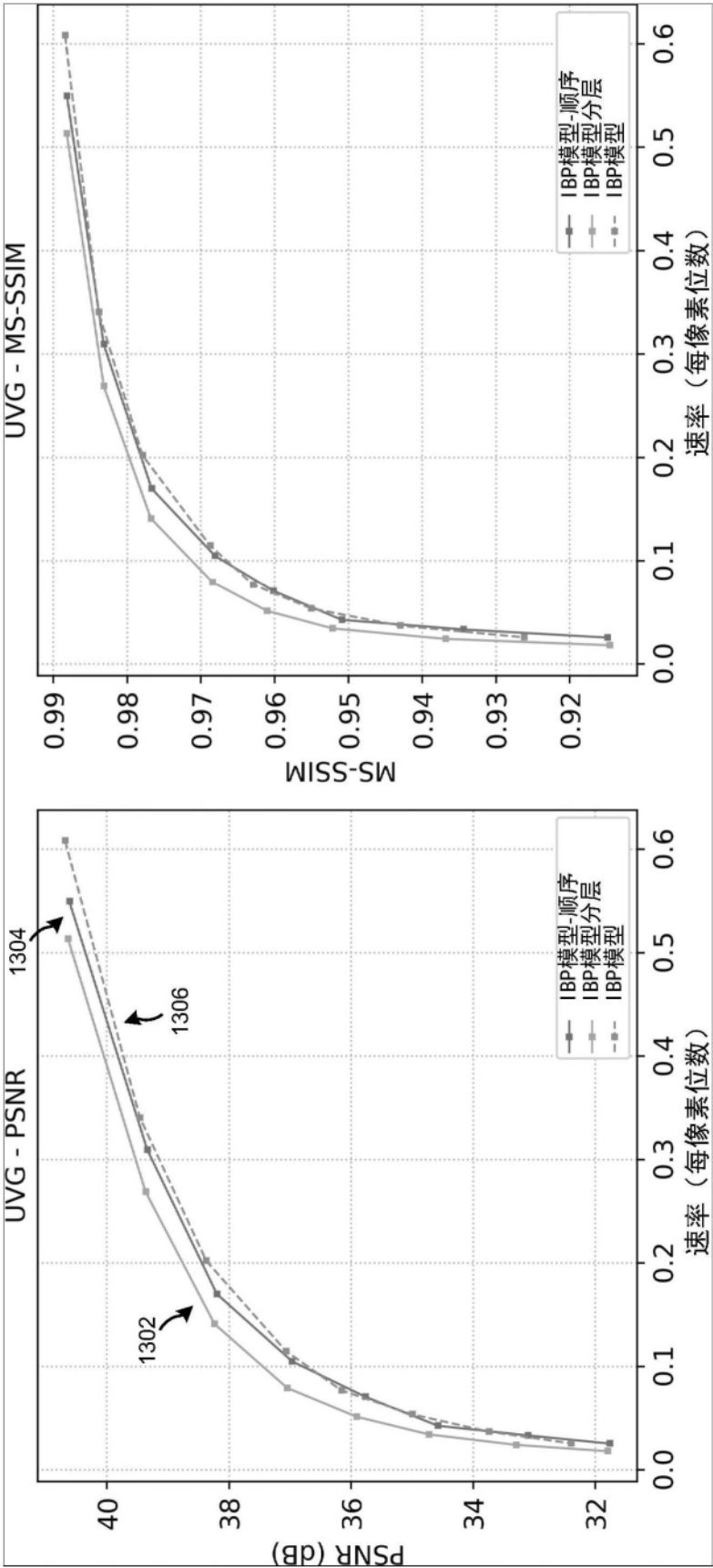


图13A

图13B

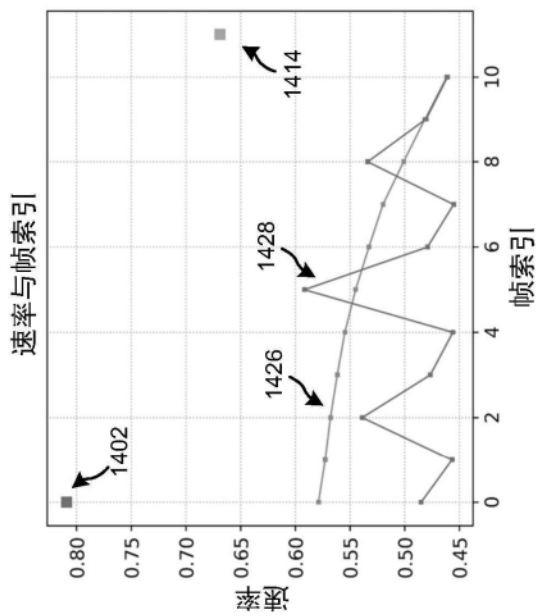


图14A

