

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7628550号
(P7628550)

(45)発行日 令和7年2月10日(2025.2.10)

(24)登録日 令和7年1月31日(2025.1.31)

(51)国際特許分類

F I

H O 4 N 19/50 (2014.01)

H O 4 N 19/90 (2014.01)

G O 6 N 3/0455(2023.01)

H O 4 N 19/50

H O 4 N 19/90

G O 6 N 3/0455

請求項の数 15 (全102頁)

(21)出願番号	特願2022-551741(P2022-551741)	(73)特許権者	595020643
(86)(22)出願日	令和3年1月15日(2021.1.15)		クゥアルコム・インコーポレイテッド
(65)公表番号	特表2023-517846(P2023-517846 A)		QUALCOMM INCORPORATED
(43)公表日	令和5年4月27日(2023.4.27)		アメリカ合衆国、カリフォルニア州 9
(86)国際出願番号	PCT/US2021/013599		2121-1714、サン・ディエゴ、
(87)国際公開番号	WO2021/178050		モアハウス・ドライブ 5775
(87)国際公開日	令和3年9月10日(2021.9.10)	(74)代理人	110003708
審査請求日	令和5年12月15日(2023.12.15)		弁理士法人鈴榮特許総合事務所
(31)優先権主張番号	62/984,673	(74)代理人	100108855
(32)優先日	令和2年3月3日(2020.3.3)		弁理士 蔵田 昌俊
(33)優先権主張国・地域又は機関	米国(US)	(74)代理人	100158805
(31)優先権主張番号	17/091,570		弁理士 井関 守三
(32)優先日	令和2年11月6日(2020.11.6)	(74)代理人	100112807
	最終頁に続く		弁理士 岡田 貴志
			最終頁に続く

(54)【発明の名称】 再帰ベースの機械学習システムを使用したビデオ圧縮

(57)【特許請求の範囲】

【請求項1】

ビデオデータを処理する方法であって、

ニューラルネットワークシステムのエンコーダ部分によって、前記ニューラルネットワークシステムの現在の時間ステップの動作のための入力ビデオフレームと、前記ニューラルネットワークシステムの前の時間ステップの動作中に前記ニューラルネットワークシステムのデコーダ部分によって生成された前に再構成されたビデオフレームと、インター予測のための前記前の時間ステップの動作からの前記前に再構成されたビデオフレームの再構成された動き推定データと、インター予測のための前記前の時間ステップの動作中に前記ニューラルネットワークシステムの前記デコーダ部分によって生成された前に再構成されたビデオフレームの再構成された残差データと、前記前の時間ステップの動作からの前記ニューラルネットワークシステムの前記デコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと、

前記ニューラルネットワークシステムの前記エンコーダ部分によって、前記ニューラルネットワークシステムの前記現在の時間ステップの動作のための出力データを生成することと、前記出力データは、前記入力ビデオフレームと、前記前の時間ステップの動作からの前記前に再構成されたビデオフレームと、前記前の時間ステップの動作からの前記再構成された動き推定データと、前記前の時間ステップの動作からの前記再構成された残差データと、前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて生成され、

前記ニューラルネットワークシステムの前記デコーダ部分によって、前記現在の時間ステップの動作のための前記出力データと、前記前の時間ステップの動作からの前記デコーダ部分の前記少なくとも1つの再帰レイヤからの前記再帰状態データとを取得することと、

前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、前記再構成されたビデオフレームは、前記入力ビデオフレームを表し、

前記ニューラルネットワークシステムの動き推定ネットワークによって、前記入力ビデオフレームと、前記前の時間ステップの動作からの前に再構成されたビデオフレームまたは前記前の時間ステップの動作からの前の入力ビデオフレームのうちの少なくとも1つと、
に少なくとも部分的に基づいて、前記現在の時間ステップの動作のための動き推定データを決定することと、

10

前記現在の時間ステップのために決定された前記動き推定データを使用して、前記前に再構成されたビデオフレームまたは前記前の入力ビデオフレームのうちの少なくとも1つの1つまたは複数のピクセルを修正することに少なくとも部分的によって、前記現在の時間ステップのためのワーピングした再構成されたビデオフレームを生成することと、

