



(19)中華民國智慧財產局

(12)發明說明書公開本

(11)公開編號：TW 202236849 A

(43)公開日：中華民國 111 (2022) 年 09 月 16 日

(21)申請案號：111104143

(22)申請日：中華民國 111 (2022) 年 01 月 28 日

(51)Int. Cl.：

*H04N19/103 (2014.01)**G06T1/20 (2006.01)**G06T7/20 (2017.01)**G06N3/06 (2006.01)*

(30)優先權：2021/03/11

美國

17/198,813

(71)申請人：美商高通公司 (美國) QUALCOMM INCORPORATED (US)

美國

(72)發明人：普雷扎 瑞薩 POURREZA, REZA (US)；柯恩 塔可 賽巴斯汀 COHEN, TACO SEBASTIAAN (NL)

(74)代理人：林怡芳

申請實體審查：無 申請專利範圍項數：52 項 圖式數：17 共 119 頁

(54)名稱

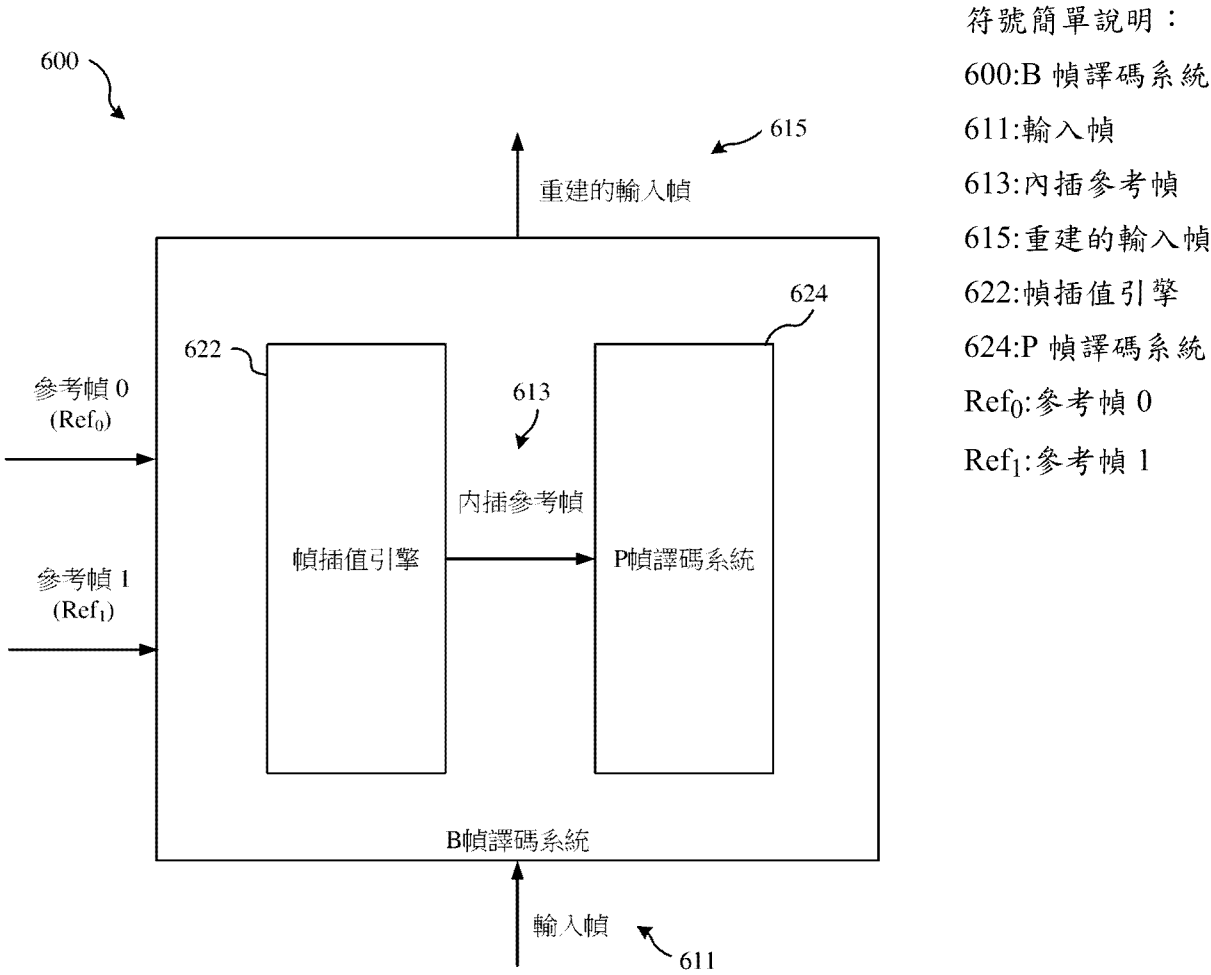
使用 P 幀譯碼系統的學習 B 幀譯碼

(57)摘要

本文描述了用於處理視頻資料的技術，例如通過使用單向譯碼系統和內插參考幀來執行學習的雙向譯碼。例如，處理可以包括：獲得第一參考幀和第二參考幀。該處理可以包括：至少部分地通過在第一參考幀和第二參考幀之間執行插值來生成第三參考幀。該處理可以包括：例如通過估計輸入幀和第三參考幀之間的運動，基於第三參考幀對輸入幀執行單向幀間預測，並至少部分地通過基於估計的運動，變形第三參考幀一個或多個像素來生成變形幀。該處理可以包括：基於變形幀和預測殘差來生成表示輸入幀的重建幀，重建幀包括雙向預測幀。

Techniques are described for processing video data, such as by performing learned bidirectional coding using a unidirectional coding system and an interpolated reference frame. For example, a process can include obtaining a first reference frame and a second reference frame. The process can include generating a third reference frame at least in part by performing interpolation between the first reference frame and the second reference frame. The process can include performing unidirectional inter-prediction on an input frame based on the third reference frame, such as by estimating motion between an input frame and the third reference frame, and generating a warped frame at least in part by warping one or more pixels of the third reference frame based on the estimated motion. The process can include generating, based on the warped frame and a predicted residual, a reconstructed frame representing the input frame, the reconstructed frame including a bidirectionally-predicted frame.

指定代表圖：



【圖 6】

【發明摘要】

【中文發明名稱】 使用P幀譯碼系統的學習B幀譯碼

【英文發明名稱】 LEARNED B-FRAME CODING USING P-FRAME CODING
SYSTEM

【中文】

本文描述了用於處理視頻資料的技術，例如通過使用單向譯碼系統和內插參考幀來執行學習的雙向譯碼。例如，處理可以包括：獲得第一參考幀和第二參考幀。該處理可以包括：至少部分地通過在第一參考幀和第二參考幀之間執行插值來生成第三參考幀。該處理可以包括：例如通過估計輸入幀和第三參考幀之間的運動，基於第三參考幀對輸入幀執行單向幀間預測，並至少部分地通過基於估計的運動，變形第三參考幀一個或多個像素來生成變形幀。該處理可以包括：基於變形幀和預測殘差來生成表示輸入幀的重建幀，重建幀包括雙向預測幀。

【英文】

Techniques are described for processing video data, such as by performing learned bidirectional coding using a unidirectional coding system and an interpolated reference frame. For example, a process can include obtaining a first reference frame and a second reference frame. The process can include generating a third reference frame at least in part by performing interpolation between the first reference frame and the second reference frame. The process can include performing unidirectional inter-prediction on an input frame based on the third reference frame, such as by estimating motion between an input frame and the third reference frame, and generating a warped

frame at least in part by warping one or more pixels of the third reference frame based on the estimated motion. The process can include generating, based on the warped frame and a predicted residual, a reconstructed frame representing the input frame, the reconstructed frame including a bidirectionally-predicted frame.

【指定代表圖】 圖 6

【代表圖之符號簡單說明】

600 B 幀譯碼系統

611 輸入幀

613 內插參考幀

615 重建的 輸入幀

622 幀插值引擎

624 P 幀譯碼系統

Ref₀ 參考幀0

Ref₁ 參考幀1

【發明說明書】

【中文發明名稱】 使用P幀譯碼系統的學習B幀譯碼

【英文發明名稱】 LEARNED B-FRAME CODING USING P-FRAME CODING
SYSTEM

【技術領域】

【0001】 概括地說，本公開內容涉及視頻譯碼（例如，對視頻資料進行編碼和/或解碼）。例如，本公開內容的各方面涉及使用單向譯碼系統（例如，單向預測幀（P 幀）編碼器-解碼器或 CODEC）的學習雙向預測幀（B 幀）譯碼。

【先前技術】

【0002】 許多設備和系統允許處理和輸出視頻資料以供消費。數位視頻資料包括大量資料，以滿足視頻品質、性能和功能方面日益增長的需求。例如，視頻資料的消費者通常想要高保真度、解析度、幀速率等等的高品質視頻。滿足這些需求通常需要大量的視頻資料，這給處理和儲存視頻資料的通訊網路和設備帶來了巨大負擔。視頻譯碼技術可以用於對視頻資料進行壓縮。視頻譯碼的一個示例目標是將視頻資料壓縮成使用較低位元率的形式，同時避免或最小化視頻品質的降低。隨著不斷發展的視頻服務變得可用，以及對大量視頻資料的需求不斷增加，需要具有更好性能和效率的譯碼技術。

【發明內容】

【0003】 本文描述的系統和技術包括使用單向（例如，P幀）譯碼系統的學習的雙向（例如，B幀）譯碼，該譯碼系統可以利用通過在參考幀之間插值生成的參考幀。根據至少一個例子，提供了一種用於處理視頻資料的方法。該方法可以包括：獲得第一參考幀和第二參考幀；至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成第三參考幀；並基於所述第三參考幀對輸入幀執行單向幀間預測。

【0004】 在另一個例子中，提供了一種用於處理視頻資料的系統。該系統可以包括具有插值引擎的雙向譯碼引擎和單向譯碼引擎。在一些情況下，該系統可以包括至少一個記憶體和耦接到所述記憶體的一個或多個處理器（例如，在電路中實現）。所述插值引擎被配置為：獲得第一參考幀和第二參考幀；至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成第三參考幀。所述單向譯碼引擎被配置為基於所述第三參考幀對輸入幀執行單向幀間預測。

【0005】 在另一個例子中，提供了一種用於處理視頻資料的非臨時性計算機可讀媒體。該非臨時性計算機可讀媒體可以包括儲存在其上的指令，當所述指令被一個或多個處理器執行時，使所述一個或多個處理器執行以下操作：獲得第一參考幀和第二參考幀；至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成第三參考幀；並基於所述第三參考幀對輸入幀執行單向幀間預測。

【0006】 根據另一個例子，提供了一種用於處理視頻資料的裝置。該裝置可以包括：用於獲得第一參考幀和第二參考幀的構件；用於至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成第三參考幀的構件；以及用於基於所述第三參考幀對輸入幀執行單向幀間預測的構件。

【0007】 在一些方面，上面所描述的方法、裝置和非臨時性計算機可讀媒體可以包括：估計所述輸入幀和所述第三參考幀之間的運動；並至少部分地通過基於所述估計的運動，變形所述第三參考幀的一個或多個像素，來生成變形幀。

【0008】 在一些方面，上面所描述的方法、裝置和非臨時性計算機可讀媒體可以包括：至少部分地通過確定所述輸入幀和所述變形幀之間的差異來確定殘差；並使用所述殘差生成預測殘差。

【0009】 在一些方面，上面所描述的方法、裝置和非臨時性計算機可讀媒體可以包括：基於所述變形幀和所述預測殘差，生成表示所述輸入幀的重建幀，所述重建幀包括雙向預測幀；並使所述重建幀儲存在記憶體中。

【0010】 在一些方面，所述運動是使用第一自動編碼器估計的，並且所述預測殘差是使用第二自動編碼器生成的。

【0011】 在一些方面，上面所描述的方法、裝置和非臨時性計算機可讀媒體可以包括：獲得第四參考幀；估計另外輸入幀和所述第四參考幀之間的運動；至少部分地通過基於所述估計的運動來變形所述第四參考幀的一個或多個像素，生成另外的變形幀；並基於所述另外的變形幀和另外的預測殘差，生成表示所述另外輸入幀的重建幀，表示所述另外輸入幀的所述重建幀包括單向預測幀。

【0012】 在一些方面，上面所描述的方法、裝置和非臨時性計算機可讀媒體可以包括：確定表示從所述第一參考幀到所述第三參考幀的像素運動的第一組運動資訊；確定表示從所述第二參考幀到所述第三參考幀的像素運動的第二組運動資訊；至少部分地通過使用所述第一組運動資訊對所述第一參考幀執行變形函數來生成第一變形資訊；至少部分地通過使用所述第二組運動資訊對所

述第二參考幀執行所述變形函數來生成第二變形資訊；並基於所述第一變形資訊和所述第二變形資訊生成所述第三參考幀。

【0013】 在一些方面，基於所述第一參考幀和所述第二參考幀之間的像素運動來確定所述第一組運動資訊和所述第二組運動資訊。

【0014】 在一些方面，所述第一組運動資訊包括第一光流圖，並且所述第二組運動資訊包括第二光流圖。

【0015】 在一些方面，所述變形函數包括雙線性插值函數。

【0016】 在一些方面，所述輸入幀與所述第三參考幀之間的所述估計的運動包括光流資料。

【0017】 在一些方面，所述輸入幀與所述第三參考幀之間的所述估計的運動包括動態卷積資料。

【0018】 在一些方面，所述輸入幀與所述第三參考幀之間的所述估計的運動包括基於塊的運動資料。

【0019】 根據至少一個例子，提供了一種用於處理視頻資料的方法。該方法可以包括：獲得內插參考幀和表示輸入幀與所述內插參考幀之間運動的運動資訊；基於所述運動資訊，估計所述輸入幀和所述內插參考幀之間的運動；至少部分地通過基於所述估計的運動，對所述內插參考幀的一個或多個像素進行變形來生成變形幀；並基於所述變形幀和預測殘差來生成表示所述輸入幀的重建幀，所述重建幀包括雙向預測幀。

【0020】 在另一個例子中，提供了一種用於處理視頻資料的系統。該系統可以包括至少一個記憶體和耦接到所述記憶體的一個或多個處理器（例如，在電路中實現）。所述至少一個處理器被配置為：獲得內插參考幀和表示輸入幀與所

述內插參考幀之間運動的運動資訊；基於所述運動資訊，估計所述輸入幀和所述內插參考幀之間的運動；至少部分地通過基於所述估計的運動，對所述內插參考幀的一個或多個像素進行變形來生成變形幀；並基於所述變形幀和預測殘差來生成表示所述輸入幀的重建幀，所述重建幀包括雙向預測幀。

【0021】 在另一個例子中，提供了一種用於處理視頻資料的非臨時性計算機可讀媒體。該非臨時性計算機可讀媒體可以包括儲存於其上的指令，當所述指令被一個或多個處理器執行時，使所述一個或多個處理器執行以下操作：獲得內插參考幀和表示輸入幀與所述內插參考幀之間運動的運動資訊；基於所述運動資訊，估計所述輸入幀和所述內插參考幀之間的運動；至少部分地通過基於所述估計的運動，對所述內插參考幀的一個或多個像素進行變形來生成變形幀；並基於所述變形幀和預測殘差來生成表示所述輸入幀的重建幀，所述重建幀包括雙向預測幀。

【0022】 根據另一個例子，提供了一種用於處理視頻資料的裝置。該裝置可以包括：用於獲得內插參考幀和表示輸入幀與所述內插參考幀之間運動的運動資訊的構件；用於基於所述運動資訊，估計所述輸入幀和所述內插參考幀之間的運動的構件；用於至少部分地通過基於所述估計的運動，對所述內插參考幀的一個或多個像素進行變形來生成變形幀的構件；以及用於基於所述變形幀和預測殘差來生成表示所述輸入幀的重建幀的構件，所述重建幀包括雙向預測幀。

【0023】 在一些方面，上面所描述的方法、裝置和非臨時性計算機可讀媒體可以包括：獲得第一參考幀和第二參考幀；並至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成所述內插參考幀。

【0024】 在一些方面，上面所描述的方法、裝置和非臨時性計算機可讀媒體可以包括：從位元流中獲取殘差；並基於所獲得的殘差來生成預測殘差。

【0025】 在一些方面，所述運動資訊是從所述位元流中獲得的。

【0026】 在一些方面，上面所描述的方法、裝置和非臨時性計算機可讀媒體可以包括：獲得第四參考幀；估計另外輸入幀和所述第四參考幀之間的運動；至少部分地通過基於所估計的運動，變形所述第四參考幀的一個或多個像素來生成另外的變形幀；並基於所述另外的變形幀和另外的預測殘差，生成表示所述另外輸入幀的重建幀，表示所述另外輸入幀的所述重建幀包括單向預測幀。

【0027】 在一些方面，裝置可以包括移動設備、可穿戴設備、擴展實境設備（例如，虛擬實境（VR）設備、擴增實境（AR）設備或混合實境（MR）設備）、個人計算機、膝上型計算機、伺服器計算機、電視、視頻遊戲控制台或其它設備，也可以是這些設備的一部分。在一些方面，該裝置包括移動設備（例如，移動電話或所謂的「智慧型電話」）。在一些方面，該裝置還包括用於擷取一個或多個圖像或視頻幀的至少一個相機。例如，該裝置可以包括用於擷取一個或多個圖像和/或包括視頻幀的一個或多個視頻的一個相機（例如，RGB相機）或多個相機。在一些方面，該裝置包括用於顯示一個或多個圖像、視頻、通知或其它可顯示資料的顯示器。在一些方面，該裝置包括發射器，其被配置為通過傳輸媒體將重建的視頻幀傳輸到至少一個設備。在一些方面，處理器包括神經處理單元（NPU）、中央處理單元（CPU）、圖形處理單元（GPU）或其它處理設備或組件。

【0028】 本概括內容既不是旨在標識所要求保護主題的關鍵或必要特徵，也不是旨在單獨用於確定所要求保護主題的範圍。通過參考本專利的整個說明書的適當部分、任何或所有附圖以及每個請求項，應當理解本發明的主題。

【0029】 通過參考以下的說明書、申請專利範圍書和附圖，前述內容以及其它特徵和實施例將變得更加顯而易見。

【圖式簡單說明】

【0030】 下面結合附圖來詳細地描述本申請的說明性實施例：

【0031】 圖1根據本公開內容的一些例子，示出了可以實現本文所描述的各种技術的示例圖像處理系統；

【0032】 圖2A示出了根據本公開內容的一些例子的全連接神經網路的示例；

【0033】 圖2B示出了根據本公開內容的一些例子的局部連接的神經網路的示例；

【0034】 圖2C示出了根據本公開內容的一些例子的卷積神經網路的示例；

【0035】 圖2D示出了根據本公開內容的一些例子的、被設計為從圖像中識別特徵的深度卷積網路（DCN）的詳細示例；

【0036】 圖3是示出根據本公開內容的一些例子的另一種示例性DCN的方塊圖；

【0037】 圖4是根據本公開內容的一些例子，示出被配置為執行單向運動壓縮和補償以生成預測的單向預測幀（P幀）譯碼系統的例子圖；

【0038】 圖5A是根據本公開內容的一些例子，示出被配置為執行雙向運動壓縮和補償以生成預測的雙向預測幀（B幀）譯碼系統的例子圖；

【0039】 圖5B是根據本公開內容的一些例子，示出被配置為使用幀內插來生成預測的B幀譯碼系統的例子圖；

【0040】圖6是根據本公開內容的一些例子，示出被配置為使用P幀譯碼系統和幀插值引擎來執行學習的B幀譯碼的B幀譯碼系統的例子圖；

【0041】圖7A是根據本公開內容的一些例子，示出可以用於執行單向譯碼和/或雙向譯碼的P幀譯碼系統的例子圖；

【0042】圖7B是根據本公開內容的一些例子，示出可以用於使用P幀譯碼系統執行雙向譯碼的B幀譯碼系統的例子圖；

【0043】圖7C是根據本公開內容的一些例子，示出可以由B幀譯碼系統使用的幀插值引擎的例子圖；

【0044】圖8是示出根據本公開內容的一些例子的運動估計技術的例子圖；

【0045】圖9是示出根據本公開內容的一些例子的I幀、B幀、I幀（IBI）譯碼技術的例子圖；

【0046】圖10是示出根據本公開內容的一些例子的I幀、B幀、P幀（IBP）譯碼技術的例子圖；

【0047】圖11是根據本公開內容的一些例子，示出用於雙向幀預測的示例順序方案的圖；

【0048】圖12是根據本公開內容的一些例子，示出用於雙向幀預測的示例分層方案的圖；

【0049】圖13A和圖13B是根據本公開內容的一些例子，示出本文所描述的B幀譯碼系統的性能結果的圖表；

【0050】圖14A-圖14F是根據本公開內容的一些例子，示出本文所描述的B幀譯碼系統的進一步性能結果的圖表；

【0051】圖15是根據本公開內容的一些例子，示出用於處理視頻資料的處理的示例流程圖；

【0052】圖16是根據本公開內容的一些例子，示出用於處理視頻資料的處理的另一個例子的流程圖；以及

【0053】圖17示出了可以用於實現本文描述的各個方面的示例計算系統。

【實施方式】

【0054】下面提供本公開內容的某些方面和實施例。對於本領域通常知識者顯而易見的是，這些方面和實施例中的一些可以獨立地應用，並且它們中的一些可以組合地應用。在下面的描述中，為了便於解釋起見，闡述了特定的細節以便提供對本申請的實施例的透徹理解。但是，顯而易見的是，可以在不具有這些特定細節的情況下實踐各個實施例。附圖和描述並不旨在是限制性的。

【0055】隨後的描述僅提供了示例性實施例，並且無意於限制本公開內容的保護範圍、適用性或配置。相反，對示例性實施例的隨後描述，將向本領域通常知識者提供用於實現示例性實施例的使能描述。應當理解的是，在不脫離所附申請專利範圍書闡述的本申請的精神和保護範圍的情況下，可以對元件的功能和佈置進行各種改變。

【0056】數位視頻資料可以包括大量資料，特別是隨著對高品質視頻資料的需求持續增長。例如，視頻資料的消費者通常希望視頻品質越來越高，具有高保真度、解析度、幀速率等。然而，滿足這些需求通常需要大量的視頻資料，這給處理和儲存視頻資料的通訊網路和設備帶來了巨大負擔。

【0057】可以使用各種技術來對視頻資料進行譯碼。在一些情況下，可以根據特定的視頻譯碼標準和/或方案來執行視頻譯碼。示例視頻譯碼標準包括高效視頻編碼（HEVC）、高級視頻編碼（AVC）、運動圖像專家組（MPEG）編

碼、通用視頻編碼（VVC）等。視頻譯碼技術的一個目標是將視頻資料壓縮成使用較低位元率的形式，同時避免或最小化視頻品質的下降。隨著對視頻服務需求的增長和新視頻服務的出現，需要具有更高效率和性能的譯碼技術。

【0058】 視頻譯碼可以使用諸如幀內預測或幀間預測之類的預測方法，它們利用存在於視頻幀或其它圖像或幀序列中的冗餘。幀內預測是使用單個視頻幀內的資料執行的，並且基於幀的空間特徵。使用幀內預測編碼的幀稱為 I 幀。幀的幀間預測是基於該幀相對於其它幀的時間特性來執行的。例如，可以通過識別包括相對於視頻幀的變化的其它視頻幀的區域和包括相對於視頻幀的冗餘的區域（例如，基本保持不變的背景區域），來執行視頻幀的幀間預測。可以去除冗餘，從而導致視頻幀的殘差。可以進一步對殘差進行編碼（例如，使用熵譯碼），並且可以將結果包括在儲存、傳輸或以其它方式輸出的位元流中。

【0059】 幀間預測的示例包括單向預測（uni-prediction）和雙向預測（bi-prediction）。單向預測包括在執行幀間預測時使用單個參考幀。使用單向預測譯碼的幀稱為 P 幀。雙向預測涉及在執行幀的幀間預測時使用兩個參考幀。使用雙向預測編碼的幀稱為 B 幀。

【0060】 在一些情況下，可以使用機器學習系統來執行視頻編碼（壓縮）和解碼（解壓縮）。一般來說，機器學習（ML）是人工智慧型（AI）的一個子集。ML 系統包括計算機系統可以用來通過依賴型樣（pattern）和推理來執行各種任務的演算法和統計模型。ML 系統的一個例子是神經網路（也稱為人工神經網路），它可以包括一組互連的人工神經元（例如，神經元模型）。神經網路可以用於各種應用和/或設備，例如圖像分析和/或計算機視覺應用、網際網路協議（IP）相機、物聯網（IoT）設備、自主車輛、服務機器人等等。

【0061】神經網路中的各個節點可以通過獲取輸入資料並對資料執行操作來模擬生物神經元。對輸入資料執行的操作的結果，有選擇地傳遞給其它神經元。權重值與網路中的每個向量和節點相關聯，並且這些值限制了輸入資料與輸出資料的關聯方式。例如，可以將每個節點的輸入資料乘以相應的權重值，然後將乘積相加。可以通過可選的偏差來調整乘積之和，並且可以將啟動函數應用於該結果，產生節點的輸出訊號或「輸出啟動」（有時稱為啟動圖或特徵圖）。最初可以由通過網路來訓練資料的迭代流，來確定這些權重值。例如，可以在網路學習如何通過其典型輸入資料特徵識別特定類別的訓練階段，建立權重值。在一個例子中，可以通過調整與神經元相關聯的參數值（例如，啟動參數和/或權重、偏差等）、添加和/或移除神經元或甚至神經元的層、添加或刪除神經元之間的邊緣等，來訓練網路以學習特定任務。

【0062】存在不同類型的神經網路，例如自動編碼器、卷積神經網路（CNN）、循環神經網路（RNN）、多層感知器（MLP）神經網路等。卷積神經網路可以包括人工神經元的集合，每個人工神經元都有一個感受野（例如，輸入空間的空間局部區域），並且共同平鋪輸入空間。卷積神經網路有許多應用。例如，CNN 可以廣泛地用於型樣識別和分類領域。RNN 的工作原理是保存層的輸出，並將該輸出反饋給輸入，以幫助預測層的結果。在 MLP 神經網路中，可以將資料饋送到輸入層，並且一個或多個隱藏層為資料提供抽象級別。然後可以基於抽象資料對輸出層進行預測。MLP 可能特別適用於為輸入分配類別或標籤的分類預測問題。

【0063】在分層神經網路架構中（當存在多個隱藏層時，稱為深度神經網路），第一層人工神經元的輸出成為第二層人工神經元的輸入，第二層人工神經

元的輸出成為第三層人工神經元的輸入，依此類推。可以訓練卷積神經網路以識別特徵層次結構。卷積神經網路架構中的計算可以分佈在一組處理節點上，可以將這些節點配置在一個或多個計算鏈中。這些多層架構可以一次訓練一層，並且可以使用反向傳播進行微調。

【0064】 在一些例子中，基於機器學習的 **P** 幀和 **B** 幀譯碼系統可以用於分別執行單向預測和雙向預測。在一些情況下，此類系統可以包括神經網路架構（例如，一個或多個深度神經網路，例如一個或多個自動編碼器）。基於機器學習的 **P** 幀譯碼系統的示例可以對當前幀和參考幀執行運動壓縮和運動補償，以確定當前幀和參考幀之間的運動預測。該運動預測可以用於修改參考幀的像素（例如，通過根據包括在運動預測中的運動向量，移動參考幀的像素），從而產生對當前幀的預測。**P** 幀譯碼系統的殘差部分可以生成表示預測幀和當前幀之間的差異的預測殘差。預測的殘差可以與預測相結合，以生成重建的當前幀。

【0065】 可以使用各種類型的基於機器學習的 **B** 幀譯碼系統來執行雙向預測。基於機器學習的 **B** 幀譯碼系統的一個示例類似於上面所描述的 **P** 幀系統，但使用兩個參考幀而不是單個參考幀來生成重建的當前幀。基於機器學習的 **B** 幀譯碼系統的另一個例子包括幀插值引擎，其基於兩個參考幀之間的插值來生成預測。然後，殘差部分確定與來自插值引擎的預測相結合的預測殘差。

【0066】 本文描述了用於使用單向（例如，**P** 幀）譯碼系統來執行學習的雙向（例如，**B** 幀）譯碼的系統、方法（也稱為過程）、裝置和計算機可讀媒體（統稱為「系統和技術」）。如本文所使用的，術語譯碼可以指代編碼（例如，壓縮）、解碼（例如，解壓縮）、或編碼和解碼兩者。例如，提供了一種雙向譯碼系統（也稱為 **B** 幀譯碼系統），其包括用於基於至少兩個參考幀來生成內插