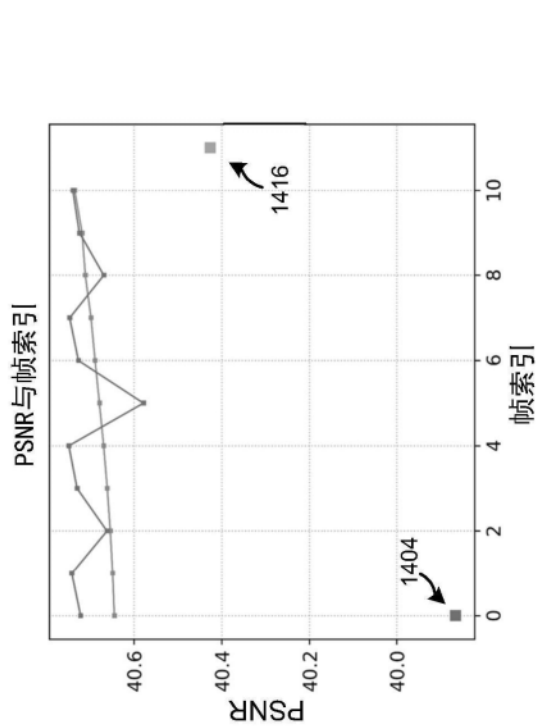


图14B

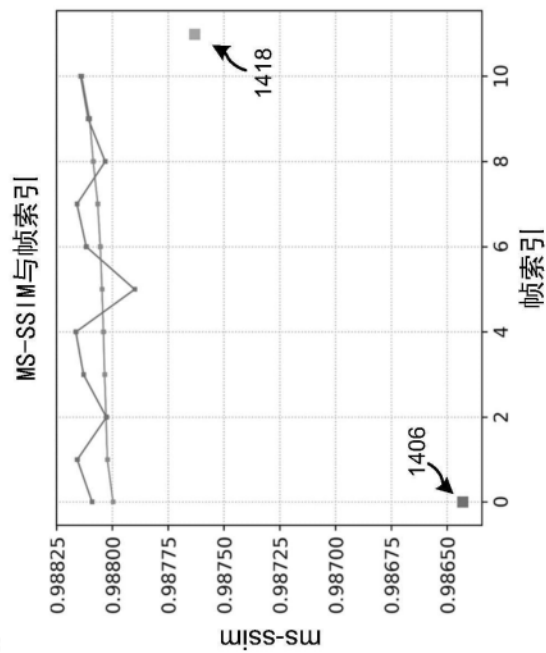
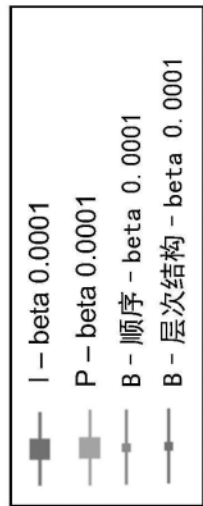