前記ニューラルネットワークシステムの前記エンコーダ部分に、前記ワーピングした再構成されたビデオフレームと前記現在の時間ステップの動作のために決定された前記動き推定データとを送ることと、

1つまたは複数のトレーニング反復について、第1の損失関数を使用して前記ニューラルネットワークシステムをトレーニングすることと、前記第1の損失関数は、1つまたは複数の入力ビデオフレームと1つまたは複数のワーピングした再構成されたビデオフレームとの間の損失を決定し、

20

前記第1の損失関数を非活性化することと、

前記1つまたは複数のトレーニング反復の後に実施される1つまたは複数の後続のトレーニング反復について、第2の損失関数を使用して前記ニューラルネットワークシステムをトレーニングすることと、前記第2の損失関数は、少なくとも1つの入力ビデオフレームと少なくとも1つの再構成されたビデオフレームとの間の損失を決定し、

を備え、

前記現在の時間ステップの動作のために生成された前記出力データは、前記現在の時間ステップの動作のために決定された前記動き推定データと前記ワーピングした再構成されたビデオフレームとに少なくとも部分的に基づいて、前記ニューラルネットワークシステムの前記エンコーダ部分によって生成される、

30

方法。

【請求項2】

前記ニューラルネットワークシステムの前記デコーダ部分によって、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための動き推定データを決定することと、

前記現在の時間ステップのために決定された前記動き推定データを使用して、前記前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することに少なくとも部分的によって、前記現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと、

40

をさらに備える、請求項1に記載の方法。

【請求項3】

前記ニューラルネットワークシステムの前記デコーダ部分によって、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成された残差データを決定することと、

前記現在の時間ステップのために決定された前記再構成された残差データに前記ワーピングした再構成されたビデオフレームを加算することに少なくとも部分的によって、前記現在の時間ステップの動作のための前記再構成されたビデオフレームを生成することと、

50

をさらに備える、請求項 2 に記載の方法。

【請求項 4】

前記ニューラルネットワークシステムの前記デコーダ部分によって、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成された残差データを決定することと、

前記デコーダ部分のポストワーピングネットワークによって、前記ワーピングした再構成されたビデオフレームと前記現在の時間ステップのために決定された前記再構成された残差データとを処理することと、

をさらに備え、

前記再構成されたビデオフレームは、前記ポストワーピングネットワークによる前記ワーピングした再構成されたビデオフレームと前記再構成された残差データとの処理に基づいて、前記現在の時間ステップの動作のために生成される、

請求項 2 に記載の方法。

【請求項 5】

前記再帰状態データは、前記少なくとも 1 つの再帰レイヤのノードの 1 つまたは複数の値を含む、請求項 1 に記載の方法。

【請求項 6】

前記ニューラルネットワークシステムの前記デコーダ部分は、複数の再帰レイヤを含み、前記複数の再帰レイヤの各再帰レイヤは、それぞれの再帰状態データを与える、請求項 1 に記載の方法。

【請求項 7】

ビデオデータを処理するための装置であって、

メモリと、

前記メモリに結合されたプロセッサと、

を備え、前記プロセッサは、

ニューラルネットワークシステムのエンコーダ部分を使用して、前記ニューラルネットワークシステムの現在の時間ステップの動作のための入力ビデオフレームと、前記ニューラルネットワークシステムの前の時間ステップの動作中に前記ニューラルネットワークシステムのデコーダ部分によって生成された前に再構成されたビデオフレームと、インター予測のための前記前の時間ステップからの前記前に再構成されたビデオフレームの再構成された動き推定データと、インター予測のための前記前の時間ステップの動作中に前記ニューラルネットワークシステムの前記デコーダ部分によって生成された前記前に再構成されたビデオフレームの再構成された残差データと、前記前の時間ステップの動作からの前記ニューラルネットワークシステムの前記デコーダ部分の少なくとも 1 つの再帰レイヤからの再帰状態データとを取得することと、