參考幀的幀插值引擎。在一些例子中，幀插值引擎可以使用與用於生成內插參考幀的參考幀相關聯的運動資訊（例如，光流或其它基於運動的技術）來生成內插參考幀。

【0067】 雙向譯碼系統可以使用單向譯碼系統（也稱為 **P** 幀譯碼系統）來生成雙向預測幀（**B** 幀）。例如，單向譯碼系統可以使用內插參考幀來生成 **B** 幀。**B** 幀可以包括通過參考來自先前幀和未來幀（例如，在顯示順序和/或解碼順序中在 **B** 幀之前和之後的幀）的資料而生成的一種幀。除了用於 **B** 幀譯碼外，單向譯碼系統還可以用於在對特定幀執行單向預測模式時，生成單向預測幀（**P** 幀）。例如，單向譯碼系統可以使用實際參考幀來生成 **P** 幀。**P** 幀可以包括通過參考來自單個參考幀的資料而生成的一種幀，該參考幀可以包括先前幀（例如，在譯碼順序中在 **B** 幀之前的幀）或未來幀（例如，在譯碼順序上在 **B** 幀之後的幀）。如本文所使用的，術語幀、圖片和圖像可以互換地使用。

【0068】 在一些例子中，單向譯碼系統可以包括視頻編碼器和視頻解碼器，它們可以使用機器學習來處理視頻資料並生成對輸入幀的預測。在一些方面，視頻編碼器可以包括運動預測系統（例如，運動自動編碼器）的編碼器神經網路，並且視頻解碼器可以包括運動預測系統的解碼器神經網路。例如，視頻編碼器（例如，運動預測系統的編碼器神經網路）可以使用機器學習來生成表示輸入幀和參考幀（例如，在執行雙向預測時的內插參考幀或實際參考幀，例如在執行單向預測時不是基於插值生成的重建的 **P** 或 **B** 幀或 **I** 幀）之間的預測運動的潛在資料（例如，位元流）。視頻解碼器（例如，運動預測系統的解碼器神經網路）可以使用機器學習來處理潛在資料，並預測或重建輸入幀和參考幀之間的運動。單向譯碼系統可以使用預測的運動，通過變形參考幀的相應部分來生成輸入幀

的預測。例如，預測的運動可以包括運動向量（例如，包括輸入幀的每個像素或像素塊的運動向量的光流圖）。變形引擎可以根據運動向量，移動參考幀的像素以生成對輸入幀的預測（稱為變形幀）。

【0069】 在一些例子中，單向譯碼系統的視頻編碼器和視頻解碼器也可以使用機器學習來生成輸入幀的殘差。在一些方面，單向譯碼系統可以包括殘差預測系統（例如，殘差自動編碼器），其包括編碼器神經網路和解碼器神經網路，這些網路可以為輸入幀生成殘差。例如，視頻編碼器（例如，殘差預測系統的編碼器神經網路）可以接收殘差作為輸入，該殘差包括輸入幀和參考幀（例如，當執行雙向預測時的內插參考幀，或執行單向預測時的實際參考幀）之間的差異。使用殘差作為輸入，視頻編碼器可以使用機器學習來生成表示殘差的潛在資料（例如，位元流）。視頻解碼器（例如，殘差預測系統的解碼器神經網路）可以使用機器學習來處理潛在資料，並預測或重建殘差。單向譯碼系統可以通過將預測的殘差添加到為輸入幀生成的預測（例如，由變形引擎生成），來生成表示輸入幀的重建幀。

【0070】 在一些例子中，系統可以將表示運動的潛在資料和表示殘差的潛在資料（例如，在編碼的視頻位元流中），傳輸到另一個設備以進行解碼。例如，另一個設備可以包括視頻解碼器（例如，如上所述的運動預測系統和殘差預測系統），該視頻解碼器可以使用機器學習，基於表示運動的潛在資料和表示殘差的潛在資料來重建輸入幀，類似於上面描述的（例如，通過處理運動潛在資料來預測輸入幀和參考幀之間的運動，通過根據預測的運動來變形參考幀來生成輸入幀的預測，處理殘差潛在資料來預測殘差，並通過將預測的殘差與為輸入幀生成的預測相加來生成重建幀）。

【0071】 將關於附圖描述系統和技術的其它方面。圖 1 示出了圖像處理系統 100 的示例實現，在一些情況下，該圖像處理系統 100 可以用於實現本文所描述的系統和技術。圖像處理系統 100 可以包括中央處理單元（CPU）102 或多核 CPU，其被配置為執行本文描述的功能中的一個或多個。參數或變量（例如，神經訊號和突觸權重）、與計算設備相關聯的系統參數（例如，具有權重的神經網路）、延遲、頻點資訊、任務資訊、圖像資料等資訊可以儲存在與神經處理單元（NPU）108 相關聯的記憶體塊中，儲存在與 CPU 102 相關聯的記憶體塊中，儲存在與圖形處理單元（GPU）104 相關聯的記憶體塊中，儲存在與數位訊號處理器（DSP）106 相關聯的記憶體塊中，儲存在記憶體塊 118 中，和/或可以分佈在多個塊中。在 CPU 102 處執行的指令可以從與 CPU 102 相關聯的程式記憶體和/或從記憶體塊 118 加載。

【0072】 圖像處理系統 100 還可以包括用於執行特定功能的其它處理塊，例如 GPU 104；DSP 106；連接塊 110（其可以包括第五代（5G）連接、第四代長期演進（4G LTE）連接、Wi-Fi 連接、USB 連接、藍牙連接等）；和/或可以例如檢測圖像特徵的多媒體處理器 112。在一些例子中，可以在 CPU 102、DSP 106 和/或 GPU 104 中實現 NPU 108。在一些情況下，圖像處理系統 100 還可以包括一個或多個感測器 114、一個或多個圖像訊號處理器（ISP）116，和/或儲存設備 120。

【0073】 在一些例子中，圖像處理系統 100 可以為一個或多個處理器實現 ARM 指令集架構。在本公開內容的一個方面，加載到 CPU 102 中的指令可以包括：用於在查找表（LUT）中搜索儲存的與輸入值和過濾器權重的乘積相對應的乘法結果的代碼。加載到 CPU 102 中的指令還可以包括：用於當檢測到乘積的

查找表命中時，在乘積的乘法運算期間禁用乘法器的代碼。此外，加載到 CPU 102 中的指令可以包括：用於在檢測到乘積的查找表未命中時，儲存輸入值和過濾器權重的計算乘積的代碼。

【0074】圖像處理系統 100 可以是計算設備或多個計算設備的一部分。在一些例子中，圖像處理系統 100 可以是電子設備（或多個電子設備）的一部分，例如相機系統（例如，數碼相機、IP 相機、攝像機、安全相機等）、電話系統（如，智慧型手機、蜂窩電話、會議系統等）、桌面型計算機、XR 設備（例如，頭戴式顯示器等）、智慧型可穿戴設備（例如，智慧型手錶、智慧型眼鏡等）、膝上型計算機或筆記本計算機、平板計算機、機頂盒、電視、顯示設備、單晶片系統（SoC）、數位媒體播放器、遊戲機、視頻流設備、伺服器、無人機、車內計算機（computer in a car）、物聯網（IoT）設備、或者任何其它適當的電子設備。

【0075】在一些實施方式中，CPU 102、GPU 104、DSP 106、NPU 108、連接塊 110、多媒體處理器 112、一個或多個感測器 114、ISP 116、記憶體塊 118 和/或儲存設備 120 可以是同一計算設備的一部分。例如，在一些情況下，CPU 102、GPU 104、DSP 106、NPU 108、連接塊 110、多媒體處理器 112、一個或多個感測器 114、ISP 116、記憶體塊 118 和/或儲存設備 120 可以整合到智慧型手機、膝上型計算機、平板計算機、智慧型可穿戴設備、視頻遊戲系統、服務器和/或任何其它計算設備中。在其它實施方式中，CPU 102、GPU 104、DSP 106、NPU 108、連接塊 110、多媒體處理器 112、一個或多個感測器 114、ISP 116、記憶體塊 118 和/或儲存設備 120 可以是兩個或更多個單獨的計算設備的一部分。

【0076】圖像處理系統 100 和/或其組件可以被配置為使用根據本文所討論的本公開內容的各方面的技術，來執行視頻壓縮和/或解壓縮（也稱為視頻編碼和/或解碼，統稱為視頻編碼）。通過使用深度學習架構和本文描述的技術來執行視頻壓縮和/或解壓縮，本公開內容的各方面可以提高設備上的視頻壓縮和/或解壓縮的效率和/或減少相關聯的資源需求和/或使用。例如，使用本文描述的視頻譯碼技術的設備可以更高效地壓縮視頻資料，可以減少以壓縮視頻資料傳輸到目的地設備的資料量，並且目的地設備可以高效地接收和解壓縮經壓縮的視頻資料。在一些例子中，本文描述的深度學習架構和技術可以減少在譯碼設備或組件（例如，編碼器和解碼器）之間交換的資料量，以對視頻內容進行譯碼。為視頻譯碼傳輸的減少的資料量可以減少延遲，提高性能，並減少計算資源（例如，頻寬、記憶體、儲存、功率、計算、硬體等）的成本或負擔。

【0077】如上所述，神經網路是機器學習系統的示例，並且可以包括輸入層、一個或多個隱藏層以及輸出層。從輸入層的輸入節點提供資料，由一個或多個隱藏層的隱藏節點執行處理，並通過輸出層的輸出節點來產生輸出。深度學習網路通常包括多個隱藏層。神經網路的每一層都可以包括特徵圖或啟動圖，其中這些圖可以包括人工神經元（或節點）。特徵圖可以包括過濾器、內核等。節點可以包括用於指示一個或多個層的節點的重要性的一個或多個權重。在一些情況下，深度學習網路可以具有一系列許多隱藏層，早期層用於確定輸入的簡單和低級特徵，而後面的層則建立更複雜和抽象特徵的層次結構。

【0078】深度學習架構可以學習特徵的層次結構。例如，如果呈現視覺資料，第一層可以進行學習，以識別輸入流中的特徵（例如，邊緣）。在另一個例子中，如果呈現聽覺資料，則第一層可以進行學習，以識別特定頻率中的頻譜功

率。第二層以第一層的輸出作為輸入，可以學習識別特徵，例如視覺資料的形狀或聽覺資料的聲音組合。例如，更高層可以進行學習，以表示視覺資料中的複雜形狀或聽覺資料中的單詞。更高的層可以進行學習，以識別常見的視覺物件和/或口語片語。

【0079】 神經網路可以被設計成具有多種連接型樣。在前饋網路中，將資訊從較低層傳遞到較高層，給定層中的每個神經元都與較高層中的神經元進行通訊。如上所述，可以在前饋網路的連續層中建立分層表示。神經網路也可以具有循環或反饋（也稱為自上而下）連接。在循環連接中，給定層中神經元的輸出可以傳遞到同一層中的另一個神經元。循環架構可以有助於識別跨越多個輸入資料塊的型樣，這些輸入資料塊按順序傳遞給神經網路。從給定層中的神經元到較低層中的神經元的連接稱為反饋（或自頂向下）連接。當對高級概念的識別可以有助於區分輸入的特定低級特徵時，具有許多反饋連接的網路會有所幫助。

【0080】 神經網路的層之間的連接可以是全連接的或局部連接的。圖 2A 示出了全連接神經網路 202 的示例。在全連接神經網路 202 中，第一隱藏層中的神經元可以將其輸出傳送到第二隱藏層中的每個神經元，使得第二層中的每個神經元將接收來自第一層中每個神經元的輸入。圖 2B 示出了局部連接的神經網路 204 的示例。在局部連接的神經網路 204 中，第一隱藏層中的神經元可以連接到第二隱藏層中的有限數量的神經元。更一般地，局部連接神經網路 204 的局部連接層可以被配置為使得層中的每個神經元將具有相同或相似的連接型樣，但連接強度可以具有不同的值（例如，210、212、214 和 216）。局部連接的連接型樣可以在較高層中產生空間上不同的感受野，這是因為給定區域中的較高層神經元可以接收通過訓練調整到網路總輸入的受限部分的屬性的輸入。

【0081】 局部連接的神經網路的一個例子是卷積神經網路。圖 2C 示出了卷積神經網路 206 的示例。卷積神經網路 206 可以被配置為使得與第二層中的每個神經元的輸入相關聯的連接強度是共享的（例如，208）。卷積神經網路可能非常適合輸入的空間位置有意義的問題。根據本公開內容的各方面，卷積神經網路 206 可以用於執行視頻壓縮和/或解壓縮的一個或多個方面。

【0082】 一種類型的卷積神經網路是深度卷積網路（DCN）。圖 2D 示出了 DCN 200 的示例，該 DCN 200 被設計為識別來自圖像擷取設備 230（例如，相機或圖像感測器）輸入的圖像 226 的特徵。在一些例子中，可以對當前示例的 DCN 200 進行訓練，以識別圖像 226 中的視覺特徵（例如，圖像 226 中的一個或多個物件或標誌）。

【0083】 在一些例子中，可以用監督學習來訓練 DCN 200。在訓練期間，可以向 DCN 200 呈現圖像（例如，圖像 226），然後可以計算前向傳遞以產生輸出 222。DCN 200 可以包括特徵提取部分和分類部分。在接收到圖像 226 時，卷積層 232 可以將卷積核（沒有示出）應用於圖像 226，以生成第一組特徵圖 218。舉一個例子，卷積層 232 的卷積核可以是 5x5 核，其生成 28x28 的特徵圖。在本例子中，因為在第一組特徵圖 218 中生成了四個不同的特徵圖，所以在卷積層 232 將四個不同的卷積核應用於圖像 226。卷積核也可以稱為濾波器或卷積濾波器。

【0084】 第一組特徵圖 218 可以由最大池化層（沒有示出）進行二次採樣以生成第二組特徵圖 220。最大池化層減小了第一組特徵圖 218 的大小。也就是說，第二組特徵圖 220 的大小（例如，14x14）小於第一組特徵圖 218 的大小（例如，28x28）。減小後的大小為後續層提供了類似的資訊，同時減少了記憶體消

耗。可以通過一個或多個後續卷積層(沒有示出)進一步卷積第二組特徵圖 220，以生成一組或多個後續特徵圖集合(沒有示出)。

【0085】在圖 2D 的例子中，對第二組特徵圖 220 進行卷積以生成第一特徵向量 224。此外，對第一特徵向量 224 進一步卷積以生成第二特徵向量 228。第二特徵向量 228 的每個特徵可以包括與圖像 226 的可能特徵相對應的標記，例如「符號」、「60」和「100」。softmax 函數(沒有示出)可以將第二特徵向量 228 中的標記轉換為機率。這樣，DCN 200 的輸出 222 是圖像 226 包括一個或多個特徵的機率。

【0086】在本例子中，輸出 222 中「sign」和「60」的機率高於輸出 222 中其它標記(例如，「30」、「40」、「50」、「70」、「80」、「90」和「100」)的機率。在訓練之前，DCN 200 產生的輸出 222 可能是不正確的。因此，可以計算輸出 222 和目標輸出之間的誤差。目標輸出是圖像 226 的基本事實(例如，「sign」和「60」)。然後，可以調整 DCN 200 的權重，使得 DCN 200 的輸出 222 更接近於目標輸出。

【0087】為了調整權重，學習演算法可以計算權重的梯度向量。梯度可以指示在調整權重時誤差將增加或減少的量。在頂層，梯度可以直接對應於連接倒數第二層中啟動的神經元和輸出層中的神經元的權重值。在較低層中，梯度可以取決於權重的值和較高層的計算誤差梯度。然後，可以調整權重以減少誤差。這種調整權重的方式可以稱為「反向傳播」，因為它涉及通過神經網路的「反向傳遞」。

【0088】在實踐中，可以在少量示例上計算權重的誤差梯度，使得計算的梯度接近真實的誤差梯度。這種近似方法可以稱為隨機梯度下降。可以重複隨機

梯度下降，直到整個系統的可實現錯誤率停止下降或直到錯誤率達到目標水平為止。在學習之後，可以向 DCN 呈現新圖像，並且通過網路的前向傳遞可以產生可以被認為是 DCN 的推斷或預測的輸出 222。

【0089】深度置信網路 (DBN) 是包含多層隱藏節點的機率模型。DBN 可以用於提取訓練資料集的分層表示。可以通過堆疊受限玻爾茲曼機 (RBM) 層來獲得 DBN。RBM 是一種人工神經網路，其可以學習一組輸入的機率分佈。由於 RBM 可以在沒有關於每個輸入應歸入的類別資訊的情況下學習機率分佈，因此 RBM 通常用於無監督學習。使用混合的無監督和監督範式，可以以無監督方式訓練 DBN 的底部 RBM 並用作特徵提取器，而可以以監督方式訓練頂部 RBM (在前一層輸入和目標類別的聯合分佈上)，並且其可以用作分類器。

【0090】深度卷積網路 (DCN) 是卷積網路的網路，其配置有另外的池化層和歸一化層。DCN 可以在許多任務上實現高性能。可以使用監督學習來訓練 DCN，其中輸入和輸出目標對於許多示例都是已知的，並用於通過使用梯度下降方法來修改網路的權重。

【0091】DCN 可以是前饋網路。此外，如上所述，在第一層中的神經元之間共享從 DCN 的第一層中的神經元到下一個更高層中的一組神經元的連接。DCN 的前饋和共享連接可以用於快速處理。例如，DCN 的計算負擔可能遠小於包含循環或反饋連接的類似大小的神經網路的計算負擔。

【0092】可以認為卷積網路的每一層的處理是空間不變的模板或基礎投影。如果首先將輸入分解到多個通道，例如彩色圖像的紅色、綠色和藍色通道，那麼可以認為在該輸入上訓練的卷積網路是三維的，兩個空間維度沿著圖像的軸，並且第三維擷取顏色資訊。可以將卷積連接的輸出視為在後續層中形成特徵

圖，其中該特徵圖的每個元素（例如，220）接收來自前一層中的一系列神經元（例如，特徵圖 218）和來自多個通道中的每一個通道的輸入。可以用非線性進一步處理（例如，校正， $\max(0, x)$ ）特徵圖中的值。來自相鄰神經元的值可以進一步合併，這對應於下採樣，並且可以提供另外的局部不變性和降維。

【0093】圖 3 是示出深度卷積網路 350 的例子的方塊圖。深度卷積網路 350 可以包括基於連通性和權重共享的多個不同類型的層。如圖 3 中所示，深度卷積網路 350 包括卷積塊 354A、354B。卷積塊 354A、354B 中的每一個可以配置有卷積層（CONV）356、歸一化層（LNorm）358 和最大池化層（MAX POOL）360。

【0094】卷積層 356 可以包括能夠應用於輸入資料 352 以生成特徵圖的一個或多個卷積濾波器。儘管僅示出了兩個卷積塊 354A、354B，但本公開內容不限於此，而是可以根據設計偏好，在深度卷積網路 350 中包括任意數量的卷積塊（例如，塊 354A、354B）。歸一化層 358 可以對卷積濾波器的輸出進行歸一化。例如，歸一化層 358 可以提供白化或橫向抑制。最大池化層 360 可以在空間上提供下採樣聚合，以用於局部不變性和降維。

【0095】例如，深度卷積網路的並行濾波器組可以加載到圖像處理系統 100 的 CPU 102 或 GPU 104 上以實現高性能和低功耗。在一些例子中，並行濾波器組可以加載在圖像處理系統 100 的 DSP 106 或 ISP 116 上。深度卷積網路 350 可以存取可能存在於圖像處理系統 100 上的其它處理塊。

【0096】深度卷積網路 350 可以包括一個或多個全連接層，例如層 362A（標記為「FC1」）和層 362B（標記為「FC2」）。深度卷積網路 350 可以包括邏輯回歸（LR）層 364。在深度卷積網路 350 的每一層 356、358、360、362、

364 之間是要更新的權重（沒有示出）。每一層（例如，356、358、360、362、364）的輸出可以用作深度卷積網路 350 中的後續層（例如，356、358、360、362、364）的輸入，以從在第一個卷積塊 354A 處提供的輸入資料 352（例如，圖像、音頻、視頻、感測器資料和/或其它輸入資料）學習分層特徵表示。深度卷積網路 350 的輸出是針對輸入資料 352 的分類得分 366。分類得分 366 可以是一組機率，其中每個機率是輸入資料包括來自一組特徵中的特徵的機率。

【0097】 另一種類型的神經網路是自動編碼器。可以對自動編碼器進行訓練（例如，使用訓練資料和一個或多個損失函數），以接收輸入並在其輸出處生成該輸入的版本（例如，實質上將其輸入複製到其輸出）。可以對自動編碼器進行訓練，以無監督的方式學習高效的資料編碼。例如，給定一個物件的圖像，自動編碼器可以首先將圖像編碼為較低維的潛在表示，然後可以將潛在表示解碼回該物件的圖像。自動編碼器可以學習（通過訓練）壓縮輸入資料，同時最小化重建誤差。

【0098】 如上所述，數位視頻資料可以包括大量的資料，這會給處理和儲存視頻資料的通訊網路和設備帶來很大的負擔。例如，錄製未壓縮的視頻內容通常會導致檔案大小隨著錄製視頻內容的解析度增加而大大增加。在一個說明性示例中，以 1080p/24 錄製的未壓縮的每通道 16 位視頻（例如，解析度為 1920 像素寬和 1080 像素高，每秒擷取 24 幀）可能每幀佔用 12.4 百萬字節，或每秒佔用 297.6 百萬字節。以每秒 24 幀的 4K 解析度錄製的未壓縮的每通道 16 位視頻，可能每幀佔用 49.8 百萬字節，或每秒佔用 1195.2 百萬字節。

【0099】 網路頻寬是大視頻檔案可能成為問題的另一個約束。例如，視頻內容通常通過無線網路（例如，通過 LTE、高級 LTE、新無線電（NR）、WiFi™、

Bluetooth™ 或其它無線網路) 傳送，並且可以構成消費者網際網路流量的很大一部分。因此，希望減少用於在這些網路中傳送視頻內容的頻寬量。

【0100】 因為未壓縮的視頻內容可能導致相當大的記憶體可能涉及用於實體儲存和相當大的頻寬用於傳輸的大型檔案，所以可以使用視頻譯碼技術來壓縮和解壓縮這樣的視頻內容，如本文進一步描述的。為了減小視頻內容的大小——以及因此儲存視頻內容所涉及的儲存量和傳送視頻內容所涉及的頻寬量——可以根據特定的視頻譯碼標準和/或方案（例如，HEVC、AVC、MPEG、VVC 等）來執行各種視頻譯碼技術。視頻譯碼可以使用諸如幀間預測或幀內預測之類的預測方法，這些方法利用了視頻圖像或序列中存在的冗餘。視頻譯碼技術的一個目標是將視頻資料壓縮成使用較低位元率的形式，同時避免或最小化視頻品質的下降。隨著對視頻服務需求的增長和新視頻服務的出現，需要具有更高譯碼效率、性能和碼率控制的譯碼技術。

【0101】 編碼設備可以根據視頻譯碼標準對視頻資料進行編碼以生成經編碼的視頻位元流。在一些例子中，經編碼的視頻位元流（或「視頻位元流」或「位元流」）可以包括一系列一個或多個編碼視頻序列。編碼設備可以通過將每個圖片分割成多個切片來生成圖片的編碼表示。切片獨立於其它切片，因此在不依賴於來自同一圖片內的其它切片的資料的情況下，對切片中的資訊進行譯碼。一個切片包括一個或多個切片段，這些切片段包括獨立切片段、以及一個或多個依賴於先前切片段的從屬切片段（如果存在的話）。在 HEVC 中，將切片劃分為亮度樣本和色度樣本的譯碼樹塊（CTB）。亮度樣本的 CTB 和色度樣本的一個或多個 CTB 以及樣本的語法稱為譯碼樹單元（CTU）。CTU 也可以稱為「樹塊」或「最大譯碼單元」（LCU）。CTU 是 HEVC 編碼的基本處理單元。一個

CTU 可以拆分為多個不同大小的譯碼單元 (CU)。CU 包含稱為譯碼塊 (CB) 的亮度和色度樣本陣列。

【0102】 可以進一步將亮度和色度 CB 拆分成預測塊 (PB)。PB 是亮度分量或色度分量的樣本塊，它使用相同的運動參數進行幀間預測或塊內複製 (IBC) 預測 (當可用或啟用時)。亮度 PB 和一個或多個色度 PB 與相關聯的語法一起形成預測單元 (PU)。對於幀間預測，在每個 PU 的位元流中用訊號通知一組運動參數 (例如，一個或多個運動向量、參考索引等)，並且這些參數用於亮度 PB 和一個或多個色度 PB 的幀間預測。運動參數也可以稱為運動資訊。也可以將 CB 劃分為一個或多個變換塊 (TB)。TB 表示顏色分量的樣本的正方形塊，在其上應用殘差變換 (例如，在某些情況下相同的二維變換) 以對預測殘差訊號進行譯碼。變換單元 (TU) 表示亮度和色度樣本的 TB 以及相應的語法元素。下面更詳細地描述變換譯碼。

【0103】 根據 HEVC 標準，可以使用 TU 來執行變換。可以基於給定 CU 內的 PU 的大小來確定 TU 的大小。TU 可以與 PU 大小相同或更小。在一些例子中，可以使用稱為殘差四叉樹 (RQT) 的四叉樹結構，將對應於 CU 的殘差樣本細分成更小的單元。RQT 的葉子節點可以對應 TU。可以將與 TU 相關聯的像素差值進行變換以產生變換係數。然後，可以由編碼設備對變換係數進行量化。

【0104】 一旦將視頻資料的圖片分割成 CU，編碼設備就使用預測模式來預測每個 PU。然後，從原始視頻資料中減去預測單元或預測塊以獲得殘差 (如下所述)。對於每個 CU，可以使用語法資料，在位元流內用訊號通知預測模式。預測模式可以包括幀內預測 (或圖片內預測) 或幀間預測 (或圖片間預測)。幀內預測利用圖片內空間相鄰樣本之間的相關性。例如，使用幀內預測，從同一圖

片中的相鄰圖像資料預測每個 PU，例如使用 DC 預測來找到 PU 的平均值，使用平面預測以將平面表面擬合到 PU，使用方向預測從相鄰資料進行推斷，或任何其它適當類型的預測。幀間預測使用圖片之間的時間相關性來導出圖像樣本塊的運動補償預測。例如，使用幀間預測，使用運動補償預測從一個或多個參考圖片中的圖像資料（按輸出順序在當前圖片之前或之後）預測每個 PU。例如，可以在 CU 級別，做出是使用圖片間預測還是圖片內預測來對圖片區域進行譯碼的決定。

【0105】 在一些例子中，向圖片的一個或多個切片分配切片類型。切片類型可以包括 I 切片、P 切片和 B 切片。I 切片（幀內，可獨立解碼）是僅通過幀內預測進行編碼的圖片切片，因此可獨立解碼，這是因為 I 切片僅需要幀內的資料來預測切片的任何預測單元或預測塊。P 切片（單向預測幀）是可以用幀內預測和用單向幀間預測進行譯碼的圖片的切片。P 切片內的每個預測單元或預測塊都使用幀內預測或幀間預測進行譯碼。當應用幀間預測時，通過一個參考圖片來預測預測單元或預測塊，並且因此參考樣本來自一幀的一個參考區域。B 切片（雙向預測幀）是可以使用幀內預測和幀間預測（例如，雙向預測或單向預測）進行譯碼的圖片切片。可以從兩個參考圖片，雙向預測 B 切片的預測單元或預測塊。每個圖片可以貢獻一個參考區域，並且可以對兩個參考區域的樣本集進行加權（例如，具有相等的權重或具有不同的權重）以產生雙向預測塊的預測訊號。如上面所解釋的，對一個圖片的切片進行獨立地譯碼。在一些情況下，可以將一個圖片只譯碼為一個切片。

【0106】 在使用幀內和/或幀間預測執行預測之後，編碼設備可以執行變換和量化。例如，在預測之後，編碼設備可以計算對應於 PU 的殘差值。殘差值可

以包括正在譯碼的當前像素塊（PU）和用於預測當前塊的預測塊（例如，當前塊的預測版本）之間的像素差值。例如，在生成預測塊（例如，使用幀間預測或幀內預測）之後，編碼設備可以通過從當前塊中減去由預測單元產生的預測塊來生成殘差塊。殘差塊包括量化當前塊的像素值和預測塊的像素值之間的差的一組像素差值。在一些例子中，可以以二維塊格式（例如，二維矩陣或像素值陣列）來表示殘差塊。在這樣的例子中，殘差塊是像素值的二維表示。