图14C



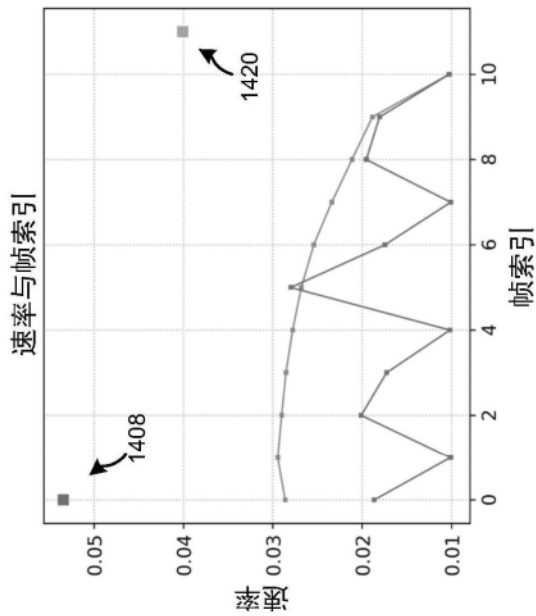


图14D

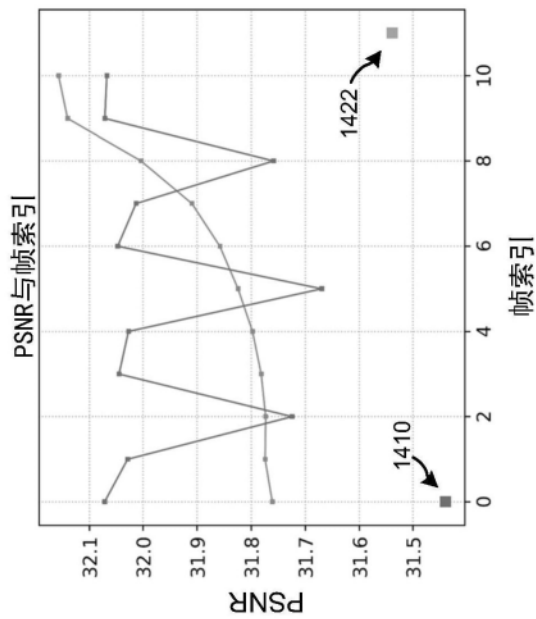