前記エンコーダ部分を使用して、前記ニューラルネットワークシステムの前記現在の時間ステップの動作のための出力データを生成することと、前記出力データは、前記入力ビデオフレームと、前記前の時間ステップの動作からの前記前に再構成されたビデオフレームと、前記前の時間ステップの動作からの前記再構成された動き推定データと、前記前の時間ステップの動作からの前記再構成された残差データと、前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて生成され、

ニューラルネットワークシステムの前記デコーダ部分を使用して、前記現在の時間ステップの動作のための前記出力データと、前記前の時間ステップの動作からの前記デコーダ部分の前記少なくとも 1 つの再帰レイヤからの前記再帰状態データとを取得することと、

前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、前記再構成されたビデオフレームは、前記入力ビデオフレームを表し、

前記ニューラルネットワークシステムの動き推定ネットワークを使用して、前記入力ビデオフレームと、前記前の時間ステップの動作からの前に再構成されたビデオフレームまた

10

20

30

40

50

は前記前の時間ステップの動作からの前の入力ビデオフレームのうちの少なくとも1つと、
に少なくとも部分的に基づいて、前記現在の時間ステップの動作のための動き推定データを決定することと、

前記現在の時間ステップのために決定された前記動き推定データを使用して、前記前に再構成されたビデオフレームまたは前記前の入力ビデオフレームのうちの少なくとも1つの1つまたは複数のピクセルを修正することによって、前記現在の時間ステップのためのワーピングした再構成されたビデオフレームを生成することと、

前記ニューラルネットワークシステムの前記エンコーダ部分に、前記ワーピングした再構成されたビデオフレームと前記現在の時間ステップの動作のために決定された前記動き推定データとを送ることと、

1つまたは複数のトレーニング反復について、第1の損失関数を使用して前記ニューラルネットワークシステムをトレーニングすることと、前記第1の損失関数は、1つまたは複数の入力ビデオフレームと1つまたは複数のワーピングした再構成されたビデオフレームとの間の損失を決定し、

前記第1の損失関数を非活性化することと、

前記1つまたは複数のトレーニング反復の後に実施される1つまたは複数の後続のトレーニング反復について、第2の損失関数を使用して前記ニューラルネットワークシステムをトレーニングすることと、前記第2の損失関数は、少なくとも1つの入力ビデオフレームと少なくとも1つの再構成されたビデオフレームとの間の損失を決定し、

を行うように構成され、

前記現在の時間ステップの動作のために生成された前記出力データは、前記現在の時間ステップの動作のために決定された前記動き推定データと前記ワーピングした再構成されたビデオフレームとに少なくとも部分的に基づいて、前記ニューラルネットワークシステムの前記エンコーダ部分によって生成される、

装置。

【請求項8】

請求項2乃至6のいずれか一項に記載の方法を実行するための手段をさらに備える、請求項7に記載の装置。

【請求項9】

ビデオデータを処理する方法であって、

ニューラルネットワークシステムのデコーダ部分によって、前記ニューラルネットワークシステムの現在の時間ステップの動作のための出力データと、前記ニューラルネットワークシステムの前の時間ステップの動作からの前記デコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと、前記出力データは、前記現在の時間ステップの動作のための入力ビデオフレームと、前記前の時間ステップの動作中に生成された前に再構成されたビデオフレームと、インター予測のための前記前の時間ステップの動作からの前記前に再構成されたビデオフレームの再構成された動き推定データと、インター予測のための前記前の時間ステップの動作中に生成された前記前に再構成されたビデオフレームの再構成された残差データと、前記再帰状態データとに少なくとも部分的に基づいて生成され、

前記ニューラルネットワークシステムの前記デコーダ部分によって、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための動き推定データを決定することと、

前記前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと、

前記現在の時間ステップのために決定された前記動き推定データを使用して、前記前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することによって、前記現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと、