【0107】 在執行預測之後可能剩餘的任何殘差資料使用塊變換進行變換，該塊變換可以基於離散余弦變換、離散正弦變換、整數變換、小波變換、其它適當的變換函數或者其任何組合。在一些情況下，一個或多個塊變換（例如，大小 32×32 、 16×16 、 8×8 、 4×4 或其它適當的大小）可以應用於每個 CU 中的殘差資料。在一些實施例中，TU 可以用於由編碼設備實現的變換和量化過程。具有一個或多個 PU 的給定 CU 還可以包括一個或多個 TU。如下文進一步詳細描述的，可以使用塊變換將殘差值變換為變換係數，然後可以使用 TU 來量化和掃描，以產生用於熵譯碼的經序列化的變換係數。

【0108】 編碼設備可以執行變換係數的量化。量化通過量化變換係數來提供進一步的壓縮，以減少用於表示這些係數的資料量。例如，量化可以減少與一些或所有係數相關聯的位深度。在一個例子中，可以在量化期間，將具有 n 位值的係數向下舍入為 m 位值，其中 n 大於 m 。

【0109】 一旦執行了量化，編碼的視頻位元流就包括量化的變換係數、預測資訊（例如，預測模式、運動向量、塊向量等）、分區資訊和任何其它適當的資料（例如，其它語法資料）。然後，編碼設備可以對編碼的視頻位元流的不同元素進行熵編碼。在一些例子中，編碼設備可以利用預定義的掃描順序來掃描量

化的變換係數，以產生可以進行熵編碼的序列化向量。在一些例子中，編碼設備可以執行自適應掃描。在掃描經量化的變換係數以形成向量（例如，一維向量）之後，編碼設備可以對該向量進行熵編碼。例如，編碼設備可以使用上下文自適應可變長度譯碼、上下文自適應二進制算術譯碼、基於語法的上下文自適應二進制算術譯碼、機率區間分割熵譯碼或者另一種適當的熵編碼技術。

【0110】 編碼設備可以儲存編碼的視頻位元流，和/或可以通過通訊鏈路將編碼的視頻位元流資料發送到接收設備，接收設備可以包括解碼設備。解碼設備可以通過對構成編碼的視頻資料的一個或多個編碼視頻序列的元素進行熵解碼（例如，使用熵解碼器）和提取，來解碼經編碼的視頻位元流資料。然後，解碼設備可以對編碼的視頻位元流資料執行重新縮放和逆變換。然後，將殘差資料傳遞到解碼設備的預測階段。解碼設備然後使用幀內預測、幀間預測、IBC 和/或其它類型的預測，來預測像素塊（例如，PU）。在一些例子中，將預測添加到逆變換的輸出（殘差資料）。解碼設備可以將解碼的視頻輸出到視頻目的地設備，該視頻目的地設備可以包括用於將解碼的視頻資料顯示給內容消費者的顯示器或其它輸出設備。

【0111】 由各種視頻譯碼標準定義的視頻譯碼系統和技術（例如，上面描述的 HEVC 視頻譯碼技術）能夠保留原始視頻內容中的大部分資訊，並且可以基於訊號處理和資訊論概念進行先驗定義。然而，雖然各種壓縮演算法可能普遍適用（例如，適用於任何類型的視頻內容），但這些視頻譯碼技術可能缺乏在基於網路的環境或其它類型的環境中有益的某些特性。例如，視頻譯碼系統應該能夠實現低延遲操作。然而，根據現有視頻譯碼標準操作的視頻譯碼系統和相關技術，可以以幾幀為單位（有時稱為圖片組或 GoP）對視頻進行編碼和傳輸，並且

可能需要等待一批視頻幀（例如，GoP）的積累以進行傳輸。與使用逐幀譯碼和傳輸所能實現的延遲相比，這種基於批處理的視頻譯碼導致更高的延遲。

【0112】 在一些情況下，可以使用基於機器學習的 P 幀譯碼系統（例如，其包括一個或多個神經網路）來執行單向預測。圖 4 是示出基於機器學習的 P 幀譯碼系統 400 的例子的圖。P 幀譯碼系統 400 的組件包括運動壓縮引擎 402、運動補償引擎 404 和殘差壓縮引擎 406。P 幀譯碼系統 400 的組件可以包括一個或多個神經網路，這些神經網路被配置為使用運動壓縮引擎 402、運動補償引擎 404 和殘差壓縮引擎 406 來執行單向幀間預測（uni-prediction）。

【0113】 例如，運動壓縮引擎 402 可以獲得輸入幀 401 和參考幀 403。運動壓縮引擎 402 可以執行基於機器學習的運動壓縮，以確定輸入幀 401 和參考幀之間的運動預測 403。運動壓縮引擎 402 可以將預測的運動資訊提供給運動補償引擎 404。運動補償引擎 404 還可以接收參考幀 403 作為輸入。運動補償引擎 404 可以使用運動預測來修改參考幀的像素，並生成輸入幀 403 的預測。例如，運動補償引擎 404 可以根據包含在運動預測中的運動向量來移動參考幀的像素。殘差壓縮引擎 406 可以生成表示預測和輸入幀 401 之間的差異的預測殘差。然後，可以將預測殘差與預測進行組合，以生成重建的輸入幀。可以輸出重建的輸入幀以用於儲存、顯示等。

【0114】 在一些情況下，可以使用基於機器學習的 B 幀譯碼系統來執行雙向預測。圖 5A 是示出 B 幀譯碼系統 500 的例子的圖。B 幀譯碼系統 500 的組件包括運動壓縮引擎 502、運動補償引擎 504 和殘差壓縮引擎 506。B 幀譯碼系統 500 可以包括一個或多個神經網路，這些神經網路被配置為使用運動壓縮引擎 502、運動補償引擎 504 和殘差壓縮引擎 506 來執行雙向幀間預測（雙向預測）。

雙向預測是基於兩個參考幀(在圖 5A 中表示為參考幀 0(Ref₀)和參考幀 1(Ref₁))而不是單個參考幀來執行。

【0115】 例如，可以將輸入幀 501 以及參考幀 Ref₀和 Ref₁作為輸入提供給運動壓縮引擎 502。運動壓縮引擎 502 可以執行基於機器學習的運動壓縮，以確定輸入幀 501 和第一參考幀 Ref₀之間的運動的第一預測以及輸入幀 501 和第二參考幀 Ref₁之間的運動的第二預測。運動補償引擎 504 可以接收參考幀 Ref₀和 Ref₁以及來自運動壓縮引擎 502 的預測運動作為輸入。運動補償引擎 504 可以使用兩個運動預測來修改參考幀 Ref₀和 Ref₁的像素並生成輸入幀 501 的預測。例如，運動補償引擎 504 可以根據第一運動預測來移動第一參考幀 Ref₀的像素，並且可以根據第二運動預測來移動第二參考幀 Ref₁的像素。運動補償引擎 504 可以合併這兩個預測以生成用於輸入幀 501 的合併預測。

【0116】 殘差壓縮引擎 506 可以生成表示合併預測和輸入幀 501 之間的差異的預測殘差。然後，可以將預測殘差與合併預測進行組合以生成重建的輸入幀，可以輸出該重建的輸入幀以用於儲存或顯示。

【0117】 圖 5B 是示出 B 幀譯碼系統 510 的另一示例的圖。B 幀譯碼系統 510 的組件包括幀插值引擎 512 和殘差壓縮引擎 516。B 幀譯碼系統 500 的組件可以包括一個或多個神經網路，這些神經網路被配置為使用幀插值引擎 512 和殘差壓縮引擎 516 執行雙向預測。

【0118】 類似於圖 5A 的 B 幀譯碼系統 500，B 幀譯碼系統 510 可以基於兩個參考幀(表示為參考幀 0(Ref₀)和參考幀 1(Ref₁))執行雙向預測。然而，插值引擎 512 可以通過在參考幀 Ref₀和參考幀 Ref₁之間執行插值來生成輸入幀 511 的預測。然後可以將預測輸出到殘差壓縮引擎 516。殘差壓縮引擎 506 可以

生成表示預測和輸入幀 511 之間的差異的預測殘差。B 幀譯碼系統 510 可以將預測的殘差與合併的預測進行組合以生成重建的輸入幀。B 幀譯碼系統 510 可以輸出重建的輸入幀以用於儲存、顯示等。

【0119】圖 5A 和圖 5B 的上述 B 幀譯碼系統在某些情況下可能不是最理想的。例如，圖 5A 的 B 幀譯碼系統 500 基於兩個參考幀 Ref_0 和 Ref_1 以及輸入幀 511，執行雙向運動壓縮以生成兩個運動預測。B 幀譯碼系統 500 基於兩個運動預測修改參考幀 Ref_0 和 Ref_1 ，然後合併修改後的幀以生成針對輸入幀 511 的預測。這樣的過程涉及訓練和推理期間的許多計算（在訓練機器學習系統並用於處理輸入視頻資料之後）。此外，（圖 5A 或圖 5B 的）單獨 B 幀譯碼系統和 P 幀譯碼系統（圖 4）需要單獨的神經網路（具有不同的參數，例如權重、偏差等）。訓練圖 5A 和圖 5B 的機器學習系統 500 和 510 還需要大量的訓練資料（例如，每個預測需要兩個參考幀）。此外，圖 5B 的 B-幀譯碼系統 510 確定了僅在線性運動假設下準確的插值，並且該預測可能沒有與輸入幀對準。在一些情況下，B 幀譯碼系統 510 可以使用非線性幀內插器；但是，在這種情況下仍可能發生錯位。B 幀譯碼系統 510 僅依靠殘差來補償未對準。

【0120】如上所述，本文描述了用於使用 P 幀譯碼系統和內插參考幀來執行學習的 B 幀譯碼的系統和技術。本文描述的系統和技術可以使用從參考幀內插（例如，使用諸如光流資訊的運動資訊進行內插）的參考幀，來高效和智慧型地預測雙向預測幀（B 幀）。這些系統和技術可以實現具有較低資源需求（例如，較低的計算和頻寬需求）的高效、高品質視頻譯碼。例如，當神經網路用於 P 幀譯碼系統時，P 幀譯碼系統需要一組神經網路參數（例如，權重、偏差等）來執行雙向預測（B 幀）並執行單向預測（P 幀）。此外，由於僅使用單個參考幀（內

插參考幀)來執行雙向預測,因此每個 B 幀的雙向預測僅需要單個運動資訊集。在一些情況下,與訓練圖 5B 的 B 幀譯碼系統 500 所需的訓練資料相比,訓練機器學習系統(例如,神經網路)可能還需要更少的訓練資料。

【0121】 B 幀譯碼系統 600 還提供改進的率失真性能,例如與圖 5A 的 B 幀譯碼系統 500 和圖 5B 的 B 幀譯碼系統 510 相比。例如,對於給定的輸入幀,當來自過去(例如, Ref_0)和未來(例如, Ref_1)的參考幀都可用時,例如在線性運動假設下, B 幀譯碼系統 600 可以確定通過對兩個參考幀進行線性插值,粗略預測輸入幀 611。該預測消耗零位,這是因為兩個參考幀已經可用並且不需要運動場或其它資訊來執行插值。圖 5A 的 B 幀譯碼系統 500 使用雙向運動(例如,流動或變形),其不利用這種資訊並且發送關於兩個參考幀的運動資訊(例如,光流)(例如,好像解碼器/接收器不知道物體運動)。如上所述,關於圖 5B 的 B 幀譯碼系統 510,幀插值引擎 512 的插值結果僅在線性運動假設下是準確的。在這樣的系統中,預測可能沒有與輸入幀對準。即使在採用非線性幀內插器時,仍可能發生未對準。在這些情況下,譯碼系統僅依靠殘差來補償未對準。結果,與圖 6 的 B 幀譯碼系統 600 相比,這種譯碼系統的譯碼效率會嚴重降低,圖 6 的 B 幀譯碼系統 600 可以使用廉價的邊資訊(例如,參考幀之間的插值)來糾正未對準,然後可以使用殘差來確定重建的輸入幀。

【0122】 圖 6 是示出被配置為執行學習的 B 幀譯碼的 B 幀譯碼系統 600 的示例的圖。B 幀譯碼系統 600 包括用於處理視頻資料的各種組件。如圖所示, B 幀譯碼系統 600 的組件包括幀插值引擎 622 和 P 幀譯碼系統 624。使用幀插值引擎 622 和 P 幀譯碼系統 624, B 幀譯碼系統 600 可以執行輸入幀 611 的雙向幀間預測(bi-prediction)以生成輸入幀 611 的重建版本(或解碼版本),其稱為重建

的輸入幀 615。幀插值引擎 622 可以基於對參考幀 0 (Ref_0) 和參考幀 1 (Ref_1) 進行處理，來生成內插參考幀 613。幀插值引擎 622 可以利用任何類型的插值技術。在一些例子中，幀插值引擎 622 可以使用與參考幀 Ref_0 和 Ref_1 相關聯的運動資訊來生成內插參考幀 613。例如，如參考圖 7C 所描述的，幀插值引擎 622 可以基於為參考幀 Ref_0 和參考幀 Ref_1 確定的雙向運動資訊（例如，前向光流和後向光流）來確定內插參考幀 615。在一些情況下，可以儲存內插參考幀 615，例如用作一個或多個其它輸入幀的參考幀。

【0123】 B 幀譯碼系統 600 可以使用 P 幀譯碼系統 624 來生成重建的輸入幀 615（作為 B 幀）。例如，如圖 6 中所示，P 幀譯碼系統 624 可以使用內插參考幀 613 作為輸入來生成重建的輸入幀 615。P 幀譯碼系統 624 可以使用任何類型的 P 幀譯碼技術。在一些例子中，P 幀譯碼系統 624 可以包括至少一個基於機器學習的視頻編碼器和至少一個基於機器學習的視頻解碼器，以處理內插參考幀 613 和輸入幀 611 來生成重建的輸入幀 615。在一些實施方式中，如下文關於圖 7A 和圖 7B 所描述的，P 幀譯碼系統 624 可以包括具有編碼器神經網路和解碼器神經網路的運動神經網路系統。該運動神經網路系統可以包括任何類型的神經網路結構，例如自動編碼器、卷積神經網路（CNN）和/或其它神經網路結構。運動神經網路系統的編碼器神經網路可以生成表示針對輸入幀 611 的運動資料的潛在資料。運動神經網路系統的解碼器神經網路可以從潛在資料重建運動，以生成針對輸入幀 611 的預測運動。P 幀譯碼系統 624 可以使用預測的運動，來變形內插參考幀 613 以生成輸入幀 611 的預測。

【0124】 P 幀譯碼系統 624 還可以包括殘差神經網路系統，該殘差神經網路系統包括編碼器神經網路和解碼器神經網路。殘差神經網路系統可以包括任

何類型的神經網路結構，例如自動編碼器、CNN 和/或其它神經網路結構。P 幀譯碼系統 624 可以例如通過確定變形的內插參考幀 613（使用預測的運動變形）和輸入幀 611 之間的差異，為輸入幀 611 生成殘差。殘差神經網路系統的編碼器神經網路可以對殘差進行處理，以生成表示殘差的殘差潛在資料。殘差神經網路系統的解碼器神經網路可以從殘差潛在資料重建殘差，以生成針對輸入幀 611 的預測殘差。P 幀譯碼系統 624 可以通過將預測殘差添加到變形的內插參考幀 613，來生成重建的輸入幀 615。

【0125】 在一些例子中，P 幀譯碼系統 624 可以用於執行雙向預測（以生成 B 幀），並且還可以用於執行單向預測（以生成 P 幀）。例如，P 幀譯碼系統 624 可以用於在確定對輸入幀執行單向預測譯碼模式時生成 P 幀，並且可以用於在確定對輸入幀執行雙向預測譯碼模式時生成 B 幀。參考圖 9-圖 12 描述了關於為某些幀選擇不同譯碼模式（例如，幀內預測、單向預測或雙向預測）的細節。當為輸入幀選擇單向預測模式時，P 幀譯碼系統 624 可以處理實際參考幀（例如，參考幀 Ref_0 或參考幀 Ref_1 ），以生成重建幀作為 P-幀。下面參考圖 7A 描述了關於使用 P 幀譯碼系統 624 進行單向預測的細節。當為輸入幀選擇雙向預測模式時，P 幀譯碼系統 624 可以處理內插參考幀（例如，內插參考幀 613）以生成重建幀作為 B 幀。下面參考圖 7B 描述了關於使用 P 幀譯碼系統 624 進行雙向預測的細節。

【0126】 雖然將 B 幀譯碼系統 600 示出為包括某些組件，但本領域通常知識者應當理解，B 幀譯碼系統 600 可以包括比圖 6 中所示的組件更多的組件。在一些情況下，B 幀譯碼系統 600 的其它組件可以包括軟體、硬體、或軟體和硬體的一種或多種組合。例如，在一些情況下，B 幀譯碼系統 600 可以包括一個或多

個顯示設備、一個或多個其它處理引擎、一個或多個圖像感測器（例如，相機）、一個或多個其它感測器（例如，一個或多個慣性量測單元（IMU）、雷達、光探測和測距（LIDAR）感測器、音頻感測器等）、和/或用於無線通訊的一個或多個無線收發器（或單獨的無線接收器和發射器），該無線通訊例如為蜂窩網路通訊（例如，使用長期演進（LTE）協議、5G/新無線電（NR）協議或其它蜂窩協議）、無線區域網（WLAN）通訊（例如，使用 802.11 Wi-Fi 協議）、藍牙或其它短程通訊、其任意組合和/或其它通訊。本領域通常知識者應當理解，B 幀譯碼系統 600 可以包括圖 6 中沒有示出的一個或多個其它軟體和/或硬體組件。

【0127】 在一些實施方式中，B 幀譯碼系統 600 的其它組件可以包括電子電路或其它電子硬體和/或可以使用電子電路或其它電子硬體來實現，其可以包括一個或多個可程式化電子電路（例如，數位訊號處理器（DSP））、微處理器、微控制器、圖形處理單元（GPU）、中央處理單元（CPU）、它們的任何組合和/或其它適當的電子電路），和/或可以包括計算機軟體、韌體或其任意組合和/或使用計算機軟體、韌體或其任意組合來執行本文所描述的各种操作。軟體和/或韌體可以包括儲存在計算機可讀儲存媒體上，並且可由實現 B 幀譯碼系統 600 的電子設備的一個或多個處理器執行的一個或多個指令。

【0128】 B 幀譯碼系統 600 可以包括計算設備或者是計算設備的一部分。例如，該計算設備可以包括移動或固定電話手機（例如，智慧型手機、蜂窩電話等）、擴展實境設備（例如，虛擬實境（VR）設備、擴增實境（AR）設備和/或混合實境設備）、圖像擷取設備（例如，相機）、車輛、顯示設備、伺服器計算機（例如，與車輛計算系統通訊）、桌上型計算機、膝上型計算機或筆記型計算

機、平板計算機、機頂盒、電視機、數位媒體播放器、視頻流設備或者任何其它適當的計算設備，或者是這些設備的一部分。

【0129】圖 7A 是說明可以用於執行 P 幀的單向譯碼和/或 B 幀的雙向譯碼的 P 幀譯碼系統 700 的示例的圖。P 幀譯碼系統 700 是圖 6 的 P 幀譯碼系統 624 的說明性示例。P 幀譯碼系統 700 包括運動預測系統 702、變形引擎 704 和殘差預測系統 706。運動預測系統 702 和殘差預測系統 706 可以包括任何類型的機器學習系統（例如，使用一個或多個神經網路）。在一個說明性示例中，運動預測系統 702 包括運動自動編碼器，並且殘差預測系統 706 包括殘差自動編碼器。雖然將圖 7A 的 P 幀譯碼系統 700 示出為包括某些組件，但本領域通常知識者應當理解，P 幀譯碼系統 700 可以包括比圖 7A 中所示的組件更少或更多的組件。

【0130】對於給定時間 t ，P 幀譯碼系統 700 接收輸入幀（表示為幀 X_t ）和參考幀 $\hat{X}_{ref}$ 。參考幀 $\hat{X}_{ref}$ 是在時間 t 之前生成的先前重建的幀（如帽子符號「 $\hat{\cdot}$ 」所示）。例如，輸入幀 X_t 是時間 t 的當前幀，並且參考幀 $\hat{X}_{ref}$ 可以是在時間 $t-n$ 重建或生成的過去幀（例如，在時間 t 在輸入幀 X_t 之前的時間或順序上），其中 n 是大於或等於 1 的整數值。在一個說明性示例中，參考幀 $\hat{X}_{ref}$ 可以是在時間上或順序上緊接在輸入幀 X_t 之前的幀，在這種情況下參考幀 $\hat{X}_{ref}$ 可以表示為 $\hat{X}_{t-1}$ 。在一些情況下，從譯碼系統的解碼圖像緩衝器（DPB）接收參考幀 $\hat{X}_{ref}$ 。

【0131】在一些情況下，輸入幀 X_t 可以是 P 幀，參考幀 $\hat{X}_{ref}$ 可以是 I 幀、P 幀或 B 幀。例如，參考幀 $\hat{X}_{ref}$ 可以由 I 幀譯碼系統重建或生成（例如，I 幀譯碼系統可以是包括 P 幀譯碼系統 700 的設備的一部分，或者是與包括 P 幀譯碼系統 700 的設備不同的設備）、由 P 幀譯碼系統 700 重建或生成（或者不同於包括 P 幀譯碼系統 700 的設備的設備的 P 幀譯碼系統）、或者由 B 幀譯碼系統來重建

或生成（例如，圖 6 的 B 幀譯碼系統 600、圖 7B 的 B 幀譯碼系統 720、或者與包括 P 幀譯碼系統 700 的設備不同的設備的 B 幀譯碼系統）。在一些情況下，可以根據圖 9 或圖 10 中所示的一組圖片（GoP）結構和/或根據順序方案（例如，圖 11 中所示的順序方案）或分層方案（例如，圖 12 中所示的分層方案）、或基於任何其它方案，來選擇參考幀 $\hat{X}_{ref}$ 。

【0132】運動預測系統 702 接收參考幀 $\hat{X}_{ref}$ 和輸入幀 X_t 。運動預測系統 702 可以確定參考幀 $\hat{X}_{ref}$ 的像素和輸入幀 X_t 的像素之間的運動（例如，由向量表示，諸如光流運動向量）。然後，運動預測系統 702 可以將運動進行編碼，並且在一些情況下將運動解碼為輸入幀 X_t 的預測運動 $\hat{f}_t$ 。例如，運動預測系統 702 的編碼器網路 705 可以確定當前幀 X_t 和參考幀 $\hat{X}_{ref}$ 之間的運動。編碼器網路 705 可以將運動編碼為潛在表示（表示為潛在 z_m ）。在一些例子中，編碼器網路 705 可以將運動映射到可以用作潛在 z_m 的潛在碼。在一些情況下，編碼器網路 705 可以通過對潛在碼執行熵譯碼，將潛在 z_m 轉換為位元流。在一些例子中，編碼器網路 705 可以量化潛在 z_m （例如，在對潛在碼執行熵譯碼之前）。量化的潛在 z_m 可以包括殘差 z_m 的量化表示。在一些情況下，潛在 z_m 可以包括表示量化碼的神經網路資料（例如，神經網路節點的啟動圖或特徵圖）。編碼器網路 705 可以儲存潛在 z_m ，將潛在 z_m 發送到運動預測系統 702 的解碼器網路 703，和/或可以將潛在 z_m 發送到可以解碼潛在 z_m 的另一個設備或系統。在接收到潛在 z_m 時，解碼器網路 703 可以對潛在 z_m 進行解碼（例如，逆熵譯碼、去量化和/或重建）以生成參考幀 $\hat{X}_{ref}$ 的像素和輸入幀 X_t 的像素之間的預測運動 $\hat{f}_t$ 。編碼器網路 705 和解碼器網路 703 可以使用訓練資料（例如，訓練圖像或幀）和一個或多個損失函數進行訓練和優化，如下所述。

【0133】 在一些例子中，預測的運動 $\hat{f}_t$ 可以包括光流資料（例如，包括一個或多個位移向量的光流圖）、動態卷積資料（例如，用於資料卷積的矩陣或核）或基於塊的運動資料（例如，每個塊的運動向量），如下面關於圖 8 所描述的。例如，在一些情況下，預測的運動 $\hat{f}_t$ 可以包括光流圖。在一個說明性示例中，光流圖可以包括輸入幀 X_t 的每個像素的運動向量（例如，第一像素的第一運動向量、第二像素的第二運動向量，等等）。運動向量表示當前幀 X_t 中的像素相對於參考幀 $\hat{X}_{ref}$ 中的相應像素的運動。

【0134】 P 幀譯碼系統的變形引擎 704 可以獲得運動預測系統 702 輸出的運動預測 $\hat{f}_t$ 。例如，變形引擎 704 可以從儲存中檢索運動預測 $\hat{f}_t$ ，或者可以從運動預測系統 702 接收運動預測 $\hat{f}_t$ 。變形引擎 704 可以使用運動預測 $\hat{f}_t$ 來變形（例如，通過執行運動補償）參考幀 $\hat{X}_{ref}$ 的像素，導致生成變形幀 $\tilde{X}_t$ 。例如，參考幀 $\hat{X}_{ref}$ 的像素可以移動到新的位置或者根據重建的運動估計資料 $\hat{f}_t$ 進行修改。在使用光流來確定參考幀 $\hat{X}_{ref}$ 的像素和輸入幀 X_t 的像素之間的運動的一個說明性示例中，變形引擎 704 可以將參考幀 $\hat{X}_{ref}$ 的像素移動光流圖的運動向量所指示的量，產生變形的幀 $\tilde{X}_t$ 。如上所述，其它基於運動的技術也可以用於確定運動預測 $\hat{f}_t$ ，以變形參考幀 $\hat{X}_{ref}$ （例如，動態卷積資料、基於塊的運動資料等）。

【0135】 如上所述，為了生成變形的幀 $\tilde{X}_t$ ，P 幀譯碼系統 700 可以通過預測 X_t 和 $\hat{X}_{ref}$ 之間的運動 $\hat{f}_t$ （例如，光流或其它運動）並且使用運動預測 $\hat{f}_t$ 變形參考幀 $\hat{X}_{ref}$ ，來執行運動補償。然而，基於運動預測 $\hat{f}_t$ 的幀預測 $\tilde{X}_t$ 可能不夠準確，無法將輸入幀 X_t 表示為重建的幀 $\hat{X}_t$ 。例如，輸入幀 X_t 描繪的場景中可能存在一個或多個遮擋區域、過度照明、缺乏照明、或導致幀預測 $\tilde{X}_t$ 不夠準確而無法用作重建幀 $\hat{X}_t$ 的其它影響。

【0136】殘差預測系統 706 可以用於校正 $\tilde{X}_t$ 的預測並生成表示幀 X_t 的重建幀 $\hat{X}_t$ 。如圖 7A 中所示，可以通過從輸入幀 X_t 中減去（使用減法運算 708 確定）預測幀 $\tilde{X}_t$ 來確定殘差 r_t 。例如，一旦變形引擎 704 確定了預測幀 $\tilde{X}_t$ ，P 幀譯碼系統 700 可以通過確定預測幀 $\tilde{X}_t$ 和輸入幀 X_t 之間的差異（使用減法運算 708），來確定殘差 r_t 。殘差預測系統 706 的編碼器網路 707 可以將殘差 r_t 編碼成表示殘差 r_t 的潛在 z_r 。在一些例子中，編碼器網路 707 可以將殘差 r_t 映射到可以用作潛在 z_r 的潛在碼。在一些情況下，編碼器網路 707 可以通過對潛在碼執行熵譯碼，將潛在 z_r 轉換為位元流。在一些例子中，編碼器網路 707 還可以量化潛在 z_r （例如，在執行熵譯碼之前）。量化的潛在 z_r 可以包括殘差 r_t 的量化表示。在一些情況下，潛在 z_r 可以包括表示量化碼的神經網路資料（例如，神經網路節點的啟動圖或特徵圖）。編碼器網路 707 可以儲存潛在 z_r ，將潛在 z_r 發送到殘差預測系統 706 的解碼器網路 709，和/或可以將潛在 z_r 發送到能夠解碼潛在 z_r 的另一個設備或系統。在接收到潛在 z_r 時，解碼器網路 709 可以對潛在 z_r 進行解碼（例如，逆熵譯碼、去量化和/或重建）以生成預測殘差 $\hat{r}_t$ 。編碼器網路 707 和解碼器網路 709 可以使用訓練資料（例如，訓練圖像或幀）和一個或多個損失函數進行訓練和優化，如下所述。