图14E

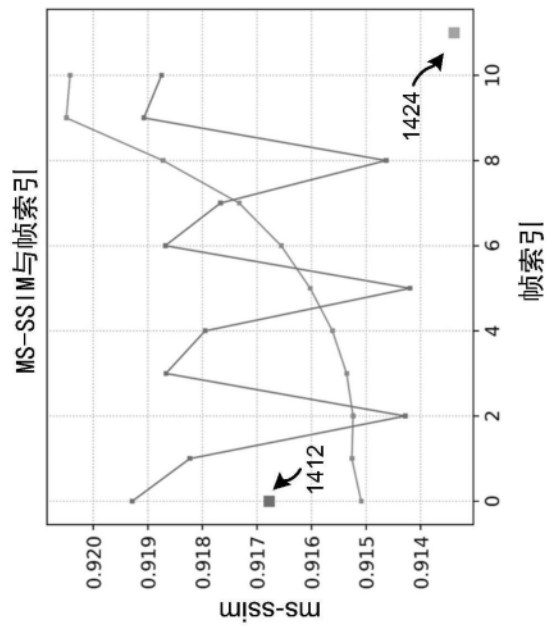


图14F

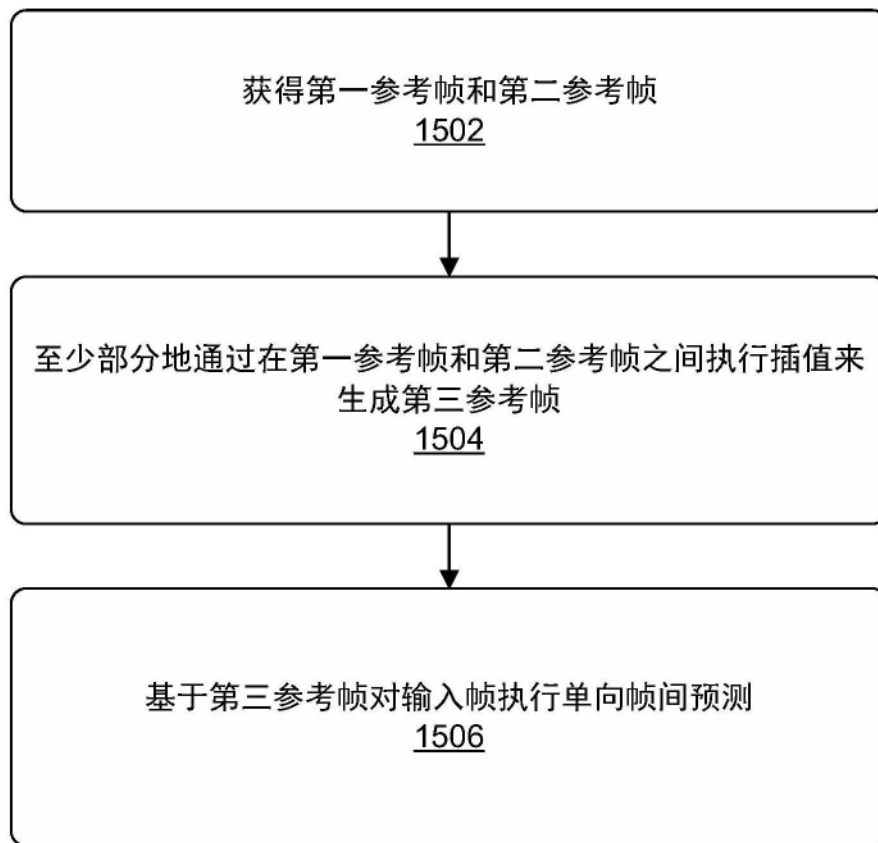
1500

图15