【0137】預測的殘差 $\hat{r}_t$ 可以與來自變形引擎 704 的幀預測 $\tilde{X}_t$ 一起使用，以生成表示時間 t 處的輸入幀 X_t 的重建幀 $\hat{X}_t$ 。例如，P 幀譯碼系統 700 可以將預測殘差 $\hat{r}_t$ 添加（使用加法運算 710）到幀預測 $\tilde{X}_t$ 以生成重建幀 $\hat{X}_t$ 。在一些情況下，殘差預測系統 706 的解碼器網路 709 可以將預測的殘差 $\hat{r}_t$ 添加到幀預測 $\tilde{X}_t$ 。重建幀 $\hat{X}_t$ 也可以稱為解碼幀。可以輸出重建的當前幀 $\hat{X}_t$ 用於儲存（例如，儲存在解碼的圖片緩衝區（DPB）或其它儲存設備中）、傳輸、顯示、用於進一步處理（例

如，作為進一步的幀間預測中的參考幀，用於後處理等）、和/或用於任何其它用途。

【0138】 如上所述，在一些情況下，P 幀譯碼系統 700 可以將表示運動的潛在資料（潛在 z_m ）和表示位元流中的殘差（潛在 z_r ）的潛在資料發送到另一個設備以進行解碼。例如，該另一個設備可以包括能夠從潛在 z_m 中解碼預測運動 $\hat{f}_t$ ，從潛在 z_r 中解碼預測殘差 $\hat{r}_t$ ，並且可以基於預測運動 $\hat{f}_t$ 和預測殘差 $\hat{r}_t$ 生成重建幀 $\hat{X}_t$ （例如，通過將預測殘差 $\hat{r}_t$ 添加到使用預測運動 $\hat{f}_t$ 生成的幀預測 $\tilde{X}_t$ ，如上所述）的視頻解碼器（例如，運動預測系統 702 和殘差預測系統 706，如上所述，或者被配置為解碼潛在 z_m 和 z_r 的另一個解碼器）。

【0139】 在一些例子中，可以使用訓練資料和一個或多個損失函數來訓練和優化運動預測系統 702 和/或殘差預測系統 706。在一些情況下，可以以端到端方式訓練運動預測系統 702 和/或殘差預測系統 706（其中，在相同的訓練過程中訓練所有的神經網路組件）。在一些方面，訓練資料可以包括訓練圖像或幀。在一些情況下，損失函數 $Loss$ 可以用於基於運動預測系統 702 和/或殘差預測系統 706 處理訓練圖像或幀的訓練。在一個例子中，可以將 $Loss$ 定義為 $Loss = D + \beta R$ ，其中 D 是幀（例如，輸入幀 X_t ）與其重建幀（例如， $D(\text{frame } X_t, \text{frame } \hat{X}_t)$ ）之間的失真， β 是控制位元率（每像素位元）的超參數， R 是用於將殘差（例如，殘差 r_t ）轉換為壓縮位元流（例如，潛在 z_r ）的位元數。在一些例子中，可以基於峰值訊雜比（PSNR）、結構相似性指數量測（SSIM）、多尺度 SSIM（MS-SSIM）等來計算失真 D 。使用訓練資料和損失函數，可以調整運動預測系統 702 和/或殘差預測系統 706 的參數（例如，權重、偏差等），直到 P 幀譯碼系統 700 達到期望的譯碼結果為止。

【0140】圖 7B 是說明可以用於使用圖 7A 的 P-幀譯碼系統 700 執行雙向譯碼的 B-幀譯碼系統 720 的示例的圖。B 幀譯碼系統 720 還包括幀插值引擎 722。雖然將 B 幀譯碼系統 720 示出為包括某些組件，但本領域通常知識者應當理解，B 幀譯碼系統 720 可以包括比圖 7B 中所示的更少或更多的組件。

【0141】插值引擎 722 可以獲得兩個參考幀，分別標記為 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 。如圖所示，P 幀譯碼系統 700 在時間 t 獲得輸入幀 X_t ，類似於上面關於圖 7A 描述的那些。輸入幀 X_t 可以是在時間上或順序上在參考幀 $\hat{X}_{ref_0}$ 之後並在參考幀 $\hat{X}_{ref_1}$ 之前的幀。在一些例子中，輸入幀 X_t 可以是時間 t 的當前幀，參考幀 $\hat{X}_{ref_0}$ 可以是在時間 $t-n$ 的過去幀（例如，在時間上或順序上在時間 t 的輸入幀 X_t 之前），並且參考幀 $\hat{X}_{ref_1}$ 可以是在時間 $t-n$ 的未來幀（例如，在時間上或順序上在輸入幀 X_t 之後），其中 n 和 m 可以表示相同或不同的值。在一個說明性示例中，參考幀 $\hat{X}_{ref_0}$ 可以是在時間 $t-1$ 的緊接輸入幀 X_t 之前的幀，並且參考幀 $\hat{X}_{ref_1}$ 可以是在時間 $t+1$ 的緊接在輸入幀 X_t 之後的幀。

【0142】在一些情況下，輸入幀 X_t 可以是 B 幀，並且參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 中的每一者可以是 I 幀、P 幀、或 B 幀。例如，參考幀 $\hat{X}_{ref_0}$ 可以是 I 幀、P 幀或 B 幀，參考幀 $\hat{X}_{ref_1}$ 可以是 I 幀、P 幀或 B 幀。在一些情況下，參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 可以是相同類型的幀，例如 B 幀、P 幀或 I 幀。在其它情況下，參考幀 $\hat{X}_{ref_0}$ 和參考幀 $\hat{X}_{ref_1}$ 可以是不同類型的幀。例如，參考幀 $\hat{X}_{ref_0}$ 可以是 I 幀，參考幀 $\hat{X}_{ref_1}$ 可以是 P 幀。在一些情況下，可以根據圖 9 或圖 10 中所示的一組圖片（GoP）結構和/或根據順序方案（例如，圖 11 中所示的順序方案）或分層方案（例如，圖 12 中所示的分層方案）、或基於任何其它方案，來選擇參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 。

【0143】 插值引擎 722 可以對兩個參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 進行插值操作以生成內插參考幀 $\hat{X}_{ref_t}$ 。例如，在一些情況下，執行插值以生成可以用作輸入幀 X_t 的參考的整個參考幀（例如，不僅是幀的部分，例如像素塊）。可以對內插參考幀 $\hat{X}_{ref_t}$ 進行儲存和/或輸出到 P-幀譯碼系統 700。幀插值引擎 722 可以使用參考幀 $\hat{X}_{ref_0}$ 和參考幀 $\hat{X}_{ref_1}$ ，執行任何類型的插值以生成內插參考幀 $\hat{X}_{ref_t}$ 。可以由幀插值引擎 722 執行的插值技術的示例可以包括例如但不限於：基於光流的幀插值、移動梯度方法、基於相位的方法、基於內核的方法和/或任何其它幀插值技術。在一個說明性示例中，為了生成內插參考幀 $\hat{X}_{ref_t}$ ，幀插值引擎 722 可以對於參考幀 $\hat{X}_{ref_0}$ 和參考幀 $\hat{X}_{ref_1}$ ，計算運動估計資訊（例如，光流資訊，例如光流圖，也稱為運動向量圖）。幀插值引擎 722 可以對運動估計資訊（例如，流圖）進行插值以確定與內插參考幀 $\hat{X}_{ref_t}$ 相關聯的時間 t 的運動估計（例如，光流）。幀插值引擎 725 可使用內插運動估計資訊（例如，內插光流）和參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 來產生內插參考幀 $\hat{X}_{ref_t}$ （例如，至少部分通過變形參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ ）。

【0144】 在一些例子中，幀插值引擎 722 可以逐個像素地確定運動估計資訊（例如，光流資訊）。例如，幀插值引擎 722 可以將參考幀 $\hat{X}_{ref_0}$ （或參考幀 $\hat{X}_{ref_0}$ 中的一個像素子集）中的每個像素映射到參考幀 $\hat{X}_{ref_1}$ （或參考幀 $\hat{X}_{ref_1}$ 中的相應像素子集）中對應像素的位置，反之亦然（例如，在兩個方向上）。在一個例子中，每個像素的運動估計資訊（例如，光流資訊）可以包括用於指示像素在參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 之間的移動的向量（例如，像素從參考幀 $\hat{X}_{ref_0}$ 移動到參考幀 $\hat{X}_{ref_1}$ 和/或像素從參考幀 $\hat{X}_{ref_1}$ 移動到參考幀 $\hat{X}_{ref_0}$ ）。在一些例子中，可以基於參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 之間的光流向量來生成光流圖。該光流圖可以包括幀中每個像素的光流向量，其中每個向量指示像素在幀之間的移動。

【0145】 在一個說明性示例中，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），其顯示像素從第一幀到第二幀的移動。在一些情況下，幀插值引擎 722 可以計算參考幀 $\hat{X}_{ref_0}$ 的二維（2D）流圖（例如，指示水平和垂直位移）和參考幀 $\hat{X}_{ref_1}$ 的 2D 流圖。在一些方面，幀插值引擎 722 可以合併 2D 流圖以生成用於內插參考幀 $\hat{X}_{ref_t}$ 的三維（3D）流圖。該 3D 流圖可以包括分別指示水平和垂直位移的水平尺寸和垂直尺寸，以及包括標識水平位移和垂直位移應用到的幀的幀索引的幀尺寸。例如，如果幀尺寸標識參考幀 $\hat{X}_{ref_0}$ ，則水平和垂直尺寸可以指示參考幀 $\hat{X}_{ref_0}$ 中像素的水平和垂直位移（例如， $\hat{X}_{ref_0}$ 中的像素應當根據光流資訊而移動多少像素）。

【0146】 幀插值引擎 722 可以使用插值後的光流（例如，3D 流圖）將參考幀 $\hat{X}_{ref_0}$ 和參考幀 $\hat{X}_{ref_1}$ 變形到輸入幀 X_t 的特定時間步長（例如，時間 t ），並且可以融合變形的幀以生成內插參考幀 $\hat{X}_{ref_t}$ 。下面關於圖 7C 來描述幀插值引擎的這種示例實現的進一步細節。

【0147】 使用內插參考幀 $\hat{X}_{ref_t}$ 和輸入幀 X_t 作為輸入，P 幀譯碼系統 700 可以執行與上面關於圖 7A 描述的那些相同或相似的操作，來生成重建幀 $\hat{X}_t$ 。例如，運動預測系統 702 的編碼器網路 705 可以確定輸入幀 X_t 和內插參考幀 $\hat{X}_{ref_t}$ 之間的運動，並且可以將運動編碼到潛在 z_m 中。運動預測系統 702 的解碼器網路 703 可以對潛在 z_m 進行解碼，以生成預測的運動 $\hat{f}_t$ 。變形引擎 704 可以使用預測的運動 $\hat{f}_t$ 來修改內插參考幀 $\hat{X}_{ref_t}$ 的像素（例如，通過執行運動補償），以生成幀預測 $\tilde{X}_t$ ，如上面參照圖 7A 所描述的。殘差預測系統 706 的編碼器網路 707 可以獲得殘差 r_t （確定為幀預測 $\tilde{X}_t$ 與輸入幀 X_t 之間的差），並且可以將殘差 r_t 編碼到潛在 z_r 中。殘差預測系統 706 的解碼器網路 709 可以對潛在 z_r 進行解碼，以生成預測

殘差 $\hat{r}_t$ 。然後，B 幀譯碼系統 720 可以將幀預測 $\tilde{X}_t$ 添加到預測殘差 $\hat{r}_t$ 以生成重建幀 $\hat{X}_t$ 。

【0148】圖 7C 是說明可以由 B 幀譯碼系統使用的示例性幀插值引擎 725 的系統流程的圖。幀插值引擎 725 是圖 7B 的幀插值引擎 722 的說明性示例。在圖 7C 的例子中，幀插值引擎 725 包括光流估計層 730、光流預測層 732、2D 變形層 734、細化層 736 和 3D 變形層 738。

【0149】如圖 7C 中所示，參考幀 $\hat{X}_{ref_0}$ 和參考幀 $\hat{X}_{ref_1}$ 輸入到光流估計層 730。在一些例子中，光流估計層 730 可以包括一個或多個神經網路或其它機器學習系統。例如，可以使用一個或多個光流估計神經網路來實現光流估計層 730。光流估計層 730 可以使用參考幀 $\hat{X}_{ref_0}$ 和參考幀 $\hat{X}_{ref_1}$ 來估計針對參考幀 $\hat{X}_{ref_0}$ 和參考幀 $\hat{X}_{ref_1}$ 的雙向運動資訊 740（例如，前向光流和後向光流）。雙向運動資訊 740 可以包括從參考幀 $\hat{X}_{ref_0}$ 到參考幀 $\hat{X}_{ref_1}$ 的光流或光流圖（表示為 $f_{0 \rightarrow 1}$ ），以及從參考幀 $\hat{X}_{ref_1}$ 到參考幀 $\hat{X}_{ref_0}$ 的光流或光流圖（標記為 $f_{1 \rightarrow 0}$ ）。

【0150】例如，雙向運動資訊 740 可以包括光流圖 $f_{0 \rightarrow 1}$ ，其將參考幀 $\hat{X}_{ref_0}$ （或參考幀 $\hat{X}_{ref_0}$ 的一個像素子集）中的每個像素的位置映射到參考幀 $\hat{X}_{ref_1}$ （或參考幀 $\hat{X}_{ref_1}$ 中的相應像素子集）中相應像素的位置。雙向運動資訊 740 還可以包括光流圖 $f_{1 \rightarrow 0}$ ，其將參考幀 $\hat{X}_{ref_1}$ （或參考幀 $\hat{X}_{ref_1}$ 的一個像素子集）中的每個像素的位置映射到參考幀 $\hat{X}_{ref_0}$ （或參考幀 $\hat{X}_{ref_0}$ 中的相應像素子集）中相應像素的位置。在一些情況下，每個像素的運動資訊 740 可以包括指示像素在參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 之間的移動的向量。在一些例子中，可以基於參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 的像素之間的光流向量，來生成光流圖 $f_{0 \rightarrow 1}$ 和 $f_{1 \rightarrow 0}$ 。光流圖可以包括針對幀中每個像素的光流向量，其中每個向量指示像素在幀之間的移動。

【0151】 光流估計層 730 可以向光流預測層 732 提供雙向運動資訊 740 (例如, $f_{0 \rightarrow 1}$ 和 $f_{1 \rightarrow 0}$)。光流預測層 732 可以獲得指示時間間隔 0 到 1 中的時間 t 的時間資訊 742 (例如, $t \in (0,1)$), 其中 0 表示參考幀 $\hat{X}_{ref_0}$ 的時間步長, 1 表示參考幀 $\hat{X}_{ref_1}$ 的時間步長, 並且時間 t 對應於要內插的幀 (例如, 內插幀 $\hat{X}_{ref_t}$) 的時間步長, 其也對應於輸入幀 (例如, 圖 7B 的輸入幀 X_t) 的時間步長。光流預測層 732 可以使用雙向運動資訊 740 和時間資訊 742 來預測 2D 光流圖 744。在一些例子中, 光流估計層 730 可以針對參考幀 $\hat{X}_{ref_0}$ 和參考幀 $\hat{X}_{ref_1}$, 對雙向運動資訊 740 (例如, $f_{0 \rightarrow 1}$ 和 $f_{1 \rightarrow 0}$) 進行插值, 以確定時間 t 的光流。在一些情況下, 時間 t 的光流可以是 $f_{0 \rightarrow 1}$ 和 $f_{1 \rightarrow 0}$ 之間的線性插值。

【0152】 在一些例子中, 2D 光流圖 744 可以包括從時間 t 到參考幀 $\hat{X}_{ref_0}$ (例如, 到參考幀 $\hat{X}_{ref_0}$ 的時間步長 0) 的光流圖 (例如, $\hat{f}_{t \rightarrow 0}^{2D}$)、以及從時間 t 到參考幀 $\hat{X}_{ref_1}$ (例如, 到參考幀 $\hat{X}_{ref_1}$ 的時間步長 1) 的光流圖 (例如, $\hat{f}_{t \rightarrow 1}^{2D}$)。例如, 2D 光流圖 744 可以包括具有從時間 t 到參考幀 $\hat{X}_{ref_0}$ 的時間序列 (例如, 圖 7C 所示示例中的時間序列 0) 的運動資訊的反向 2D 流圖 $\hat{f}_{t \rightarrow 0}^{2D}$ 、以及具有從時間 t 到參考幀 $\hat{X}_{ref_1}$ 的時間序列 (例如, 圖 7C 所示示例中的時間序列 1) 的運動資訊的前向 2D 流圖 $\hat{f}_{t \rightarrow 1}^{2D}$ 。

【0153】 在一個說明性示例中, 光流預測層 732 可以根據式 1 和式 2 來預測 2D 光流圖 744, 如下所示:

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

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

【0154】 在上面的式 1 和式 2 中， t 表示在參考幀 $\hat{X}_{ref_0}$ 的時間步長和參考幀 $\hat{X}_{ref_1}$ 的時間步長（例如，0 到 1 區間的時間步長）之間的內插幀 $\hat{X}_{ref_t}$ （對應於輸入幀 X_t ）的時間步長， $f_{0 \rightarrow 1}$ 表示從參考幀 $\hat{X}_{ref_0}$ 到參考幀 $\hat{X}_{ref_1}$ 的光流， $f_{1 \rightarrow 0}$ 表示從參考幀 $\hat{X}_{ref_1}$ 到參考幀 $\hat{X}_{ref_0}$ 的光流， $\hat{f}_{t \rightarrow 0}^{2D}$ 表示標識從 t 到與參考幀 $\hat{X}_{ref_0}$ 相關聯的時間步長（例如，圖 7C 所示示例中的時間步長 0）的反向運動的 2D 光流圖， $\hat{f}_{t \rightarrow 1}^{2D}$ 表示標識從 t 到與參考幀 $\hat{X}_{ref_1}$ 相關聯的時間步長（例如，圖 7C 所示示例中的時間步長 1）的前向運動的 2D 光流圖。

【0155】 2D 變形層 734 可以獲得 2D 光流圖 744、參考幀 $\hat{X}_{ref_0}$ 和參考幀 $\hat{X}_{ref_1}$ ，並且可以生成變形幀 746。變形幀 746 可以包括對應於參考幀 $\hat{X}_{ref_0}$ 的變形幀（ W_{x0} ）。2D 變形層 734 可以通過將 2D 光流圖 $\hat{f}_{t \rightarrow 0}^{2D}$ 應用於參考幀 $\hat{X}_{ref_0}$ 來生成變形幀 W_{x0} 。變形幀 746 還可以包括對應於參考幀 $\hat{X}_{ref_1}$ 的變形幀（ W_{x1} ）。2D 變形層 734 可以通過將 2D 光流圖 $\hat{f}_{t \rightarrow 1}^{2D}$ 應用於參考幀 $\hat{X}_{ref_1}$ 來生成變形幀 W_{x1} 。2D 變形層 734 可以實現一個或多個變形函數，以將參考幀 $\hat{X}_{ref_0}$ 變形到時間步長 t ，並且將參考幀 $\hat{X}_{ref_1}$ 變形到時間步長 t 。

【0156】 在一些例子中，為了生成變形幀 W_{x0} ，2D 變形層 734 可以使用為參考幀 $\hat{X}_{ref_0}$ 計算的 2D 光流圖 $\hat{f}_{t \rightarrow 0}^{2D}$ ，以將參考幀 $\hat{X}_{ref_0}$ 中的像素移動到對應於和/或相對於時間 t 的各個位置。類似地，為了生成變形幀 W_{x1} ，2D 變形層 734 可以使用為參考幀 $\hat{X}_{ref_1}$ 計算的 2D 光流圖 $\hat{f}_{t \rightarrow 1}^{2D}$ ，將參考幀 $\hat{X}_{ref_1}$ 中的像素移動到對應於和/或相對於時間 t 的位置。在一些情況下，變形幀 746 可以包括用於參考幀 $\hat{X}_{ref_0}$ 的 2D 幀（其考慮了在時間 t 的像素（和/或像素位置）的垂直和/或水平位移）和用於參考幀 $\hat{X}_{ref_1}$ 的 2D 幀（其考慮了在時間 t 的像素（和/或像素位置）的垂直和/或水平位移）。

【0157】細化層 736 可以獲得 2D 光流圖 744、變形幀 746 以及參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 作為輸入，並且可以使用輸入資料來執行細化操作。在一些例子中，細化層 736 可以是幀插值引擎 725 的可選組件。在一些情況下，細化層 736 可以包括一個或多個神經網路或其它機器學習系統。在一個說明性示例中，細化層 736 可以包括一個或多個插值卷積神經網路 (CNN)。在一些例子中，由細化層 736 執行的細化操作可以包括：細化和/或校正 2D 光流圖 744 中的值。例如，在一些情況下，對應於參考幀 $\hat{X}_{ref_0}$ 的時間步長 0 和對應於參考幀 $\hat{X}_{ref_1}$ 的時間步長 1 之間的一些或全部移動可能不是線性的。細化層 736 可以檢測和校正這種非線性運動。細化層 736 可以生成解釋這種非線性的細化 2D 光流。另外地或替代地，細化層 736 可以細化或校正任何變形幀 746 (和/或參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 中的任何一個) 中的偽影，細化或校正任何變形幀 746 (和/或參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 中的任何一個) 的一部分 (例如，邊緣)，進行細化或校正以解決任何變形幀 746 (和/或參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 中的任何一個) 中的遮擋，和/或執行圖像資料和/或與變形幀 746 和/或參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 相關聯的光流資訊的任何其它細化。

【0158】在一些例子中，細化操作還可以包括：將 2D 光流圖 744 合併到用於內插幀 $\hat{X}_{ref_t}$ 的 3D 光流圖 748 中。例如，如先前所描述的，2D 光流圖 744 可以包括用於參考幀 $\hat{X}_{ref_0}$ 的 2D 光流圖 (例如， $\hat{f}_{t \rightarrow 0}^{2D}$) 和用於參考幀 $\hat{X}_{ref_1}$ 的 2D 光流圖 (例如， $\hat{f}_{t \rightarrow 1}^{2D}$)。細化層 736 可以合併 2D 光流圖 744 (例如，2D 光流圖 $\hat{f}_{t \rightarrow 0}^{2D}$ 和 2D 光流圖 $\hat{f}_{t \rightarrow 1}^{2D}$) 以生成 3D 光流圖 748。3D 光流圖 748 可以將特定幀 (例如，參考幀 $\hat{X}_{ref_0}$ 和參考幀 $\hat{X}_{ref_1}$) 映射到對應的光流資訊。例如，3D 光流圖 748 可以包括兩個空間維度 (例如，X 和 Y 軸或者垂直和水平維度) 和包括幀資訊的第

三維度（例如，幀維度）。第三維度中的幀資訊可以識別運動資訊（例如，垂直和/或水平位移）在兩個空間維度上適用的幀（例如，參考幀 $\hat{X}_{ref_0}$ 或參考幀 $\hat{X}_{ref_1}$ ）。

【0159】 例如，如果第三維度（例如，幀維度）標識參考幀 $\hat{X}_{ref_0}$ ，則空間維度中的對應值（例如，在 X 和 Y 軸或垂直和水平維度上）可以指示參考幀 $\hat{X}_{ref_0}$ 中像素的水平和/或垂直位移（例如，參考幀 $\hat{X}_{ref_0}$ 中的像素應基於空間維度中的光流資訊來移動多少）。類似地，如果第三維度（例如，幀維度）標識參考幀 $\hat{X}_{ref_1}$ ，則空間維度中的對應值（例如，在 X 和 Y 軸或垂直和水平維度上）可以指示參考幀 $\hat{X}_{ref_1}$ 中像素的水平和/或垂直位移（例如，參考幀 $\hat{X}_{ref_1}$ 中的像素應基於空間維度中的光流資訊來移動多少）。

【0160】 在一些情況下，細化層 736 可以使用 2D 光流圖 744 將 2D 光流圖 744 合併到 3D 光流圖 748 中。在一些例子中，細化層 736 可以使用通過細化 2D 光流圖 744 生成的經細化將 2D 光流，將 2D 光流圖 744 合併到 3D 光流圖 748 中。在一些情況下，細化層 736 可以使用參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 、變形幀 746 和/或 2D 光流圖 744（或細化的 2D 光流）作為輸入，來生成 3D 光流圖 748。

【0161】 3D 變形層 738 可以使用 3D 光流圖 748 和參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ （和/或變形幀 746）來生成與時間步長 t 相對應（與輸入幀 X_t 的時間步長相對應）的內插參考幀 $\hat{X}_{ref_t}$ 。例如，3D 變形層 738 可以使用 3D 光流圖 748 對參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ （或變形幀 746）進行變形，並融合變形的幀（例如，變形參考幀 $\hat{X}_{ref_0}$ 和 $\hat{X}_{ref_1}$ 或變形幀 746），以生成內插參考幀 $\hat{X}_{ref_t}$ 。在一些情況下，3D 變形層 738 可以實現一個或多個變形函數。在一個說明性示例中，可以使用下式來生成內插幀：

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

式 (3)

【0162】 其中 $g(\cdot, \cdot)$ 是後向變形函數(在一些情況下可以使用雙線性插值來實現) , 參數 α_0 控制兩個輸入圖像 I_0 和 I_1 的貢獻, 並取決於時間一致性和遮擋推理, 而 $\odot$ 項表示元素乘法(其導致基於 α_0 參數的輸入圖像的內容感知權重) 。

【0163】 在一些情況下, 可以使用一個或多個損失函數來訓練和/或優化光流估計層 730、光流預測層 732、2D 變形層 734、細化層 736 和/或 3D 變形層 738。在一些情況下, 光流估計層 730、光流預測層 732、2D 變形層 734、細化層 736 和 3D 變形層 738 可以使用一個或多個損失函數進行端到端訓練。在一個說明性示例中, 可以使用上面所描述的損失 ($Loss = D + \beta R$) 。

【0164】 雖然關於使用光流插值來描述了圖 7C, 但應當注意, 光流插值在本文中用作說明性示例以用於解釋目的。本領域通常知識者應當認識到, 圖 7B 的幀插值引擎 722 和/或圖 7C 的幀插值引擎 725 可以實現任何其它幀插值技術。

【0165】 圖 8 是說明可以執行以確定參考幀(例如, 從參考幀 $\hat{X}_{ref_0}$ 到參考幀 $\hat{X}_{ref_1}$ 或反之亦然) 之間的運動資訊的不同類型的運動估計的圖。圖 8 中的運動估計可以用於確定內插參考幀(例如, 如關於圖 7B 和圖 7C 所描述的) 和/或對參考幀進行變形(例如, 通過圖 7A 和圖 7B 的變形引擎 704) 。在圖 8 中, 項 x 表示可以從其估計運動的參考幀, 項 f 表示運動估計, 並且項 y 表示可以如下計算的變形幀: $y = f(x)$ 。

【0166】 一種類型的運動估計是基於塊的運動估計技術 802。可以在逐塊的基礎上執行基於塊的運動估計。例如, 對於幀 y 中的每個塊, 運動估計 f 定義了幀 x 中相應塊的位置。在一個說明性示例中, 運動估計 f 可以包括指示幀 y 中

的塊相對於幀 x 中的對應塊的位移（例如，水平和垂直位移）的運動向量。通過確定塊之間的相似性（例如，像素值的相似性），可以確定來自幀 x 的塊對應於幀 y 中的塊。

【0167】 可以執行的另一種類型的運動估計是光流運動估計技術 804。可以在逐個像素的基礎上執行光流運動估計。例如，對於幀 y 中的每個像素，運動估計 f 定義了幀 x 中相應像素的位置。每個像素的運動估計 f 可以包括指示像素在幀之間的移動的向量。在一些情況下，可以基於幀之間的光流向量的計算，來生成光流圖（也稱為運動向量圖）。光流圖可以包括針對幀中每個像素的光流向量，其中每個向量指示像素在幀之間的移動。在一個說明性示例中，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），顯示像素從第一幀到第二幀的移動。

【0168】 在某些情況下，光流圖可以包括幀中少於所有像素的向量。例如，可以在幀之間計算密集光流，以生成針對幀中每個像素的光流向量，該向量可以包含在密集光流圖中。在一些例子中，每個光流圖可以包括 2D 向量場，每個向量是位移向量，其顯示了點從第一幀到第二幀的移動。

【0169】 如上所述，可以在幀序列中的幀之間計算光流向量或光流圖。兩個幀可以包括作為連續擷取的幀的兩個直接相鄰的幀，也可以包括在幀序列中相隔一定距離（例如，在彼此的兩個幀內、在彼此的三個幀內、或者任何其它適當的距離）的兩個幀。在一個說明性示例中，幀 x 中的像素 $I(x, y, t)$ 可以在幀 y 中移動某個距離或位移 $(\Delta x, \Delta y)$ 。

【0170】 可以執行的另一種類型的運動估計是動態卷積（也稱為自適應卷積）運動估計技術 806。可以逐個像素地執行動態卷積運動估計。例如，對於幀

y 中的每個像素，運動估計 f 定義了幀 x 中像素鄰域的線性組合。如圖 8 中所示，對於幀 y 中某個像素位置處的像素，圍繞幀 x 中相同像素位置處的像素周圍的像素鄰域定義一個內核或矩陣。對幀 y 中的位置 $(2, 7)$ （對應於（列號，行號））的像素執行圖 8 中的示例性動態卷積運動估計，其中幀 y 中的左上像素位置在位置 $(0, 0)$ 。圍繞幀 x 中位置 $(2, 7)$ 處的像素來定義內核。例如，如圖所示，將大小為 5×5 的內核應用於幀 x 中位置 $(2, 7)$ 處的像素周圍的像素鄰域。

【0171】內核或矩陣中的每個條目可以包括一個數字（例如，圖 8 的示例中的一組 5×5 數字）。內核中的數字集可以描述如何組合來自在幀 x 中的該像素位置處的像素周圍的像素，其中該像素位置對應於在幀 y 中正在映射和/或重建的像素的相同像素位置。

【0172】圖 9 和圖 10 是示出圖片組（GoP）結構（對應於 GoP 中的幀的順序）的不同示例的圖。圖 9 示出了 I 幀、B 幀、I 幀（IBI）譯碼技術的示例。例如，IBI 技術可以用於傳統譯碼系統（例如，根據諸如 HEVC 或 VVC 之類的一種或多種視頻譯碼標準進行操作）中的隨機存取模式。如圖所示，對於第一 GoP（標記為 GoP1），譯碼系統（例如，圖 6 的譯碼系統 600）可以將第一幀和最後幀譯碼（例如，編碼/壓縮和/或解碼/解壓縮）為 I 幀，並且可以將中間幀壓縮為 B 幀。如標記為「Ref」的列中的箭頭所示，兩個 I 幀用作 B 幀壓縮的參考幀。標記有「順序」的列表示首先對 I 幀進行譯碼，然後對中間的 B 幀進行譯碼。在第二 GoP（標記為 GoP2）中，可以使用來自 GoP1 的參考幀對 B 幀進行譯碼，因此不需要將 GoP2 中的第一幀譯碼為 I 幀。可以使用 GoP1 中的最後幀（I 幀）並使用最後幀（也是 I 幀）作為參考幀，對 GoP2 中的 B 幀進行譯碼。圖 6 的 B 幀譯碼系統 600 或圖 7B 的 B 幀譯碼系統 720 可以用於對 B 幀進行譯碼。例如，

圖 9 的 GoP1 中的第一個 I 幀可以是圖 7B 中的參考幀 $\hat{X}_{ref_0}$ ，圖 9 的 GoP1 中的最後一個 I 幀可以是圖 7B 中的參考幀 $\hat{X}_{ref_1}$ 。圖 9 的 GoP1 中的 B 幀可以是圖 7B 中的重建幀 $\hat{X}_t$ ，在這種情況下，可以使用 GoP1 中的兩個 I 幀來生成內插幀($\hat{X}_{ref_t}$)。

【0173】圖 10 是示出 I 幀、B 幀、P 幀 (IBP) 譯碼技術的示例的圖。例如，因為本文所描述的 B 幀譯碼系統 (例如，圖 7B 的 B 幀譯碼系統 720) 使用 P 幀譯碼系統 (例如，圖 7A 和圖 7B 的 P 幀譯碼系統 700)，當 P 幀譯碼系統用於單向預測和雙向預測時，如果參數 (例如，權重、偏差等) 是常見的，則可以使用相同的組件生成 P 幀，而不增加額外的複雜性。在圖 10 的例子中，每個 GoP 的最後一幀不必是 I 幀 (如圖 9 的 IBI GoP 結構中的情況)，而是可以譯碼為 P 幀。與圖 9 的譯碼順序相比，圖 10 中的譯碼順序也不同。例如，首先將圖 1 中的 GoP1 的第一幀譯碼 (例如，編碼/壓縮和/或解碼/解壓縮) 為 I 幀。使用 I 幀作為參考幀，可以將 GoP1 中的最後一個幀譯碼為 P 幀 (例如，通過圖 7A 的 P 幀譯碼系統 700)。然後，GoP1 中的 I 幀和 P 幀可以用於將中間幀譯碼為 B 幀 (例如，使用圖 6 的 B 幀譯碼系統 600 或圖 7B 的 B 幀譯碼系統 720)。對於 GoP2，GoP1 的最後一幀可以用作 (例如，由圖 7A 的 P-幀譯碼系統 700 用作) 參考幀，以用於將 GoP2 的最後一幀譯碼為 P-幀。GoP1 的最後一幀和 GoP2 的最後一幀可以用作 (例如，由圖 6 的 B 幀譯碼系統 600 或圖 7B 的 B 幀譯碼系統 720 用作) 參考幀，以用於將 GoP2 的中間幀譯碼為 B 幀。

【0174】圖 11 和圖 12 是說明 GoP 內的幀譯碼順序的圖。例如，一旦譯碼系統 (例如，I 幀譯碼系統和/或 P 幀譯碼系統) 確定了用於 GoP 的兩個參考幀 (例如，圖 11 和圖 12 的 GoP 中的第一個幀和最後一個幀)，B 幀譯碼系統 (例如，由圖 6 的 B 幀譯碼系統 600 或圖 7B 的 B 幀譯碼系統 720) 可以開始將 GoP

的中間幀譯碼為 B-幀。圖 11 和圖 12 的示例譯碼順序定義了對那些 B 幀進行譯碼（例如，編碼/壓縮和/或解碼/解壓縮）的順序的示例。在一個說明性示例中，圖 11 和圖 12 的 GoP 的第一個幀和最後一個幀可以包括 I 幀，類似於圖 9 中所示的。在另一個說明性示例中，圖 11 和圖 12 的 GoP 的第一幀可以包括 I 幀，並且最後幀可以包括 P 幀，類似於圖 10 所示的。

【0175】 圖 11 示出了用於雙向預測的順序方案的例子。在圖 11 的順序方案中，從左到右順序地譯碼 B 幀（例如，由圖 6 的 B 幀譯碼系統 600 或圖 7B 的 B 幀譯碼系統 720 進行譯碼）。如圖所示，在將 GoP 的第一幀和最後一幀用作參考幀的情況下，首先使用雙向預測對序號為 2 的 B 幀進行譯碼。接下來在將前一個 B 幀（序號為 2）和 GoP 的最後一幀用作參考幀的情況下，使用雙向預測對序號為 3 的 B 幀進行譯碼。在序號為 3 的 B 幀之後，對序號為 4 的 B 幀進行譯碼，然後是序號為 5 的 B 幀，以此類推。

【0176】 圖 12 示出了用於雙向幀預測的分層方案的例子。在圖 12 的分層方案中，一旦兩個初始參考幀（GoP 的第一幀和最後一幀）可用，譯碼系統（例如，圖 6 的 B 幀譯碼系統 600 或圖 7B 的 B 幀譯碼系統 720）就可以在雙向預測中使用這兩個初始參考幀，將第五幀（序號為 2 的中間幀）譯碼為第一 B 幀。然後，譯碼系統可以使用圖 12 中「Ref」列指示的參考幀，對 GoP 中的第三幀或第七幀（都具有序號 3）進行譯碼。例如，可以使用 GoP 的第一幀（其可以是 I 幀）和 GoP 的第五幀（其可以是 B 幀）作為參考幀，對第三幀（序號為 3）進行譯碼，並且可以使用 GoP 的最後一幀（其可以是 I 幀或 P 幀）和 GoP 的第五幀（B 幀）作為參考幀，對第七幀（序號為 3）進行譯碼。然後，可以使用由圖 12 中的「Ref」列指示的參考幀，以任何順序對標有序號 4 的幀進行譯碼。

【0177】圖 13A 和圖 13B 是示出圖 6 的 B 幀譯碼系統 600（例如，使用圖 7B 的 B 幀譯碼系統 720 實現）相對於不使用內插參考幀的基線譯碼系統的性能的圖表。在圖 13A 和圖 13B 的圖表中說明的所有三個譯碼系統是使用相同訓練條件（例如，使用相同的訓練資料集、相同的損失函數和相同的訓練迭代次數等）訓練的基於神經網路的譯碼系統。這些圖表通過繪製速率（表示為每像素位數）作為峰值訊雜比（PSNR）的函數來表示性能。線 1302 示出了當使用圖 12 的分層譯碼順序時，圖 7B 的 B 幀譯碼系統 720 的性能。線 1304 示出了當使用圖 11 的順序譯碼順序時，B 幀譯碼系統 720 的性能。線 1306 示出了基線譯碼系統的性能。這些圖的左上角（較低的速率，較高的 PSNR）表示更好的性能。如圖 13A 和圖 13B 中所示，線 1302（使用分層譯碼的 B 幀譯碼系統 720）表示最高性能，其次是線 1304（使用順序譯碼的 B 幀譯碼系統 720），其次是線 1306（基線譯碼系統）。

【0178】圖 14A 至圖 14F 是依據每幀速率（如標記為速率與幀索引的圖表所示）、每幀的 PSNR（如標記為 PSNR 與幀索引的圖表所示）、以及每幀的多尺度結構相似性（MS-SSIM）（如標記為 MS-SSIM 與幀索引的圖表所示），示出圖 6 的 B 幀譯碼系統 600（例如，使用圖 7B 的 B 幀譯碼系統 720 實現）的性能的圖表。這些圖表的 x 軸包括表示幀順序的幀索引，並且 y 軸包括各種性能指標（例如，速率、PSNR 和 MS-SSIM）。在圖 14A 至圖 14F 的例子中，每個 GoP 中的幀數是 12，並且這些圖示出了用於生成結果的視頻序列中所有 GoP 的平均值。

【0179】使用幀內預測，將每個 GoP 序列中的第一幀（具有幀索引 0）生成為 I 幀（例如，如圖 9 和圖 10 的 GoP 結構所示）。圖 14A 中的點 1402、圖

14B 中的點 1404、圖 14C 中的點 1406、圖 14D 中的點 1408、圖 14E 中的點 1410 和圖 14E 中的點 1412 表示 I 幀的度量（速率、PSNR 或 MS-SSIM）。例如，圖 14A 中的點 1402 顯示了用於該 I 幀的壓縮的速率（位元數）。使用單向預測，將每個 GoP 序列中的最後一幀（具有幀索引 11）生成為 P 幀（例如，由圖 7A 的 P 幀譯碼系統 700 生成）。圖 14A 中的點 1414、圖 14B 中的點 1416、圖 14C 中的點 1418、圖 14D 中的點 1420、圖 14E 中的點 1422 和圖 14E 中的點 1424 表示 P 幀的度量（速率、PSNR 或 MS-SSIM）。例如，如圖 14A 中所示，與用於 I 幀的位元相比，更少的位元用於 P 幀。

【0180】將第一幀和最後一幀之間的幀譯碼為 B 幀（例如，由圖 7B 的 B 幀譯碼系統 720 執行）。顯示了 B 幀的兩組結果，一組使用圖 11 的順序譯碼順序，並且另一組使用圖 12 的分層譯碼順序。參照圖 14A 作為示例，線 1426 表示使用順序譯碼順序的結果，其中首先將幀索引為 1 的幀譯碼為 B 幀，接著是幀索引為 2 的幀，依此類推。圖 15A 中的線 1428 表示使用分層譯碼順序的結果，其中首先將幀索引為 5 的幀譯碼為 B 幀，接著是幀索引為 2 的幀，接著是幀索引為 8 的幀，以此類推。如圖 14A 中所示，基於參考幀到 B 幀的距離對於後續譯碼的 B 幀較小，較少的位元用於後續譯碼的 B 幀，如圖 12 中所示。例如，使用 I 幀（GoP 的第一幀）和 P 幀（GoP 的最後一幀）作為參考幀，對幀索引為 5 的幀進行雙向預測，而使用幀索引為 2 的 I 幀和 B 幀作為參考幀，對幀索引為 1 的幀進行雙向預測。圖 14B-F 中光滑的線條表示使用圖 12 的分層譯碼順序的相應結果，並且圖 14B-F 中的鋸齒線表示使用圖 11 的順序譯碼順序的相應結果。

【0181】圖 15 是說明用於使用本文描述的技術來執行幀的雙向預測的處理 1500 的示例的流程圖。在框 1502 處，處理 1500 包括：獲得第一參考幀和第

二參考幀。在一個說明性示例中，第一參考幀可以包括圖 6 的參考幀 0 (Ref_0)，第二參考幀可以包括圖 6 的參考幀 1 (Ref_1)。在另一個說明性示例中，第一參考幀可以包括圖 7B 和/或圖 7C 的參考幀 0 ($\hat{X}_{ref0}$)，第二參考幀可以包括圖 7B 和/或圖 7C 的參考幀 1 ($\hat{X}_{ref1}$)。

【0182】在框 1504 處，處理 1500 包括：至少部分地通過在第一參考幀和第二參考幀之間執行插值來生成第三參考幀。例如，處理 1500 可以包括：確定表示從第一參考幀到第三參考幀的像素運動的第一組運動資訊。處理 1500 還可以包括：確定表示從第二參考幀到第三參考幀的像素運動的第二組運動資訊。在一些方面，第一組運動資訊和第二組運動資訊是基於第一參考幀和第二參考幀之間的像素運動來確定的。在一些情況下，第一組運動資訊包括第一光流圖，而第二組運動資訊包括第二光流圖。參考圖 7C 的描述作為一個說明性示例，第一運動資訊可以包括上面使用式 (1) 定義的 $\hat{f}_{t \rightarrow 0}^{2D}$ ，第二運動資訊可以包括上面使用式 (2) 定義的 $\hat{f}_{t \rightarrow 1}^{2D}$ 。處理 1500 可以包括：至少部分地通過使用第一組運動資訊對第一參考幀執行變形函數，來生成第一變形資訊。在一些方面，變形函數包括雙線性插值函數，例如上面關於圖 7C 所描述的。處理 1500 還可以包括：至少部分地通過使用第二組運動資訊對第二參考幀執行變形函數來生成第二變形資訊。處理 1500 可以包括：基於第一變形資訊和第二變形資訊來生成第三參考幀。

【0183】在框 1506 處，處理 1500 包括：基於第三參考幀對輸入幀執行單向幀間預測。例如，為了執行單向幀間預測，處理 1500 可以包括：估計輸入幀和第三參考幀之間的運動。在一個說明性示例中，預測運動可以包括圖 7B 的預測運動 $\hat{f}_t$ 。在一些情況下，輸入幀和第三參考幀之間的估計運動包括光流資料(例

如，使用圖 8 中所示的光流運動估計技術 804 來確定）。在一些方面，輸入幀和第三參考幀之間的估計運動包括動態卷積資料（例如，使用圖 8 所示的動態卷積運動估計技術 806 來確定）。在一些方面，輸入幀和第三參考幀之間的估計運動包括基於塊的運動資料（例如，使用圖 8 中所示的基於塊的運動估計技術 802 來確定）。處理 1500 還可以包括：至少部分地通過基於估計的運動，變形第三參考幀的一個或多個像素來生成變形幀。例如，參照圖 7B 作為說明性示例，變形引擎 704 可以基於預測運動 $\hat{f}_t$ 中包括的運動向量（或其它運動資料）來變形內插參考幀 $\hat{X}_{ref_t}$ 的像素。

【0184】 在一些例子中，為了進一步執行單向幀間預測，處理 1500 可以包括：至少部分地通過確定輸入幀和變形幀之間的差異來確定殘差。處理 1500 還可以包括：使用殘差來生成預測殘差。參照圖 7B 作為說明性示例，P 幀譯碼系統 700 可以生成殘差 r_t 作為幀 t 預測 $\tilde{X}_t$ 和輸入幀 X_t 之間的差。殘差預測系統 706 可以使用上面所描述的技術來生成預測殘差 $\hat{r}_t$ 。

【0185】 在一些方面，使用第一神經網路（例如，第一自動編碼器或其它類型的神經網路）來估計運動，並且使用第二神經網路（例如，第二自動編碼器或其它類型的神經網路）來生成預測殘差。在一個說明性示例中，第一神經網路可以包括圖 7B 的運動預測系統 702，第二神經網路可以包括圖 7B 的殘差預測系統 706。在一些例子中，處理 1500 可以包括：基於變形幀和預測的殘差，生成表示輸入幀的重建幀。該重建幀包括雙向預測幀。再次參照圖 7B 作為說明性示例，P 幀譯碼系統 720 可以通過將幀 t 預測 $\tilde{X}_t$ （通過對內插參考幀 $\hat{X}_{ref_t}$ 進行變形來生成）與預測的殘差 $\hat{r}_t$ 相加來生成重建幀 t ($\hat{X}_t$)。在一些情況下，處理 1500

還可以包括：使重建幀儲存在記憶體（例如，解碼圖片緩衝器（DPB）或其它記憶體）中。在一些例子中，儲存的重建幀可以用於其它幀的幀間預測。

【0186】如上所述，本文所描述的 P 幀譯碼系統（例如，圖 7A 的 P 幀譯碼系統 700）可以使用內插參考幀（例如，內插參考幀 $\hat{X}_{ref_t}$ ）來執行雙向預測，也可以執行單向預測。例如，為了執行單向預測，處理 1500 可以包括：獲得第四參考幀（例如，參考幀 $\hat{X}_{ref}$ ）。處理 1500 還可以包括：估計另外輸入幀和第四參考幀之間的運動（例如，使用運動預測系統 702）。在一個說明性示例中，估計的運動可以包括圖 7A 中所示的預測運動 $\hat{f}_t$ 。處理 1500 可以包括：至少部分地通過基於估計的運動來變形第四參考幀的一個或多個像素，來生成另外的變形幀（例如，使用變形引擎 704）。再次參照圖 7A 作為說明性示例，另外的變形幀可以包括幀預測 $\tilde{X}_t$ 。處理 1500 可以包括：基於另外的變形幀和另外的預測殘差，生成表示另外輸入幀的重建幀，在這種情況下，表示另外輸入幀的重建幀包括單向預測幀。參照圖 7A 作為說明性示例，P 幀譯碼系統 720 可以通過將幀 t 預測 $\tilde{X}_t$ （通過變形參考幀 $\hat{X}_{ref}$ 生成）添加到預測殘差 $\hat{r}_t$ 來生成重建幀 t ($\hat{X}_t$)。在一些情況下，處理 1500 可以包括：使表示另外輸入幀的重建幀儲存在記憶體（例如，解碼圖片緩衝器（DPB）或其它記憶體）中。在一些例子中，儲存的表示另外輸入幀的重建幀可以用於其它幀的幀間預測。

【0187】圖 16 是說明用於使用本文描述的技術，來執行幀的雙向預測的處理 1600 的另一示例的流程圖。在一些情況下，處理 1600 的至少一部分可以由運動預測系統 702 的解碼器網路 703 和殘差預測系統 706 的解碼器網路 709 來執行。在框 1602 處，處理 1600 包括：獲得內插參考幀和表示輸入幀與內插參考幀之間的運動的運動資訊。在一個說明性示例中，內插參考幀可以包括圖 7B 的

內插參考幀 $\hat{X}_{ref_t}$ 。在一些方面，運動資訊是從位元流中獲得的。例如，運動資訊可以包括圖 7B 中所示的潛在 z_m 。在框 1604 處，處理 1600 包括：基於運動資訊，估計輸入幀和內插參考幀之間的運動。例如，運動預測系統 702 的解碼器網路 703 可以基於潛在 z_m ，估計輸入幀 X_t 和內插參考幀 $\hat{X}_{ref_t}$ 之間的運動。估計的運動可以包括圖 7B 的預測運動 $\hat{f}_t$ 。

【0188】 在框 1606 處，處理 1600 包括：至少部分地通過基於估計的運動，對內插參考幀的一個或多個像素進行變形，來生成變形幀。例如，參照圖 7B 作為說明性示例，變形引擎 704 可以基於包括在預測運動 $\hat{f}_t$ 中的運動向量（或其它運動資料），對內插參考幀 $\hat{X}_{ref_t}$ 的像素進行變形。

【0189】 在框 1608 處，處理 1600 包括：基於變形幀和預測殘差，生成表示輸入幀的重建幀。該重建幀包括雙向預測幀。再次參照圖 7B 作為說明性示例，P 幀譯碼系統 720 可以通過將幀 t 預測 $\tilde{X}_t$ （通過變形內插參考幀 $\hat{X}_{ref_t}$ 產生）與預測殘差 $\hat{r}_t$ 相加來產生重建幀 t ($\hat{X}_t$)。在一些情況下，處理 1500 還可以包括：使重建幀儲存在記憶體（例如，解碼圖片緩衝器（DPB）或其它記憶體）中。在一些例子中，儲存的重建幀可以用於其它幀的幀間預測。

【0190】 在一些例子中，處理 1600 可以包括：獲得第一參考幀和第二參考幀。處理 1600 還可以包括：至少部分地通過在第一參考幀和第二參考幀之間執行插值來生成內插參考幀。本文描述的技術（例如，關於圖 7C 和/或圖 15 描述的技術）可以用於生成內插參考幀。

【0191】 在一些例子中，處理 1600 可以包括：從位元流中獲得殘差。例如，殘差預測系統 706 的解碼器網路 709 可以從位元流中獲得潛在 z_r 。潛在 z_r 表示殘差 r_t 。處理 1600 還可以包括：基於獲得的殘差來生成預測殘差。例如，參照圖

7B 作為說明性示例，解碼器網路 709 可以使用潛在 z_r 作為輸入來生成預測殘差 $\hat{r}_t$ 。

【0192】 在一些方面，類似於上面關於處理 1500 所描述的，處理 1600 可以包括：獲得第四參考幀，估計其它輸入幀和第四參考幀之間的運動，並至少部分地通過基於估計的運動，對第四參考幀的一個或多個像素進行變形，來生成另外的變形幀。處理 1600 可以包括：基於另外變形幀和另外的預測殘差，生成表示另外輸入幀的重建幀，表示另外輸入幀的重建幀包括單向預測幀。

【0193】 在一些例子中，處理 1500 和/或處理 1600 可以包括：針對一個或多個訓練迭代，使用損失函數（例如， $Loss$ ）來訓練神經網路系統。在一些情況下，損失函數可以確定一個或多個重建的 B 幀與一個或多個重建的 B 幀表示的一個或多個幀之間的損失。在一些例子中，第一損失函數可以等於 $D + \beta R$ ，如先前所描述的。

【0194】 在一些例子中，本文所描述的處理（例如，處理 1500、處理 1600 和/或本文所描述的其它處理）可以由計算設備或裝置（例如，具有圖 17 所示的計算設備架構 1700 的計算設備）來執行。該計算設備可以包括任何適當的設備，例如自主車輛計算機、機器人設備、移動設備（例如，移動電話）、桌面型計算設備、平板計算設備、可穿戴設備、伺服器設備、視頻戲設備、擴展實境設備（例如，虛擬實境（VR）設備、擴增實境（AR）設備或混合實境（MR）設備）、照相機設備、機頂盒設備、和/或具有資源能力來執行本文所描述的處理的任何其它計算設備，這些處理包括處理 800。在一些例子中，該計算設備可以包括移動設備、可穿戴設備、XR 設備、個人計算機、膝上型計算機、視頻伺服器、電視、照相機、機頂盒、視頻遊戲控制台、或其它設備。在一些例子中，處理 1500

和/或處理 1600 可以由具有實現圖 6 的 B 幀譯碼系統 600 和/或圖 7 的 B 幀譯碼系統 720 的計算設備架構 1700 的計算設備來執行。

【0195】 在一些情況下，計算設備或裝置可以包括各種組件，例如一個或多個輸入設備、一個或多個輸出設備、一個或多個處理器、一個或多個微處理器、一個或多個微型計算機、一個或多個發射器、接收器、或者組合的發射器接收器（例如，稱為收發器）、一個或多個照相機、一個或多個感測器、和/或被配置為執行本文所描述的處理的步驟的其它組件。在一些例子中，計算設備可以包括顯示器、被配置為傳送和/或接收資料的網路介面、它們的任何組合、和/或其它組件。網路介面可以被配置為傳送和/或接收基於網際網路協議（IP）的資料或其它類型的資料。

【0196】 可以利用電路來實現計算設備的組件。例如，這些組件可以包括和/或可以使用電子電路或其它電子硬體來實現，這些電子硬體可以包括一個或多個可程式化電子電路（例如，微處理器、圖形處理單元（GPU）、數位訊號處理器（DSP）、中央處理單元（CPU）、神經處理單元（NPU）、專用積體電路（ASIC）、現場可程式化閘陣列（FPGA）和/或其它適當的電子電路），和/或可以包括和/或使用計算機軟體、韌體或它們的任何組合來實現，以執行本文所描述的各种操作。

【0197】 將處理 1500 和處理 1600 示出為邏輯流程圖，其操作表示可以以硬體、計算機指令或其組合來實現的一系列操作。在計算機指令的上下文中，這些操作表示儲存在一個或多個計算機可讀儲存媒體上的計算機可執行指令，當所述計算機可執行指令被一個或多個處理器執行時，執行所敘述的操作。通常，計算機可執行指令包括執行特定功能或實現特定資料類型的例程、程式、物件、

組件、資料結構等等。描述這些操作的順序不應被理解為是限制性的，並且可以以任何順序和/或並行地組合任意數量的所描述的操作以實現處理。

【0198】 另外，可以在配置有可執行指令的一個或多個計算機系統的控制下執行本文所描述的處理（包括處理 1500 和處理 1600），並且可以將這些處理實現為在一個或多個處理器上共同執行的代碼、通過硬體執行的代碼或者其組合（例如，可執行指令、一個或多個計算機程式或一個或多個應用程式）。如上所述，可以例如以包括多個可由一個或多個處理器執行的指令的計算機程式的形式，將代碼儲存在計算機可讀或機器可讀儲存媒體上。計算機可讀或機器可讀儲存媒體可以是非臨時性的。

【0199】 圖 17 示出了可以實現本文所描述各種技術的示例性計算設備的示例性計算設備架構 1700。在一些例子中，該計算設備可以包括移動設備、可穿戴設備、XR 設備、個人計算機、膝上型計算機、視頻伺服器、視頻遊戲控制台、機器人設備、機頂盒、電視機、相機、伺服器或其它設備。例如，計算設備架構 1700 可以實現圖 6 的 B 幀譯碼系統 600、圖 7A 的 P 幀譯碼系統 700 和/或圖 7B 的 B 幀譯碼系統 720。示出了使用諸如匯流排之類的連接 1705 彼此電通訊的計算設備架構 1700 的組件。示例性計算設備架構 1700 包括處理單元（CPU 或處理器）1710 和計算設備連接 1705，該計算設備連接 1705 將包括計算設備記憶體 1715（例如，唯讀記憶體（ROM）1720 和隨機存取記憶體（RAM）1725）的各種計算設備組件耦接到處理器 1710。

【0200】 計算設備架構 1700 可以包括高速記憶體的高速緩存，該高速記憶體與處理器 1710 直接連接、緊密接近或作為處理器 1710 的一部分而整合。計算設備架構 1700 可以將資料從記憶體 1715 和/或儲存設備 1730 複製到高速緩存

1712 以供處理器 1710 快速存取。用此方式，高速緩存可以提供性能提升，從而避免了處理器 1710 在等待資料時的延遲。這些和其它模組可以控制或者被配置為控制處理器 1710 執行各種動作。其它計算設備記憶體 1715 也可以可用。記憶體 1715 可以包括具有不同性能特徵的多種不同類型的記憶體。處理器 1710 可以包括任何通用處理器和硬體或儲存在儲存設備 1730 中的軟體服務（例如，服務 1 1732、服務 2 1734 和服務 3 1736），並且被配置為控制處理器 1710 以及專用處理器，其中將軟體指令合併到處理器設計方案中。處理器 1710 可以是包含多個核或處理器、匯流排、記憶體控制器、高速緩存等等的自包含系統。多核處理器可以是對稱的或非對稱的。