1600

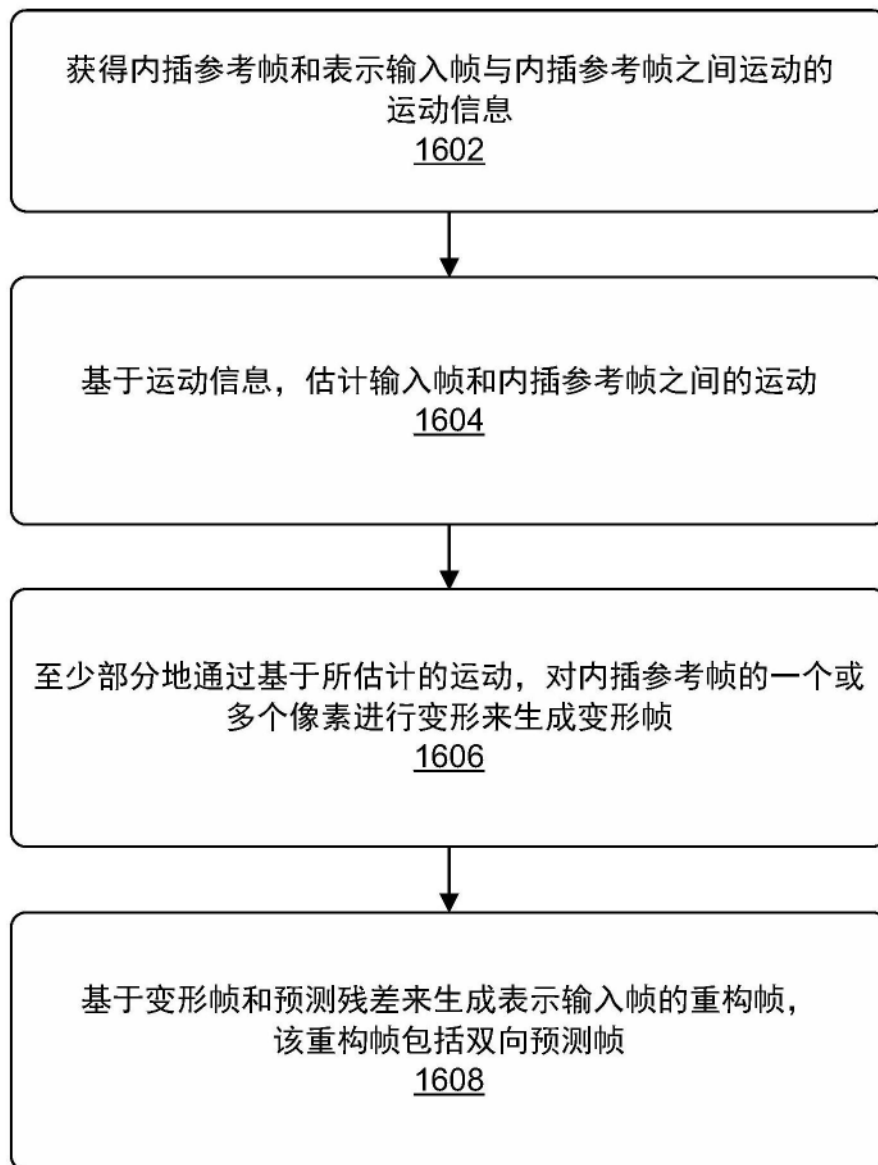


图16

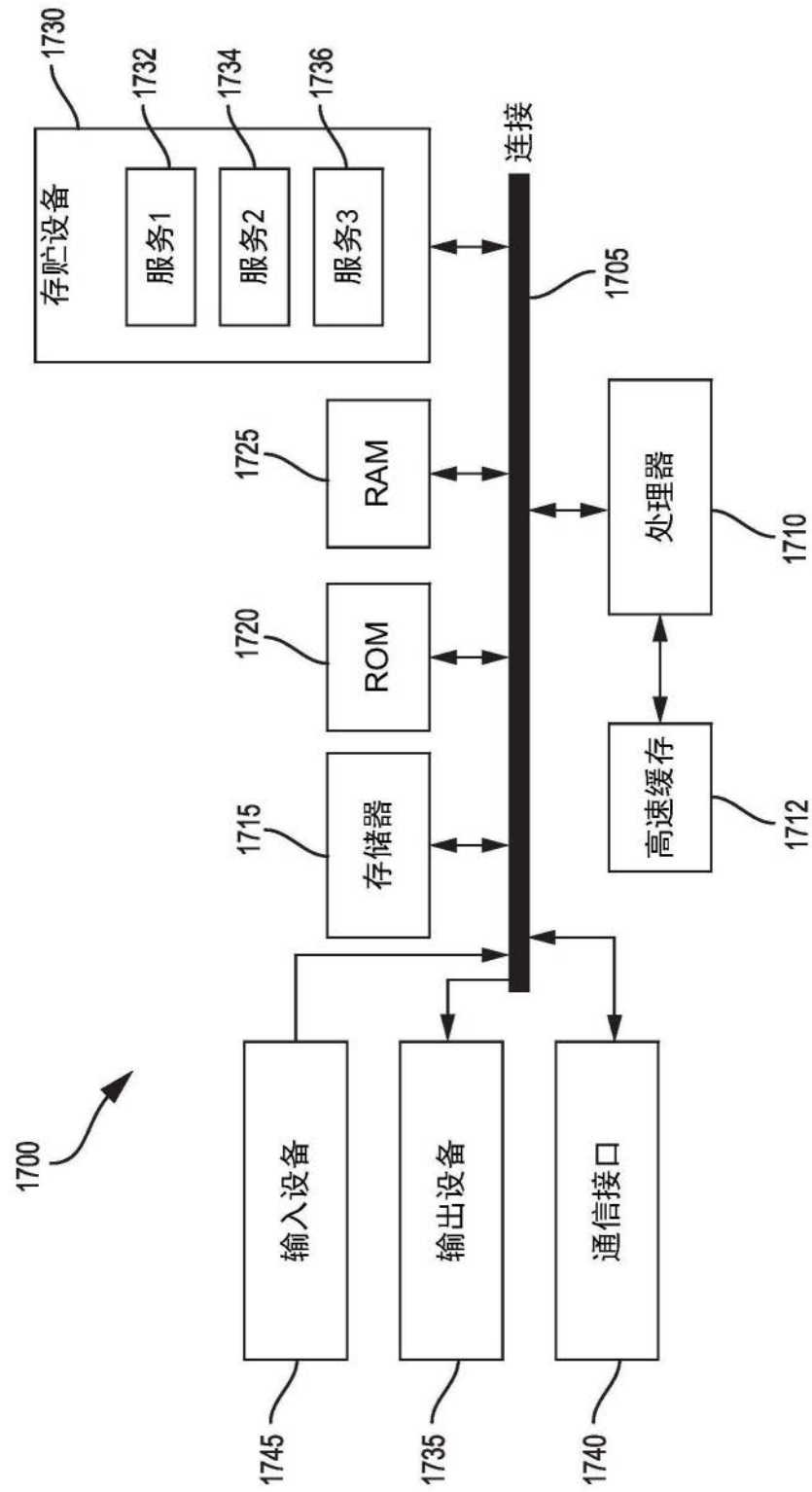


图17