【0201】為了使用戶能夠與計算設備架構 1700 進行互動，輸入設備 1745 可以表示任意數量的輸入裝置，例如用於語音的麥克風、用於手勢或圖形輸入的觸摸敏感型屏幕、鍵盤、滑鼠、運動輸入、語音等等。輸出設備 1735 也可以是本領域通常知識者已知的眾多輸出裝置中的一種或多種，例如顯示器、投影儀、電視、揚聲器設備等等。在一些實例中，多模式計算設備可以使用戶能夠提供多種類型的輸入，以與計算設備架構 1700 進行通訊。通訊介面 1740 通常可以控制和管理用戶輸入以及計算設備輸出。對於在任何特定硬體裝置上的操作沒有限制，因此在改進的硬體或軟體裝置得到開發時，本文的基本功能可以輕鬆地替換為改進的硬體或軟體裝置。

【0202】儲存設備 1730 是非易失性記憶體，並且可以是能夠儲存能夠由計算機存取的資料的硬碟或其它類型的計算機可讀媒體，例如磁帶、快閃記憶體卡、固態儲存設備、數位通用磁碟、盒式磁帶、隨機存取記憶體（RAM）1725、唯讀記憶體（ROM）1720 以及其混合。儲存設備 1730 可以包括用於控制處理器

1710 的服務 1732、服務 1734 和服務 1736。可以預期其它硬體或軟體模組。儲存設備 1730 可以連接到計算設備連接 1705。在一個方面，執行特定功能的硬體模組可以包括結合必要的硬體組件（例如，處理器 1710、連接 1705、輸出設備 1735 等等）儲存在計算機可讀媒體中的用於執行該功能的軟體組件。

【0203】 術語「設備」並不限於一個或特定數量的實體物件（例如，一個智慧型手機、一個控制器、一個處理系統等）。如本文所使用的，設備可以包括具有可以實施本公開內容的至少一些部分的一個或多個部分的任何電子設備。雖然說明書和示例使用術語「設備」來描述本公開內容的各個方面，但術語「設備」並不限於特定的配置、類型或物件數量。此外，術語「系統」不限於多個組件或特定實施例。例如，系統可以在一個或多個印刷電路板或其它基板上實現，並且可以具有可移動或靜態組件。儘管描述和示例使用術語「系統」來描述本公開內容的各個方面，但術語「系統」並不限於特定的配置、類型或物件數量。

【0204】 在以上的描述中提供了特定細節，以提供對本文提供的實施例和示例的透徹理解。但是，本領域通常知識者應當理解，可以在沒有這些特定細節的情況下實踐這些實施例。為了清楚解釋起見，在一些實例中，可以將本文的技術呈現為包括單獨的功能塊，該單獨的功能塊包括設備、設備組件、以軟體體現的方法或例程、或者硬體和軟體的組合。除了圖中所示的和/或本文所述的組件以外，還可以使用其它組件。例如，可以將電路、系統、網路、過程和其它組件示出為方塊圖形式的組件，以免在不必要的細節上使實施例晦澀難懂。在其它實例中，可以在沒有不必要的細節的情況下，示出公知的電路、過程、算法、結構和技術，以避免使實施例晦澀難懂。

【0205】 在上文將各個實施例描述為過程或方法，將過程或方法描繪為流程圖、流程、資料流程圖、結構圖或方塊圖。雖然流程圖可以將操作描述為順序過程，但是許多操作可以並行或同時地執行。另外，可以重新安排這些操作的順序。流程在其操作完成時終止，但可以有未包含在附圖中的其它步驟。過程可以對應於方法、函數、過程、子例程、子程式等等。當過程對應於函數時，其終止可以對應於該函數返回到調用函數或主函數。

【0206】 可以使用儲存在計算機可讀媒體中或者從計算機可讀媒體中可用的計算機可執行指令，來實現根據上述示例的過程和方法。例如，這樣的指令可以包括引起或配置通用計算機、專用計算機或處理設備執行特定功能或功能組的指令和資料。可以通過網路存取使用的計算機資源部分。計算機可執行指令可以是例如二進制、中間格式指令（例如，匯編語言）、韌體、源代碼等等。

【0207】 術語「計算機可讀媒體」包括但不限於便攜式或非便攜式儲存設備、光儲存設備、以及能夠儲存、包含或攜帶指令和/或資料的各種其它媒體。計算機可讀媒體可以包括在其中能夠儲存資料的非臨時性媒體，並且不包括無線地或通過有線連接傳播的載波和/或臨時性電子訊號。非臨時性媒體的例子可以包括但不限於磁碟或磁帶、諸如快閃記憶體、記憶體或儲存設備、磁碟或光碟之類的光學儲存媒體、快閃記憶體、配備有非易失性記憶體的 USB 設備、網路儲存設備、壓縮光碟（CD）或數位通用光碟（DVD）、其任何適當的組合等等。計算機可讀媒體上可以儲存有代碼和/或機器可執行指令，它們可以表示過程、函數、子程式、程式、例程、子例程、模組、軟體包、類、或者指令、資料結構或程式語句的任何組合。可以通過傳遞和/或接收資訊、資料、自變量、參數或記憶體內容，將代碼段耦接到另一個代碼段或硬體電路。可以通過包括記憶體共

享、訊息傳遞、符記 (token) 傳遞、網路傳輸等等的任何適當方式，來傳遞、轉發或傳輸資訊、自變量、參數、資料等等。

【0208】 在一些實施例中，計算機可讀儲存設備、媒體和記憶體可以包括包含位元流等等的電纜或無線訊號。但是，當提及時，非臨時性計算機可讀儲存媒體明確地排除諸如能量、載波訊號、電磁波和訊號本身之類的媒體。

【0209】 實現根據這些公開內容的過程和方法的設備可以包括硬體、軟體、韌體、中間件、微代碼、硬體描述語言或者其任意組合，並且可以採用多種形式中的任何一種。當以軟體、韌體、中間件或微代碼來實現時，用於執行必要任務的程式碼或代碼段（例如，計算機程式產品）可以儲存在計算機可讀媒體或機器可讀媒體中。處理器可以執行必要的任務。外形尺寸的典型示例包括膝上型計算機、智慧型手機、移動電話、平板設備或者其它小型外形的個人計算機、個人數位助理、機架安裝設備、獨立設備等等。本文所描述的功能還可以體現在外圍設備或附加卡中。作為進一步的示例，還可以在單個設備中執行的不同芯片或不同過程之間，在電路板上實現這種功能。

【0210】 指令、用於傳送這樣的指令的媒體、用於執行它們的計算資源、以及用於支援這樣的計算資源的其它結構，是用於提供本公開內容中描述的功能的示例性裝置。

【0211】 在前面的描述中，參考本申請的特定實施例描述了本申請的各方面，但是本領域通常知識者應當認識到，本申請並不受此限制。因此，雖然本文已經詳細描述了本申請的說明性實施例，但是應當理解，可以以其它方式不同地體現和採用本發明的構思，並且所附申請專利範圍旨在被解釋為包括這樣的變型，除非受到現有技術的限制。可以單獨或聯合地使用上述申請的各種特徵和方

面。此外，在不脫離本說明書的保護範圍的情況下，可以在本文所述之外的任何數量的環境和應用中使用實施例。因此，說明書和附圖應被認為是說明性的而不是限制性的。為了說明起見，以特定順序描述了方法。應當理解，在替代實施例中，可以以與所描述的順序不同的順序來執行所述方法。

【0212】 本領域通常知識者應當理解，可以將本文中使用的「<」和「>」符號或術語分別替換為「≤」和「≥」符號，而不脫離本說明書的範圍。

【0213】 在將組件描述為「被配置為」執行某些操作的情況下，可以例如通過以下方式來實現這種配置：設計電子電路或其它硬體以執行該操作，通過對可程式化電子電路（例如，微處理器或其它適當的電子電路）進行程式化以執行該操作，或其任意組合。

【0214】 片語「耦接到」指代直接或間接地實體連接到另一個組件的任何組件、和/或直接或間接地與另一個組件通訊的任何組件（例如，通過有線或無線連接來連接到另一個組件和/或其它適當的通訊介面）。

【0215】 引用一個集合中的「至少一個」和/或一個集合中的「一個或多個」的聲明語言或其它語言，指示該集合中的一個成員或該組中的多個成員（具有任何組合）滿足申請專利範圍。例如，引用「A 和 B 中的至少一個」或者「A 或 B 中的至少一個」的聲明語言意味著 A、B、或 A 和 B。在另一個例子中，引用「A、B 和 C 中的至少一個」或「A、B 或 C 中的至少一個」的聲明語言意味著 A、B、C、或 A 和 B、或 A 和 C、或 B 和 C、或 A 和 B 和 C。一個集合中的「至少一個」和/或一個集合中的「一個或多個」的語言，並不將該集合限制於該集合中列出的項。例如，引用「A 和 B 中的至少一個」或者「A 或 B 中的至少一個」的

聲明語言可以意味著 A、B、或 A 和 B，並且可以另外地包括 A 和 B 的集中沒有列出的項。

【0216】 結合本文所公開實施例描述各種示例性的邏輯框、模組、電路和演算法步驟均可以實現成電子硬體、計算機軟體、韌體或者其組合。為了清楚地表示硬體和軟體之間的這種可交換性，上面對各種示例性的部件、框、模組、電路和步驟均圍繞其功能進行了總體描述。至於這種功能是實現成硬體還是實現成軟體，取決於特定的應用和對整個系統所施加的設計約束條件。熟練的通常知識者可以針對每個特定應用，以變通的方式實現所描述的功能，但是，這種實現決策不應解釋為背離本公開內容的保護範圍。

【0217】 本文所描述的技術還可以利用電子硬體、計算機軟體、韌體或者其任何組合來實現。可以利用諸如通用計算機、無線通訊設備手持裝置或者具有多種用途（其包括在無線通訊設備手持裝置和其它設備中的應用）的積體電路設備等等之類的各種設備中的任何一種來實現這種技術。描述成模組或組件的任何特徵可以在積體邏輯設備中一起實現，也可以分別作為離散但可互操作的邏輯設備來實現。如果以軟體實施，則這些技術可以至少部分地由計算機可讀資料儲存媒體來實現，所述計算機可讀資料儲存媒體包括程式碼，所述程式碼包括在被執行時執行上述方法、演算法和/或操作中的一個或多個的指令。計算機可讀資料儲存媒體可以形成計算機程式產品的一部分，該計算機程式產品可以包括封裝材料。計算機可讀媒體可以包括記憶體或資料儲存媒體，例如，諸如同步動態隨機存取記憶體（SDRAM）之類的隨機存取記憶體（RAM）、唯讀記憶體（ROM）、非易失性隨機存取記憶體（NVRAM）、電可擦除可程式化唯讀記憶體（EEPROM）、快閃記憶體、磁或光資料儲存媒體等等。另外地或替代地，可

以至少部分地通過計算機可讀通訊媒體來實現這些技術，該計算機可讀通訊媒體以指令或資料結構的形式（例如，傳播的訊號或波）攜帶或傳遞程式碼，並且程式碼可以由計算機存取、讀取和/或執行。

【0218】 程式碼可以由處理器執行，該處理器可以包括一個或多個處理器，例如一個或多個數位訊號處理器(DSP)、通用微處理器、專用積體電路(ASIC)、現場可程式化邏輯陣列(FPGA)或者其它等效的積體邏輯電路或離散邏輯電路。這種處理器可以被配置為執行本公開內容中描述的任何技術。通用處理器可以是微處理器，但是，該處理器也可以是什麼常規處理器、控制器、微控制器或狀態機。處理器也可以實現為計算設備的組合，例如，DSP和微處理器的組合、若干微處理器、與DSP核心結合的一個或多個微處理器、或者任何其它這樣的配置。因此，本文所使用的術語「處理器」可以指代任何的前述結構、前述結構的任何組合、或者適合於實施本文所描述的技術的任何其它結構或裝置。

【0219】 本公開內容的說明性示例包括：

【0220】 方面 1：一種用於處理視頻資料的系統，所述系統包括：雙向譯碼引擎，其包括插值引擎和單向譯碼引擎，其中：所述插值引擎被配置為獲得第一參考幀和第二參考幀，並至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成第三參考幀；所述單向譯碼引擎被配置為基於所述第三參考幀，對輸入幀執行單向幀間預測。

【0221】 方面 2：根據方面 1 所述的系統，其中，所述單向譯碼引擎包括運動引擎和變形引擎，並且其中：所述運動引擎被配置為估計所述輸入幀和所述第三參考幀之間的運動；所述變形引擎被配置為至少部分地通過基於所估計的運動來變形所述第三參考幀的一個或多個像素，以生成變形幀。

【0222】 方面 3：根據方面 2 所述的系統，其中，所述單向譯碼引擎還包括殘差引擎，並且其中：所述殘差引擎被配置為：至少部分地通過確定所述輸入幀和所述變形幀之間的差異來確定殘差；並使用所述殘差來生成預測殘差。

【0223】 方面 4：根據方面 3 所述的系統，還包括記憶體和耦接到所述記憶體的至少一個處理器，其中所述單向譯碼引擎被配置為：基於所述變形幀和所述預測殘差，生成表示所述輸入幀的重建幀，所述重建幀包括雙向預測幀；並且其中，所述至少一個處理器被配置為使所述重建幀被儲存在所述記憶體中。

【0224】 方面 5：根據方面 4 所述的系統，其中，所述至少一個處理器包括神經處理單元（NPU）。

【0225】 方面 6：根據方面 3 至 5 中的任何一項所述的系統，其中，所述運動引擎包括第一神經網路，並且所述殘差引擎包括第二神經網路。

【0226】 方面 7：根據方面 6 所述的系統，其中，所述第一神經網路包括第一自動編碼器，並且所述第二神經網路包括第二自動編碼器。

【0227】 方面 8：根據方面 1 至 7 中的任何一項所述的系統，其中，所述單向譯碼引擎被配置為：獲得第四參考幀；估計另外輸入幀和所述第四參考幀之間的運動；至少部分地通過基於所述估計的運動來變形所述第四參考幀的一個或多個像素，生成另外的變形幀；並基於所述另外的變形幀和另外的預測殘差，生成表示所述另外輸入幀的重建幀，表示所述另外輸入幀的所述重建幀包括單向預測幀。

【0228】 方面 9：根據方面 1 至 8 中的任何一項所述的系統，其中，為了至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成所述第三參考幀，所述插值引擎被配置為：確定表示從所述第一參考幀到所述第三參

考幀的像素運動的第一組運動資訊；確定表示從所述第二參考幀到所述第三參考幀的像素運動的第二組運動資訊；至少部分地通過使用所述第一組運動資訊對所述第一參考幀執行變形函數來生成第一變形資訊；至少部分地通過使用所述第二組運動資訊對所述第二參考幀執行所述變形函數來生成第二變形資訊；並基於所述第一變形資訊和所述第二變形資訊生成所述第三參考幀。

【0229】 方面 10：根據方面 9 所述的系統，其中，基於所述第一參考幀和所述第二參考幀之間的像素運動來確定所述第一組運動資訊和所述第二組運動資訊。

【0230】 方面 11：根據方面 9 或 10 中的任何一項所述的系統，其中，所述第一組運動資訊包括第一光流圖，並且所述第二組運動資訊包括第二光流圖。

【0231】 方面 12：根據方面 9 至 11 中的任何一項所述的系統，其中，所述變形函數包括雙線性插值函數。

【0232】 方面 13：根據方面 2 至 12 中的任何一項所述的系統，其中，所述輸入幀與所述第三參考幀之間的所述估計的運動包括光流資料。

【0233】 方面 14：根據方面 2 至 12 中的任何一項所述的系統，其中，所述輸入幀與所述第三參考幀之間的所述估計的運動包括動態卷積資料。

【0234】 方面 15：根據方面 2 至 12 中的任何一項所述的系統，其中，所述輸入幀與所述第三參考幀之間的所述估計的運動包括基於塊的運動資料。

【0235】 方面 16：根據方面 1 至 15 中的任何一項所述的系統，其中，所述系統是移動設備、擴展實境設備或電視之一。

【0236】 方面 17：根據方面 1 至 16 中的任何一項所述的系統，還包括被配置為擷取一個或多個視頻幀的顯示器和相機中的至少一個。

【0237】 方面 18：一種用於處理視頻資料的方法，所述方法包括：獲得第一參考幀和第二參考幀；至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成第三參考幀；並基於所述第三參考幀對輸入幀執行單向幀間預測。

【0238】 方面 19：根據方面 18 所述的方法，還包括：估計所述輸入幀和所述第三參考幀之間的運動；並至少部分地通過基於所估計的運動，變形所述第三參考幀的一個或多個像素，來生成變形幀。

【0239】 方面 20：根據方面 19 所述的方法，還包括：至少部分地通過確定所述輸入幀和所述變形幀之間的差異來確定殘差；並使用所述殘差生成預測殘差。

【0240】 方面 21：根據方面 20 所述的方法，還包括：基於所述變形幀和所述預測殘差，生成表示所述輸入幀的重建幀，所述重建幀包括雙向預測幀；並使所述重建幀儲存在記憶體中。

【0241】 方面 22：根據方面 20 或 21 中的任何一項所述的方法，其中，所述運動是使用第一自動編碼器估計的，並且其中所述預測殘差是使用第二自動編碼器生成的。

【0242】 方面 23：根據方面 18 至 22 中的任何一項所述的方法，還包括：獲得第四參考幀；估計另外輸入幀和所述第四參考幀之間的運動；至少部分地通過基於所估計的運動來變形所述第四參考幀的一個或多個像素，生成另外的變形幀；並基於所述另外的變形幀和另外的預測殘差，生成表示所述另外輸入幀的重建幀，表示所述另外輸入幀的所述重建幀包括單向預測幀。

【0243】 方面 24：根據方面 18 至 23 中的任何一項所述的方法，還包括：確定表示從所述第一參考幀到所述第三參考幀的像素運動的第一組運動資訊；確定表示從所述第二參考幀到所述第三參考幀的像素運動的第二組運動資訊；至少部分地通過使用所述第一組運動資訊對所述第一參考幀執行變形函數來生成第一變形資訊；至少部分地通過使用所述第二組運動資訊對所述第二參考幀執行所述變形函數來生成第二變形資訊；並基於所述第一變形資訊和所述第二變形資訊生成所述第三參考幀。

【0244】 方面 25：根據方面 24 所述的方法，其中，基於所述第一參考幀和所述第二參考幀之間的像素運動來確定所述第一組運動資訊和所述第二組運動資訊。

【0245】 方面 26：根據方面 24 或 25 中的任何一項所述的方法，其中，所述第一組運動資訊包括第一光流圖，並且所述第二組運動資訊包括第二光流圖。

【0246】 方面 27：根據方面 24 至 26 中的任何一項所述的方法，其中，所述變形函數包括雙線性插值函數。

【0247】 方面 28：根據方面 19 至 27 中的任何一項所述的方法，其中，所述輸入幀與所述第三參考幀之間的所述估計的運動包括光流資料。

【0248】 方面 29：根據方面 19 至 27 中的任何一項所述的方法，其中，所述輸入幀與所述第三參考幀之間的所述估計的運動包括動態卷積資料。

【0249】 方面 30：根據方面 19 至 27 中的任何一項所述的方法，其中，所述輸入幀與所述第三參考幀之間的所述估計的運動包括基於塊的運動資料。

【0250】 方面 31：一種其上儲存有指令的非臨時性計算機可讀媒體，當所述指令被一個或多個處理器執行時，使所述一個或多個處理器執行以下操作：獲

得第一參考幀和第二參考幀；至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成第三參考幀；並基於所述第三參考幀對輸入幀執行單向幀間預測。

【0251】 方面 32：根據方面 31 所述的非臨時性計算機可讀媒體，還包括當被所述一個或多個處理器執行時，使得所述一個或多個處理器執行方面 19 至 30 中的任何一項所述的操作的指令。

【0252】 方面 33：一種用於處理視頻資料的裝置，包括：用於獲得第一參考幀和第二參考幀的構件；用於至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成第三參考幀的構件；用於基於所述第三參考幀對輸入幀執行單向幀間預測的構件。

【0253】 方面 34：根據方面 33 所述的裝置，還包括用於執行方面 19 至 30 中的任何一項所述的操作的構件。

【0254】 方面 35：一種用於處理視頻資料的系統，所述系統包括：記憶體；以及耦接到所述記憶體的至少一個處理器，所述至少一個處理器被配置為：獲得內插參考幀和表示輸入幀與所述內插參考幀之間運動的運動資訊；基於所述運動資訊，估計所述輸入幀和所述內插參考幀之間的運動；至少部分地通過基於所述估計的運動，對所述內插參考幀的一個或多個像素進行變形來生成變形幀；並基於所述變形幀和預測殘差來生成表示所述輸入幀的重建幀，所述重建幀包括雙向預測幀。

【0255】 方面 36：根據方面 35 所述的系統，其中，所述至少一個處理器被配置為：獲得第一參考幀和第二參考幀；並至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成所述內插參考幀。

【0256】 方面 37：根據方面 35 所述的系統，其中，所述至少一個處理器被配置為：從位元流中獲取殘差；並基於所獲得的殘差來生成預測殘差。

【0257】 方面 38：根據方面 37 所述的系統，其中，所述運動資訊是從所述位元流中獲得的。

【0258】 方面 39：根據方面 35 所述的系統，其中，所述至少一個處理器被配置為：獲得第四參考幀；估計另外輸入幀和所述第四參考幀之間的運動；至少部分地通過基於所估計的運動，變形所述第四參考幀的一個或多個像素來生成另外的變形幀；並基於所述另外的變形幀和另外的預測殘差，生成表示所述另外輸入幀的重建幀，表示所述另外輸入幀的所述重建幀包括單向預測幀。

【0259】 方面 40：一種用於處理視頻資料的方法，所述方法包括：獲得內插參考幀和表示輸入幀與所述內插參考幀之間運動的運動資訊；基於所述運動資訊，估計所述輸入幀和所述內插參考幀之間的運動；至少部分地通過基於所估計的運動，對所述內插參考幀的一個或多個像素進行變形來生成變形幀；並基於所述變形幀和預測殘差來生成表示所述輸入幀的重建幀，所述重建幀包括雙向預測幀。

【0260】 方面 41：根據方面 40 所述的方法，還包括：獲得第一參考幀和第二參考幀；並至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成所述內插參考幀。

【0261】 方面 42：根據方面 40 所述的方法，還包括：從位元流中獲取殘差；並基於所獲得的殘差來生成預測殘差。

【0262】 方面 43：根據方面 42 所述的方法，其中，所述運動資訊是從所述位元流中獲得的。

【0263】 方面 44：根據方面 40 所述的方法，還包括：獲得第四參考幀；估計另外輸入幀和所述第四參考幀之間的運動；至少部分地通過基於所估計的運動，變形所述第四參考幀的一個或多個像素來生成另外的變形幀；並基於所述另外的變形幀和另外的預測殘差，生成表示所述另外輸入幀的重建幀，表示所述另外輸入幀的所述重建幀包括單向預測幀。

【0264】 方面 45：一種其上儲存有指令的非臨時性計算機可讀媒體，當所述指令被一個或多個處理器執行時，使所述一個或多個處理器執行以下操作：獲得內插參考幀和表示輸入幀與所述內插參考幀之間運動的運動資訊；基於所述運動資訊，估計所述輸入幀和所述內插參考幀之間的運動；至少部分地通過基於所述估計的運動，對所述內插參考幀的一個或多個像素進行變形來生成變形幀；並基於所述變形幀和預測殘差來生成表示所述輸入幀的重建幀，所述重建幀包括雙向預測幀。

【0265】 方面 46：根據方面 45 所述的非臨時性計算機可讀媒體，還包括當被所述一個或多個處理器執行時，使得所述一個或多個處理器執行根據方面 41 至 44 中的任何一項所述的操作的指令。

【0266】 方面 47：一種用於處理視頻資料的裝置，包括：用於獲得內插參考幀和表示輸入幀與所述內插參考幀之間運動的運動資訊的構件；用於基於所述運動資訊，估計所述輸入幀和所述內插參考幀之間的運動的構件；用於至少部分地通過基於所述估計的運動，對所述內插參考幀的一個或多個像素進行變形來生成變形幀的構件；用於基於所述變形幀和預測殘差來生成表示所述輸入幀的重建幀的構件，所述重建幀包括雙向預測幀。

【0267】 方面 48：根據方面 47 所述的裝置，還包括用於執行根據方面 41 至 44 中的任何一項所述的操作的構件。

【符號說明】

【0268】

- 100 圖像處理系統
- 102 中央處理單元（CPU）
- 104 圖形處理單元（GPU）
- 106 數位訊號處理器（DSP）
- 108 神經處理單元（NPU）
- 110 連接塊
- 112 多媒體處理器
- 114 感測器
- 116 圖像訊號處理器（ISP）
- 118 儲存記憶體塊
- 120 儲存設備
- 200 深度卷積網路（DCN）
- 202 全連接神經網路
- 204 局部連接的神經網路
- 206 卷積神經網路
- 208 連接強度
- 210 值

212	值
214	值
216	值
218	第一組特徵圖
220	第二組特徵圖
222	輸出
224	第一特徵向量
226	圖像
228	第二特徵向量
230	圖像擷取設備
232	卷積層
350	深度卷積網路
352	輸入資料
354A	卷積塊
354B	卷積塊
356	卷積層 (CONV)
358	歸一化層 (LNorm)
360	最大池化層 (MAX POOL)
362A	層
362B	層
364	邏輯回歸 (LR) 層
366	分類得分

- 400 P 幀譯碼系統
- 401 輸入幀
- 402 運動壓縮引擎
- 403 參考幀
- 404 運動補償引擎
- 406 殘差壓縮引擎
- 500 B 幀譯碼系統
- 501 輸入幀
- 502 運動壓縮引擎
- 504 運動補償引擎
- 506 殘差壓縮引擎
- 510 B 幀譯碼系統
- 511 輸入幀
- 512 幀插值引擎
- 516 殘差壓縮引擎
- 600 B 幀譯碼系統
- 611 粗略預測輸入幀
- 613 內插參考幀
- 615 輸入幀
- 622 幀插值引擎
- 624 P 幀譯碼系統
- 700 P 幀譯碼系統

- 702 運動預測系統
- 703 解碼器網路
- 704 變形引擎
- 705 編碼器網路
- 706 殘差預測系統
- 707 編碼器網路
- 708 減法運算
- 709 解碼器網路
- 710 加法運算
- 720 B 幀譯碼系統
- 722 幀插值引擎
- 725 幀插值引擎
- 730 光流估計層
- 732 光流預測層
- 734 2D 變形層
- 736 細化層
- 738 3D 變形層
- 740 雙向運動資訊
- 742 時間資訊
- 744 2D 光流圖
- 746 變形幀
- 748 3D 光流圖

- 802 基於塊的運動估計技術
- 804 光流運動估計技術
- 806 動態卷積運動估計技術
- 1302 線
- 1304 線
- 1402 點
- 1404 點
- 1406 點
- 1408 點
- 1410 點
- 1412 點
- 1414 點
- 1416 點
- 1418 點
- 1420 點
- 1422 點
- 1424 點
- 1426 線
- 1428 線
- 1500 處理
- 1502 框
- 1504 框

1506 框

1600 處理

1602 框

1604 框

1606 框

1608 框

1700 計算設備架構

1705 連接

1710 處理單元（CPU 或處理器）

1712 高速緩存

1715 儲存記憶體

1720 唯讀儲存記憶體（ROM）

1725 隨機存取儲存記憶體（RAM）

1730 儲存設備

1732 服務 1

1734 服務 2

1735 輸出設備

1736 服務 3

1740 通訊介面

1745 輸入設備

【發明申請專利範圍】

【請求項1】 一種用於處理視頻資料的系統，所述系統包括：

雙向譯碼引擎，其包括插值引擎和單向譯碼引擎，其中：

所述插值引擎被配置為：

獲得第一參考幀和第二參考幀；以及

至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成第三參考幀；以及

所述單向譯碼引擎被配置為：

基於所述第三參考幀，對輸入幀執行單向幀間預測。

【請求項2】 根據請求項1所述的系統，其中，所述單向譯碼引擎包括運動引擎和變形引擎，並且其中：

所述運動引擎被配置為估計所述輸入幀和所述第三參考幀之間的運動；以及

所述變形引擎被配置為至少部分地通過基於所述估計的運動來變形所述第三參考幀的一個或多個像素，以生成變形幀。

【請求項3】 根據請求項2所述的系統，其中，所述單向譯碼引擎還包括殘差引擎，並且其中：

所述殘差引擎被配置為：

至少部分地通過確定所述輸入幀和所述變形幀之間的差異來確定殘差；以及

使用所述殘差來生成預測殘差。

【請求項4】 根據請求項3所述的系統，還包括記憶體和耦接到所述記憶體的至少一個處理器，其中，所述單向譯碼引擎被配置為：

基於所述變形幀和所述預測殘差，生成表示所述輸入幀的重建幀，所述重建幀包括雙向預測幀；以及

其中，所述至少一個處理器被配置為使所述重建幀被儲存在所述記憶體中。

【請求項5】 根據請求項4所述的系統，其中，所述至少一個處理器包括神經處理單元（NPU）。

【請求項6】 根據請求項3所述的系統，其中，所述運動引擎包括第一神經網路，並且所述殘差引擎包括第二神經網路。

【請求項7】 根據請求項6所述的系統，其中，所述第一神經網路包括第一自動編碼器，並且所述第二神經網路包括第二自動編碼器。

【請求項8】 根據請求項1所述的系統，其中，所述單向譯碼引擎被配置為：

獲得第四參考幀；

估計另外輸入幀和所述第四參考幀之間的運動；

至少部分地通過基於所述估計的運動來變形所述第四參考幀的一個或多個像素，生成另外的變形幀；以及

基於所述另外的變形幀和另外的預測殘差，生成表示所述另外輸入幀的重建幀，表示所述另外輸入幀的所述重建幀包括單向預測幀。

【請求項9】 根據請求項1所述的系統，其中，為了至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成所述第三參考幀，所述插值引擎被配置為：

確定表示從所述第一參考幀到所述第三參考幀的像素運動的第一組運動資訊；

確定表示從所述第二參考幀到所述第三參考幀的像素運動的第二組運動資訊；

至少部分地通過使用所述第一組運動資訊對所述第一參考幀執行變形函數來生成第一變形資訊；

至少部分地通過使用所述第二組運動資訊對所述第二參考幀執行所述變形函數來生成第二變形資訊；以及

基於所述第一變形資訊和所述第二變形資訊生成所述第三參考幀。

【請求項10】 根據請求項9所述的系統，其中，所述第一組運動資訊和所述第二組運動資訊是基於所述第一參考幀和所述第二參考幀之間的像素運動來確定的。

【請求項11】 根據請求項9所述的系統，其中，所述第一組運動資訊包括第一光流圖，並且所述第二組運動資訊包括第二光流圖。

【請求項12】 根據請求項9所述的系統，其中，所述變形函數包括雙線性插值函數。

【請求項13】 根據請求項2所述的系統，其中，所述輸入幀與所述第三參考幀之間的所述估計的運動包括光流資料。

【請求項14】 根據請求項2所述的系統，其中，所述輸入幀與所述第三參考幀之間的所述估計的運動包括動態卷積資料。

【請求項15】 根據請求項2所述的系統，其中，所述輸入幀與所述第三參考幀之間的所述估計的運動包括基於塊的運動資料。

【請求項16】 根據請求項1所述的系統，其中，所述系統是移動設備、擴展實境設備或電視之一。

【請求項17】 根據請求項1所述的系統，還包括：

被配置為擷取一個或多個視頻幀的顯示器和相機中的至少一個。

【請求項18】 一種用於處理視頻資料的方法，所述方法包括：

獲得第一參考幀和第二參考幀；

至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成第三參考幀；以及

基於所述第三參考幀對輸入幀執行單向幀間預測。

【請求項19】 根據請求項18所述的方法，還包括：

估計所述輸入幀和所述第三參考幀之間的運動；以及

至少部分地通過基於所述估計的運動，變形所述第三參考幀的一個或多個像素，來生成變形幀。

【請求項20】 根據請求項19所述的方法，還包括：

至少部分地通過確定所述輸入幀和所述變形幀之間的差異來確定殘差；以及
使用所述殘差生成預測殘差。

【請求項21】 根據請求項20所述的方法，還包括：

基於所述變形幀和所述預測殘差，生成表示所述輸入幀的重建幀，所述重建幀包括雙向預測幀；以及
使所述重建幀儲存在記憶體中。

【請求項22】 根據請求項20所述的方法，其中，所述運動是使用第一自動編碼器估計的，並且其中所述預測殘差是使用第二自動編碼器生成的。

【請求項23】 根據請求項18所述的方法，還包括：

獲得第四參考幀；
估計另外輸入幀和所述第四參考幀之間的運動；
至少部分地通過基於所述估計的運動來變形所述第四參考幀的一個或多個像素，生成另外的變形幀；以及
基於所述另外的變形幀和另外的預測殘差，生成表示所述另外輸入幀的重建幀，表示所述另外輸入幀的所述重建幀包括單向預測幀。

【請求項24】 根據請求項18所述的方法，還包括：

確定表示從所述第一參考幀到所述第三參考幀的像素運動的第一組運動資訊；

確定表示從所述第二參考幀到所述第三參考幀的像素運動的第二組運動資訊；

至少部分地通過使用所述第一組運動資訊對所述第一參考幀執行變形函數來生成第一變形資訊；

至少部分地通過使用所述第二組運動資訊對所述第二參考幀執行所述變形函數來生成第二變形資訊；以及

基於所述第一變形資訊和所述第二變形資訊生成所述第三參考幀。

【請求項25】 根據請求項24所述的方法，其中，所述第一組運動資訊和所述第二組運動資訊是基於所述第一參考幀和所述第二參考幀之間的像素運動來確定的。

【請求項26】 根據請求項24所述的方法，其中，所述第一組運動資訊包括第一光流圖，並且所述第二組運動資訊包括第二光流圖。

【請求項27】 根據請求項24所述的方法，其中，所述變形函數包括雙線性插值函數。

【請求項28】 根據請求項19所述的方法，其中，所述輸入幀與所述第三參考幀之間的所述估計的運動包括光流資料。

【請求項29】 根據請求項19所述的方法，其中，所述輸入幀與所述第三參考幀之間的所述估計的運動包括動態卷積資料。

【請求項30】 根據請求項19所述的方法，其中，所述輸入幀與所述第三參考幀之間的所述估計的運動包括基於塊的運動資料。

【請求項31】 一種其上儲存有指令的非臨時性計算機可讀媒體，當所述指令被一個或多個處理器執行時，使所述一個或多個處理器執行以下操作：

獲得第一參考幀和第二參考幀；

至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成第三參考幀；以及

基於所述第三參考幀對輸入幀執行單向幀間預測。

【請求項32】 根據請求項31所述的非臨時性計算機可讀媒體，還包括當被所述一個或多個處理器執行時，使得所述一個或多個處理器執行以下操作的指令：

估計所述輸入幀和所述第三參考幀之間的運動；以及

至少部分地通過基於所述估計的運動，變形所述第三參考幀的一個或多個像素，來生成變形幀。

【請求項33】 根據請求項32所述的非臨時性計算機可讀媒體，還包括當被所述一個或多個處理器執行時，使得所述一個或多個處理器執行以下操作的指令：

至少部分地通過確定所述輸入幀和所述變形幀之間的差異來確定殘差；以及

使用所述殘差生成預測殘差。

【請求項34】 根據請求項33所述的非臨時性計算機可讀媒體，還包括當被所述一個或多個處理器執行時，使得所述一個或多個處理器執行以下操作的指令：

基於所述變形幀和所述預測殘差，生成表示所述輸入幀的重建幀，所述重建幀包括雙向預測幀；以及

使所述重建幀儲存在記憶體中。

【請求項35】 根據請求項33所述的非臨時性計算機可讀媒體，其中，所述運動是使用第一自動編碼器估計的，並且其中所述預測殘差是使用第二自動編碼器生成的。

【請求項36】 根據請求項31所述的非臨時性計算機可讀媒體，還包括當被所述一個或多個處理器執行時，使得所述一個或多個處理器執行以下操作的指令：

獲得第四參考幀；

估計另外輸入幀和所述第四參考幀之間的運動；

至少部分地通過基於所述估計的運動來變形所述第四參考幀的一個或多個像素，生成另外的變形幀；以及

基於所述另外的變形幀和另外的預測殘差，生成表示所述另外輸入幀的重建幀，表示所述另外輸入幀的所述重建幀包括單向預測幀。

【請求項37】 根據請求項31所述的非臨時性計算機可讀媒體，還包括當被所述一個或多個處理器執行時，使得所述一個或多個處理器執行以下操作的指令：

確定表示從所述第一參考幀到所述第三參考幀的像素運動的第一組運動資訊；

確定表示從所述第二參考幀到所述第三參考幀的像素運動的第二組運動資訊；

至少部分地通過使用所述第一組運動資訊對所述第一參考幀執行變形函數來生成第一變形資訊；

至少部分地通過使用所述第二組運動資訊對所述第二參考幀執行所述變形函數來生成第二變形資訊；以及

基於所述第一變形資訊和所述第二變形資訊生成所述第三參考幀。

【請求項38】 一種用於處理視頻資料的系統，所述系統包括：

記憶體；以及

耦接到所述記憶體的至少一個處理器，所述至少一個處理器被配置為：

獲得內插參考幀和表示輸入幀與所述內插參考幀之間的運動的運動資訊；

基於所述運動資訊，估計所述輸入幀和所述內插參考幀之間的所述運動；

至少部分地通過基於所述估計的運動，對所述內插參考幀的一個或多個像素進行變形來生成變形幀；以及

基於所述變形幀和預測殘差來生成表示所述輸入幀的重建幀，所述重建幀包括雙向預測幀。

【請求項39】 根據請求項38所述的系統，其中，所述至少一個處理器被配置為：

獲得第一參考幀和第二參考幀；以及

至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成所述內插參考幀。

【請求項40】 根據請求項38所述的系統，其中，所述至少一個處理器被配置為：

從位元流中獲取殘差；以及
基於獲取的殘差來生成所述預測殘差。

【請求項41】 根據請求項40所述的系統，其中，所述運動資訊是從所述位元流中獲得的。

【請求項42】 根據請求項38所述的系統，其中，所述至少一個處理器被配置為：

獲得第四參考幀；

估計另外輸入幀和所述第四參考幀之間的運動；

至少部分地通過基於所述估計的運動，變形所述第四參考幀的一個或多個像素來生成另外的變形幀；以及

基於所述另外的變形幀和另外的預測殘差，生成表示所述另外輸入幀的重建幀，表示所述另外輸入幀的所述重建幀包括單向預測幀。

【請求項43】 一種用於處理視頻資料的方法，所述方法包括：

獲得內插參考幀和表示輸入幀與所述內插參考幀之間的運動的運動資訊；

基於所述運動資訊，估計所述輸入幀和所述內插參考幀之間的所述運動；

至少部分地通過基於所述估計的運動，對所述內插參考幀的一個或多個像素進行變形來生成變形幀；以及

基於所述變形幀和預測殘差來生成表示所述輸入幀的重建幀，所述重建幀包括雙向預測幀。

【請求項44】 根據請求項43所述的方法，還包括：

獲得第一參考幀和第二參考幀；以及

至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成所述內插參考幀。

【請求項45】 根據請求項43所述的方法，還包括：

從位元流中獲取殘差；以及

基於所獲取的殘差來生成所述預測殘差。

【請求項46】 根據請求項45所述的方法，其中，所述運動資訊是從所述位元流中獲得的。

【請求項47】 根據請求項43所述的方法，還包括：

獲得第四參考幀；

估計另外輸入幀和所述第四參考幀之間的運動；

至少部分地通過基於所述估計的運動，變形所述第四參考幀的一個或多個像素來生成另外的變形幀；以及

基於所述另外的變形幀和另外的預測殘差，生成表示所述另外輸入幀的重建幀，表示所述另外輸入幀的所述重建幀包括單向預測幀。

【請求項48】 一種其上儲存有指令的非臨時性計算機可讀媒體，當所述指令被一個或多個處理器執行時，使所述一個或多個處理器執行以下操作：

獲得內插參考幀和表示輸入幀與所述內插參考幀之間的運動的運動資訊；

基於所述運動資訊，估計所述輸入幀和所述內插參考幀之間的所述運動；

至少部分地通過基於所述估計的運動，對所述內插參考幀的一個或多個像素進行變形來生成變形幀；以及

基於所述變形幀和預測殘差來生成表示所述輸入幀的重建幀，所述重建幀包括雙向預測幀。

【請求項49】 根據請求項48所述的非臨時性計算機可讀媒體，還包括當被所述一個或多個處理器執行時，使得所述一個或多個處理器執行以下操作的指令：

獲得第一參考幀和第二參考幀；以及

至少部分地通過在所述第一參考幀和所述第二參考幀之間執行插值來生成所述內插參考幀。

【請求項50】 根據請求項48所述的非臨時性計算機可讀媒體，還包括當被所述一個或多個處理器執行時，使得所述一個或多個處理器執行以下操作的指令：

從位元流中獲取殘差；以及

基於所獲取的殘差來生成所述預測殘差。

【請求項51】 根據請求項50所述的非臨時性計算機可讀媒體，其中，所述運動資訊是從所述位元流中獲得的。

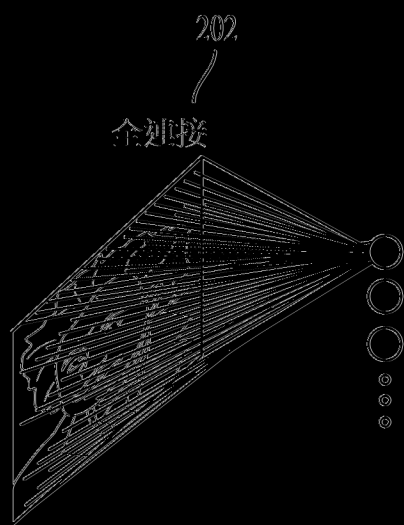
【請求項52】 根據請求項48所述的非臨時性計算機可讀媒體，還包括當被所述一個或多個處理器執行時，使得所述一個或多個處理器執行以下操作的指令：

獲得第四參考幀；

估計另外輸入幀和所述第四參考幀之間的運動；

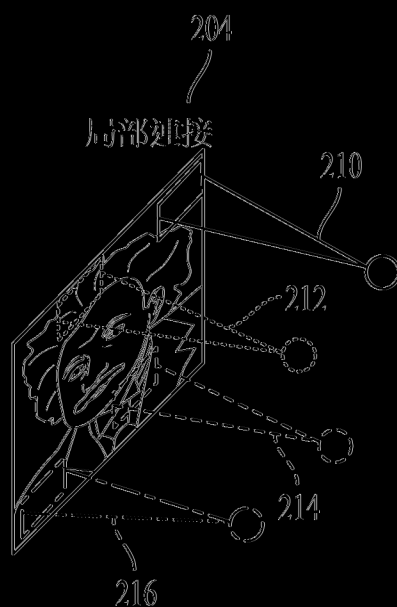
至少部分地通過基於所述估計的運動，變形所述第四參考幀的一個或多個像素來生成另外的變形幀；以及

基於所述另外的變形幀和另外的預測殘差，生成表示所述另外輸入幀的重建幀，表示所述另外輸入幀的所述重建幀包括單向預測幀。



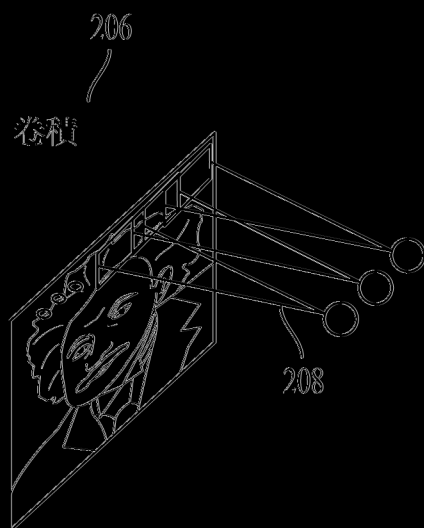
全連接

[(^{3 6}_{1 4})] 2A)



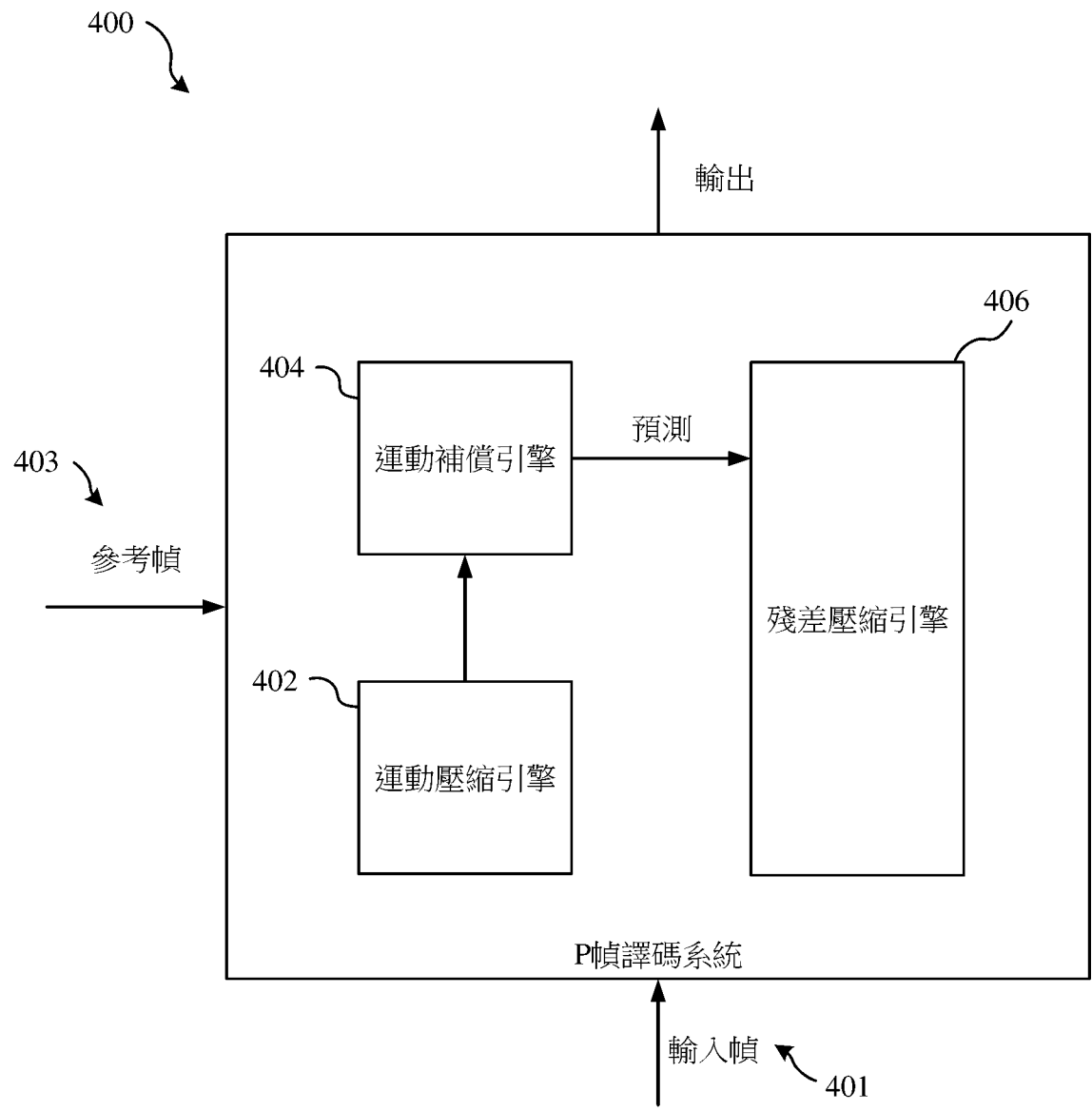
局部連接

[(^{3 6}_{1 4})] 2B)

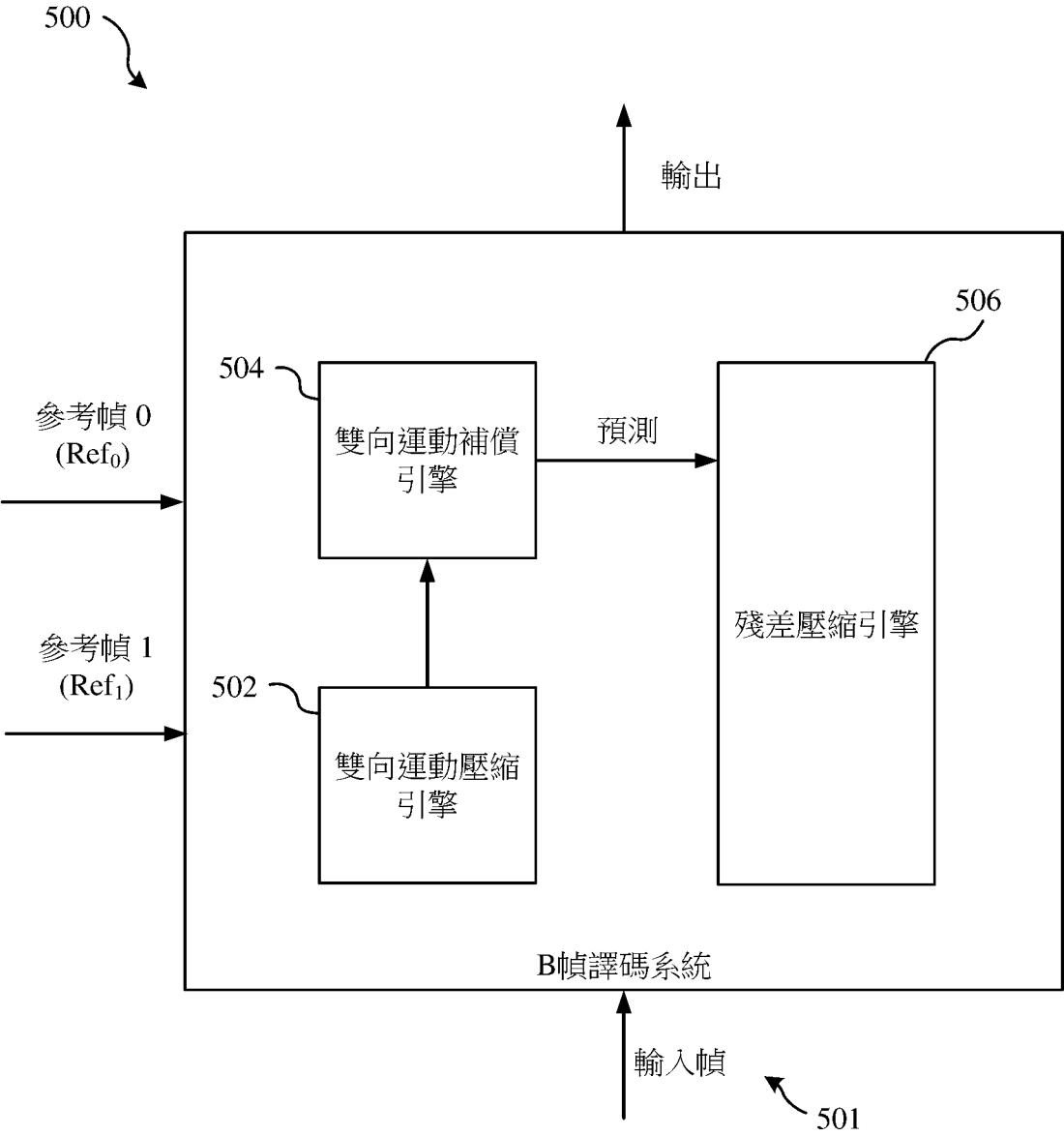


卷積

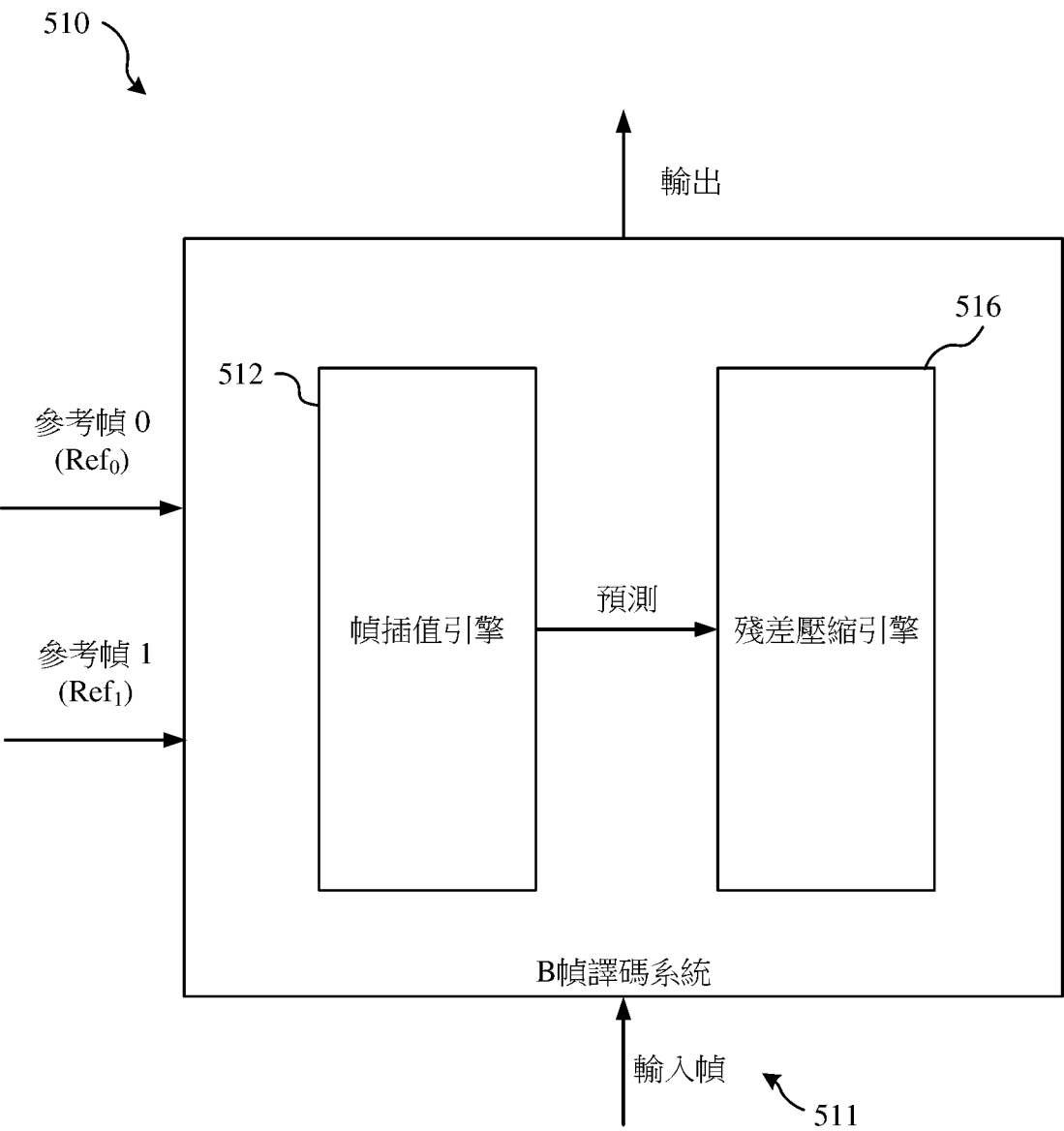
[(^{3 6}_{1 4})] 2C)



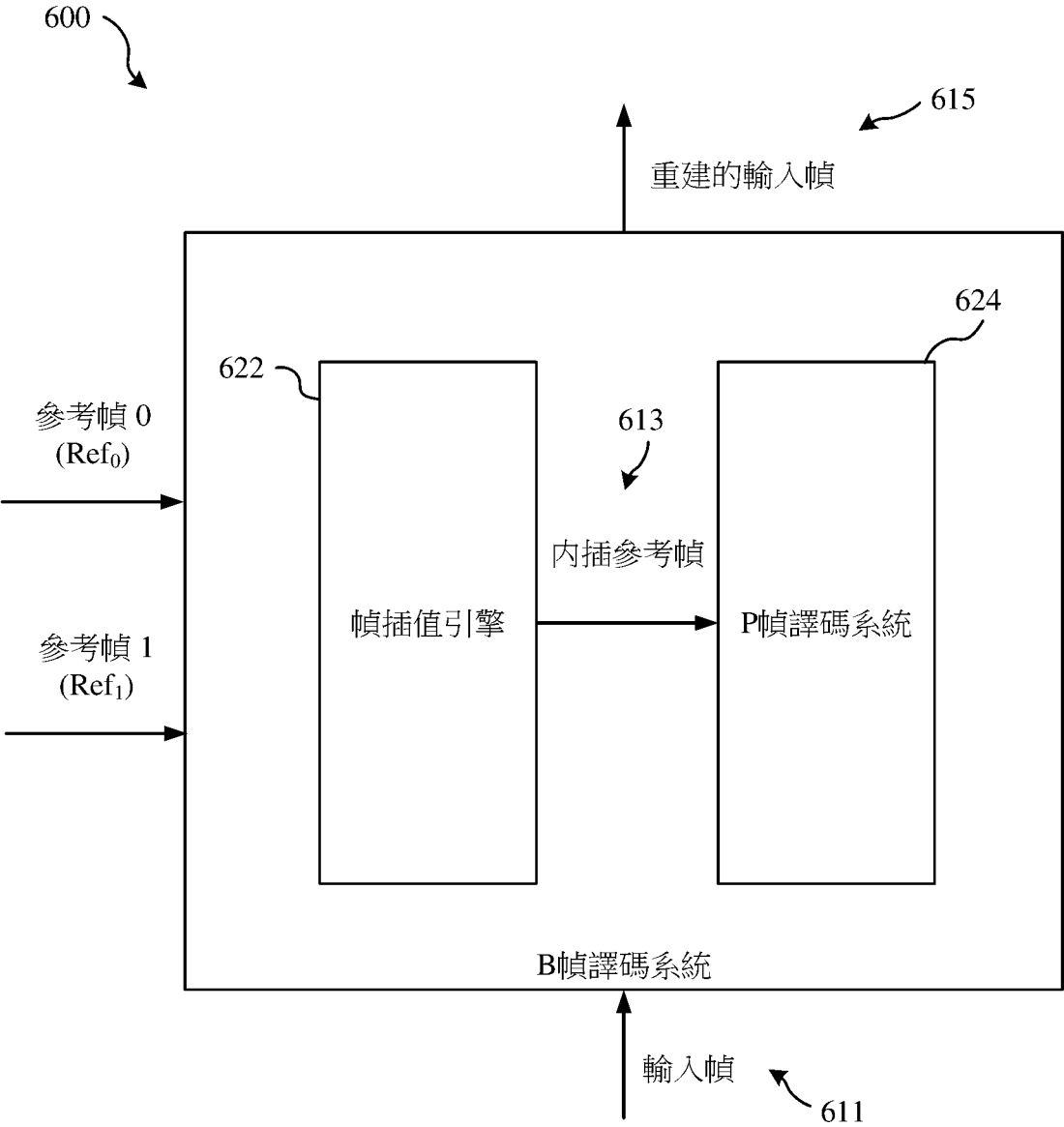
【圖 4】



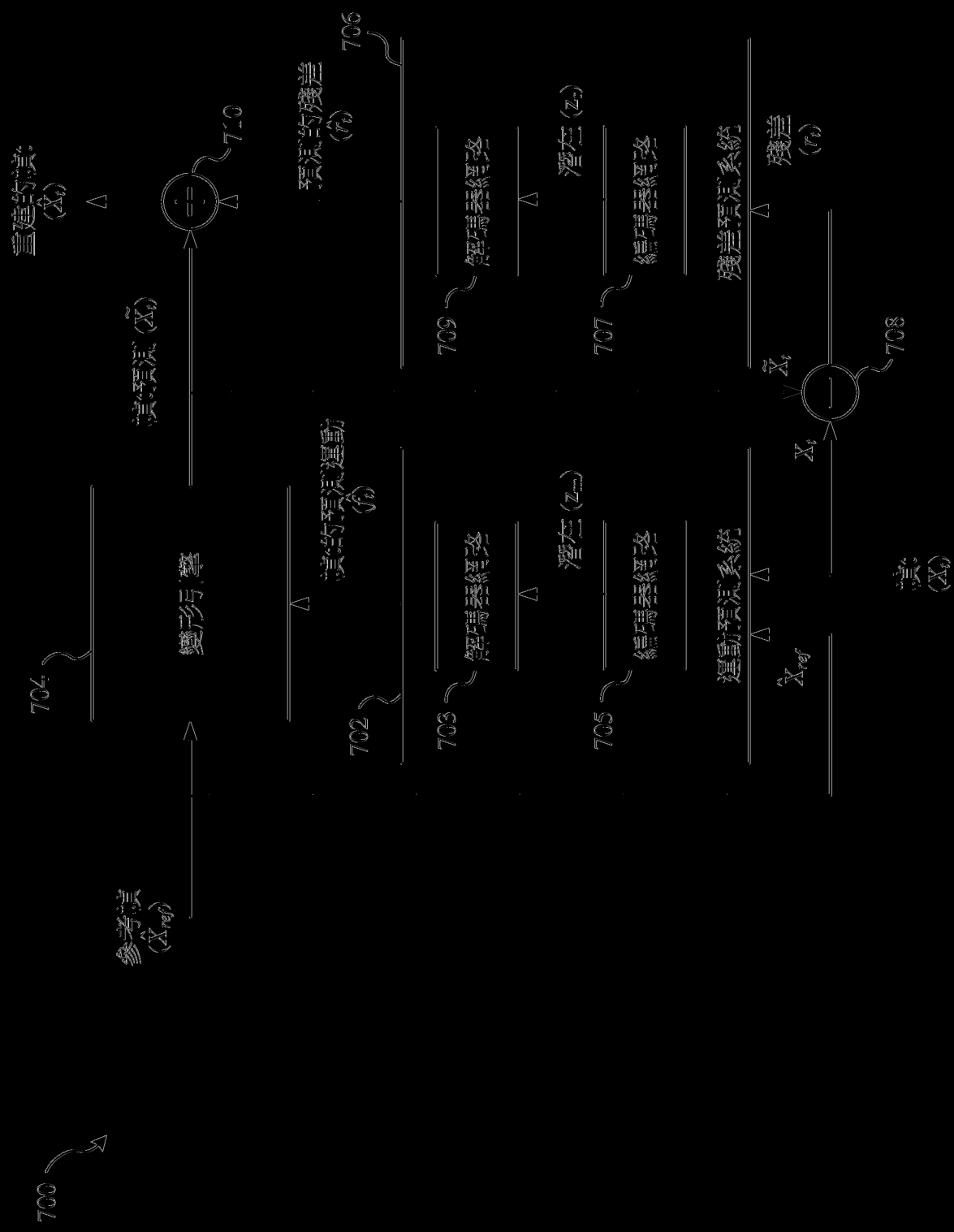
【圖 5A】



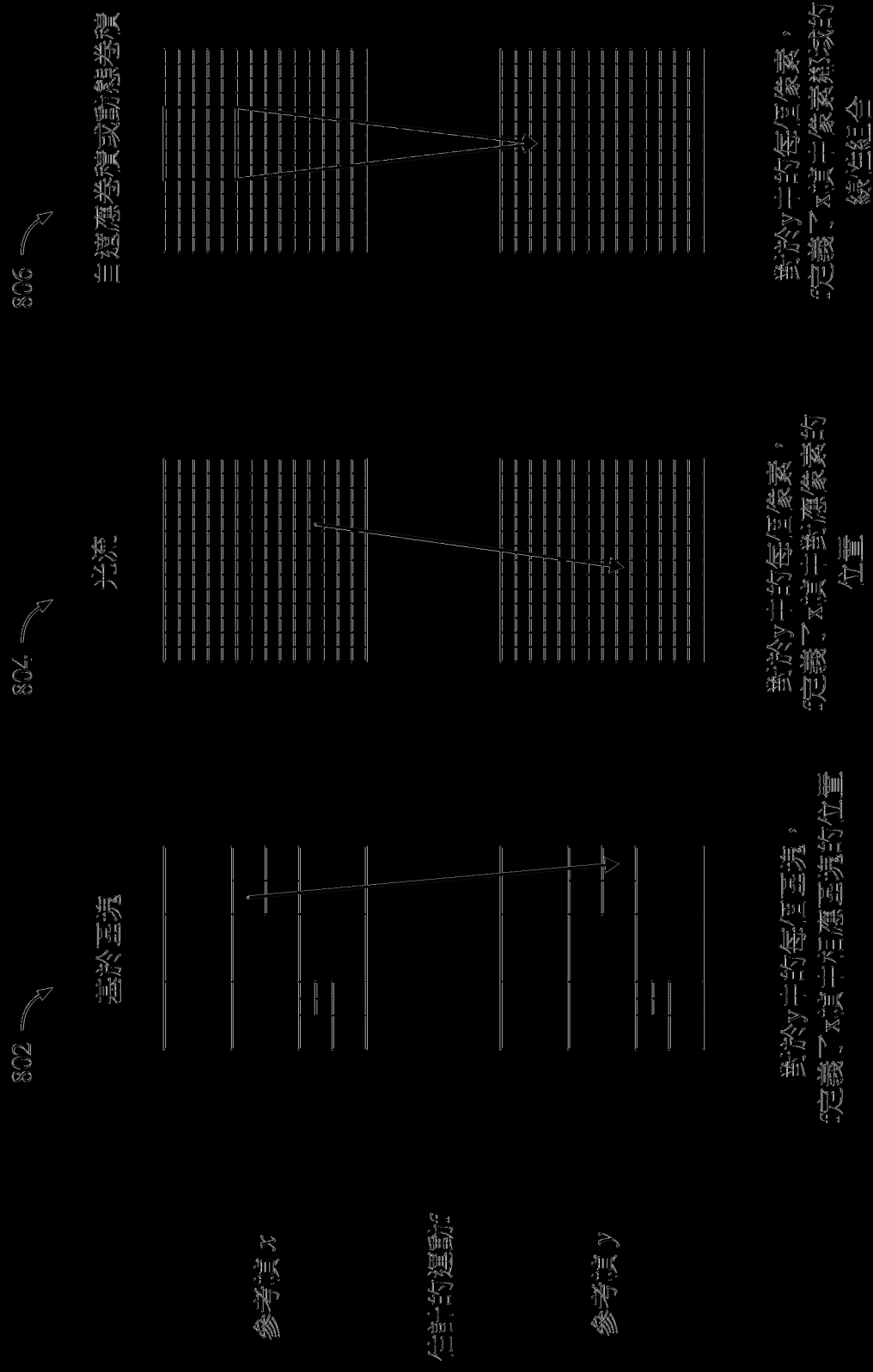
【圖 5B】



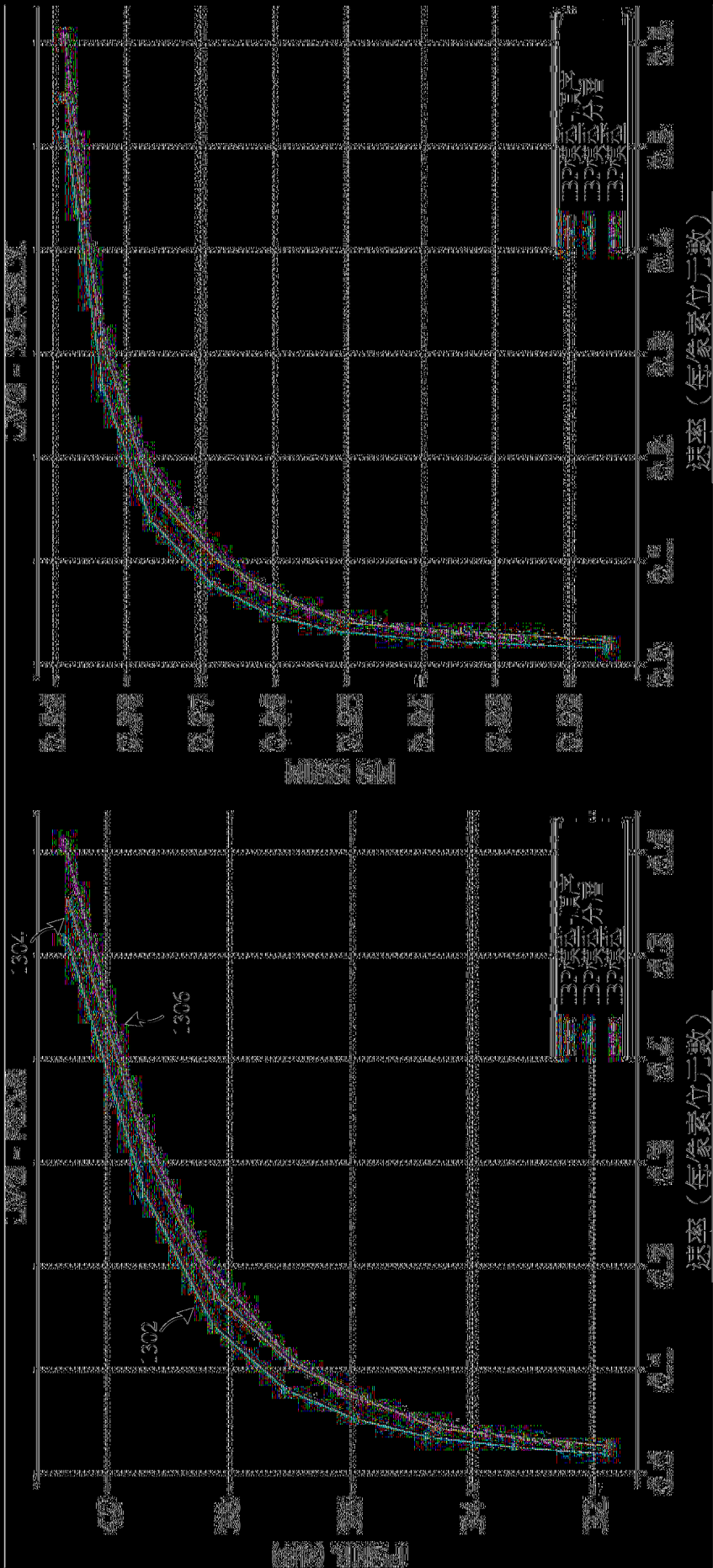
【圖 6】



[圖 7A]

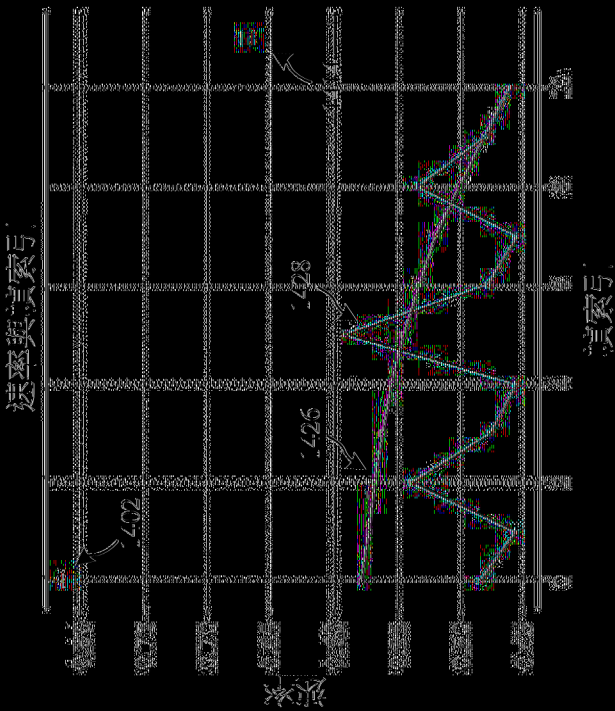


[圖 8]

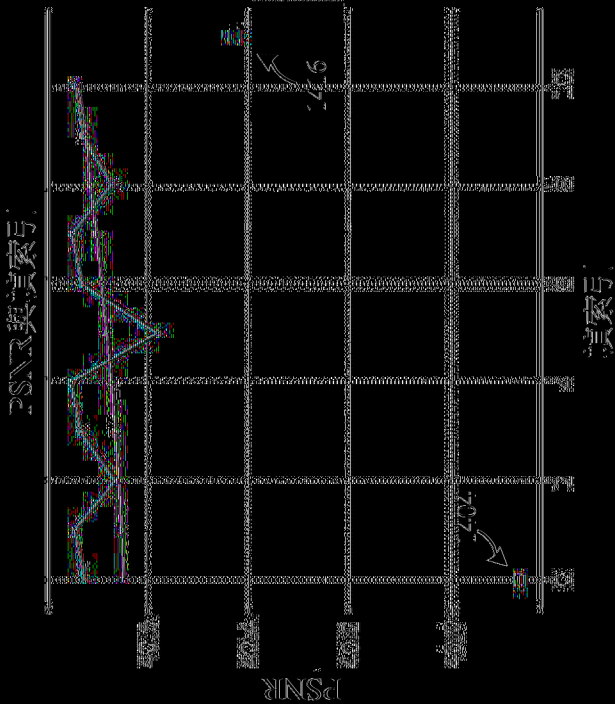


[圖 13B]

[圖 13A]

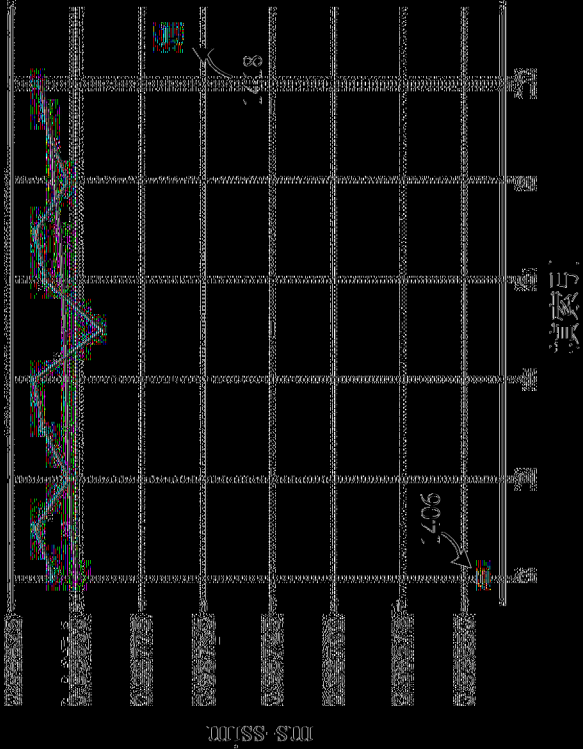


[圖 14A]



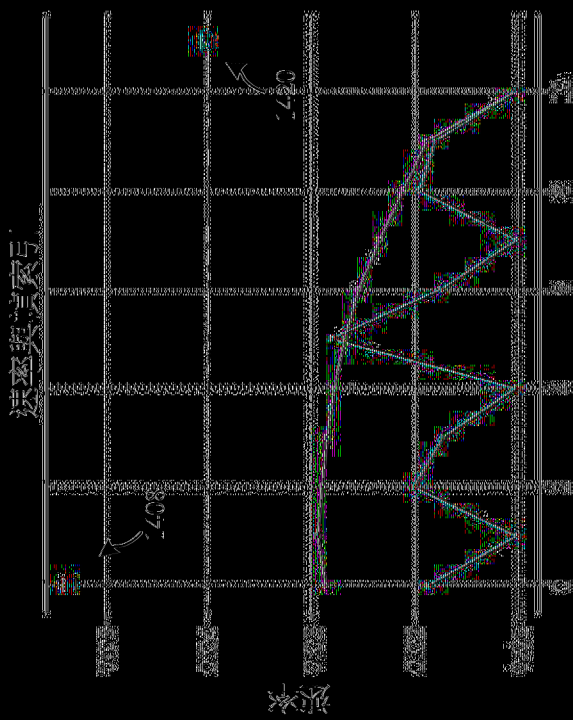
[圖 14B]

速率與搜索



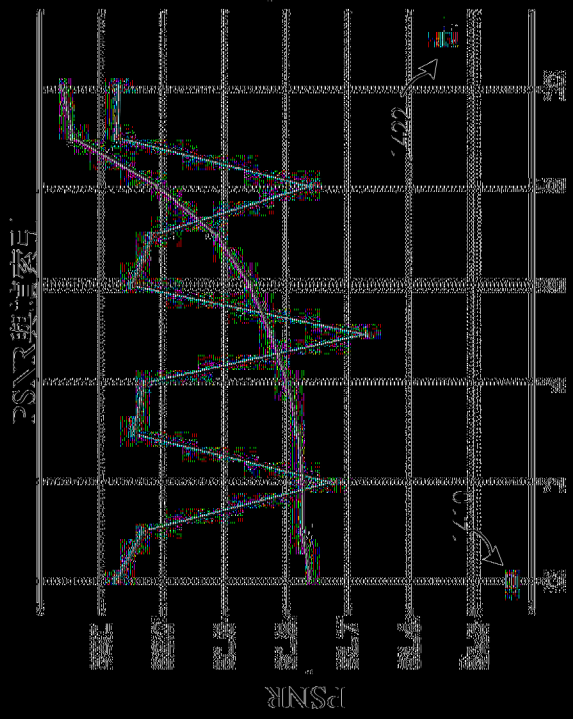
[圖 14C]

	1-速率 0.0001
	2-速率 0.0001
	3-速率 0.0001
	3-層次結構-速率 0.0001



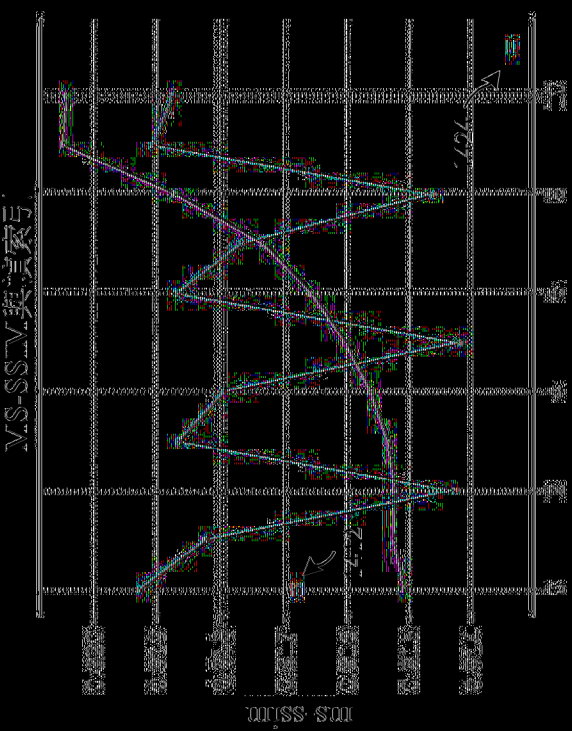
誤差

〔圖 14D〕



誤差

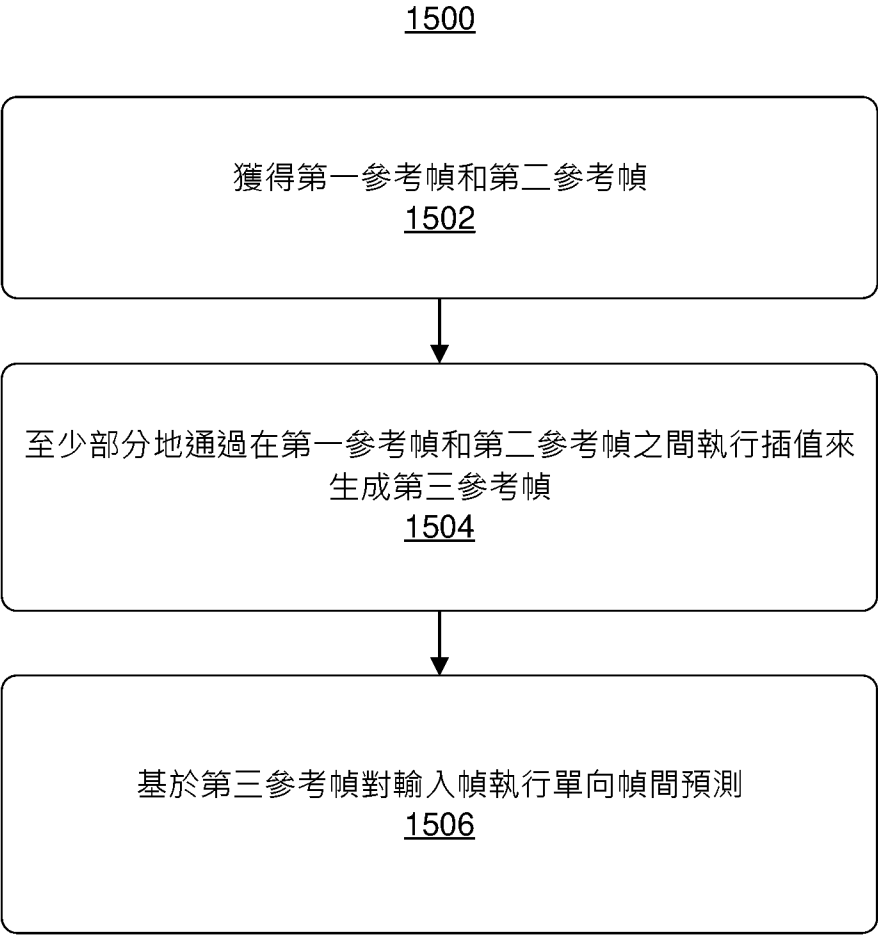
〔圖 14E〕



誤差

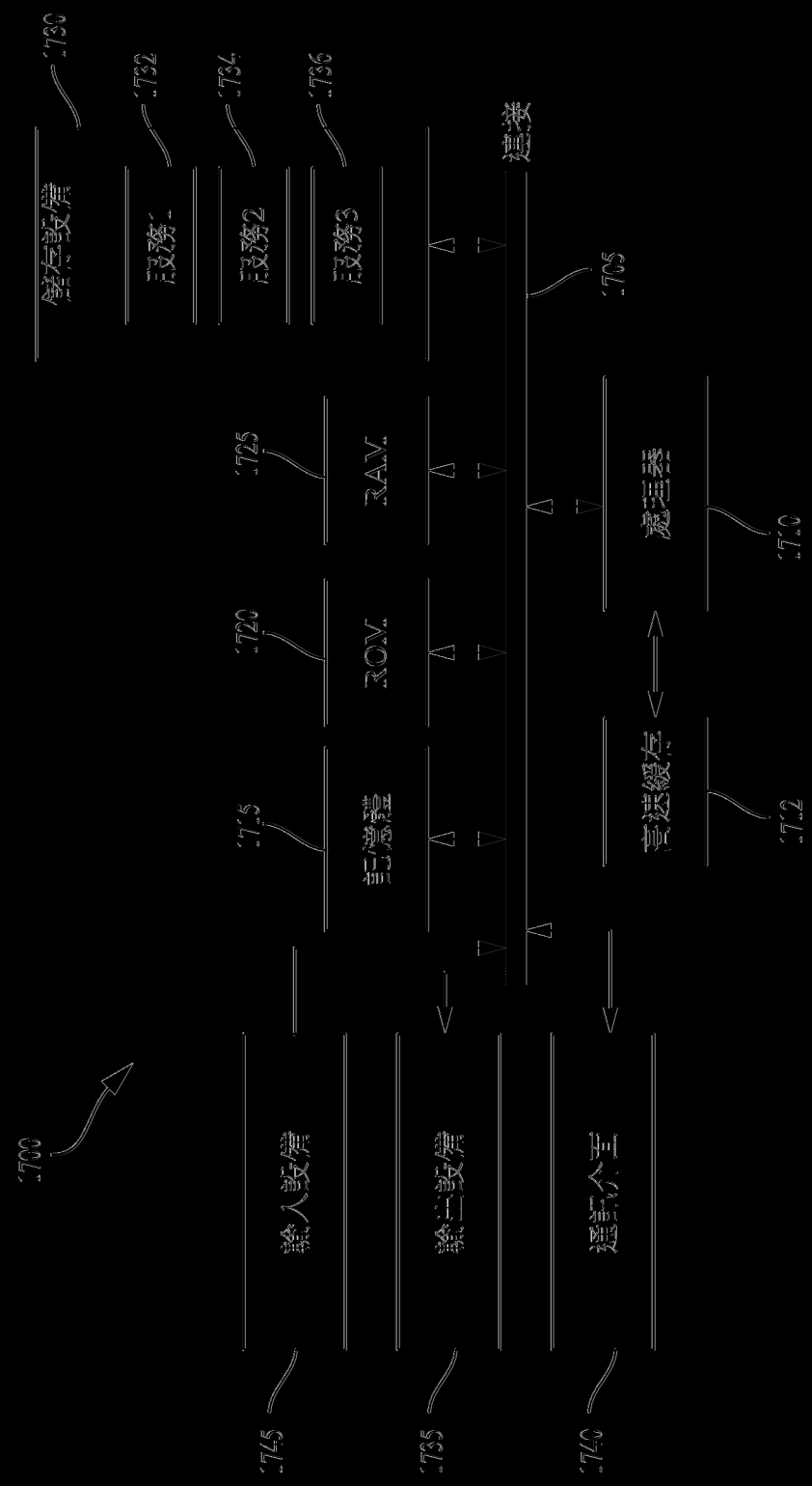
〔圖 14F〕

1	5eta 0.0128
2	5eta 0.0128
3	順序-5eta 0.0128
3	層次結構-5eta 0.0128



【圖 15】





[圖 17]