前記ワーピングした再構成されたビデオフレームに少なくとも部分的に基づいて、前記

10

20

30

40

50

現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、
1つまたは複数のトレーニング反復について、第1の損失関数を使用して前記ニューラル
ネットワークシステムをトレーニングすることと、前記第1の損失関数は、1つまたは複
数の入力ビデオフレームと1つまたは複数のワーピングした再構成されたビデオフレーム
との間の損失を決定し、

前記第1の損失関数を非活性化することと、

前記1つまたは複数のトレーニング反復の後に実施される1つまたは複数の後続のトレ
ーニング反復について、第2の損失関数を使用して前記ニューラルネットワークシステムを
トレーニングすることと、前記第2の損失関数は、少なくとも1つの入力ビデオフレーム
と少なくとも1つの再構成されたビデオフレームとの間の損失を決定し、

10

を備える方法。

【請求項10】

前記ニューラルネットワークシステムの前記デコーダ部分によって、前記出力データと
前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて
、前記現在の時間ステップの動作のための再構成された残差データを決定することと、

前記現在の時間ステップのために決定された前記再構成された残差データに前記ワー
ピングした再構成されたビデオフレームを加算することに少なくとも部分的によって、前記
現在の時間ステップの動作のための前記再構成されたビデオフレームを生成することと、
をさらに備える、請求項9に記載の方法。

【請求項11】

20

前記現在の時間ステップの動作中に、前記前の時間ステップの動作からの前記再構成さ
れた残差データを使用して、前記前に再構成されたビデオフレームを修正することと、

前記現在の時間ステップの動作のための最終の出力ビデオフレームとして、前記修正さ
れた前に再構成されたビデオフレームを出力することと、

をさらに備える、請求項10に記載の方法。

【請求項12】

前記ニューラルネットワークシステムの前記デコーダ部分によって、前記出力データと
前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて
、前記現在の時間ステップの動作のための再構成された残差データを決定することと、

前記デコーダ部分のポストワーピングネットワークによって、前記現在の時間ステップ
のために決定された前記再構成された残差データと前記ワーピングした再構成されたビデ
オフレームとを処理することと、

30

前記ポストワーピングネットワークによる前記ワーピングした再構成されたビデオフレ
ームと前記再構成された残差データとの処理に少なくとも部分的に基づいて、前記現在の
時間ステップの動作のための前記再構成されたビデオフレームを生成することと、

をさらに備える、請求項9に記載の方法。

【請求項13】

前記現在の時間ステップの動作のために決定された前記動き推定データまたは前記前の
時間ステップの動作のために決定された動き推定を使用して、前記前の時間ステップの動
作からの前記再帰状態データを修正することに少なくとも部分的によって、前記現在の時
間ステップの動作のためのワーピングした再帰状態データを生成することと、

40

前記ワーピングした再帰状態データに少なくとも部分的に基づいて、前記現在の時間ス
テップの動作のための前記再構成されたビデオフレームを生成することと、

をさらに備える、請求項9に記載の方法。

【請求項14】

ビデオデータを処理するための装置であって、

メモリと、

前記メモリに結合されたプロセッサと、

を備え、前記プロセッサは、

ニューラルネットワークシステムのデコーダ部分を使用して、前記ニューラルネットワ

50

ークシステムの現在の時間ステップの動作のための出力データと、前記ニューラルネットワークシステムの前の時間ステップの動作からの前記デコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと、前記出力データは、前記現在の時間ステップの動作のための入力ビデオフレームと、前記前の時間ステップの動作中に生成された前に再構成されたビデオフレームと、インター予測のための前記前の時間ステップの動作からの前記前に再構成されたビデオフレームの再構成された動き推定データと、前記前の時間ステップの動作中に生成された再構成された残差データと、前記再帰状態データとに少なくとも部分的に基づいて生成され、

前記ニューラルネットワークシステムの前記デコーダ部分を使用して、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための動き推定データを決定することと、

前記前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと、

前記現在の時間ステップのために決定された前記動き推定データを使用して、前記前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することに少なくとも部分的によって、前記現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと、

前記ワーピングした再構成されたビデオフレームに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、

1つまたは複数のトレーニング反復について、第1の損失関数を使用して前記ニューラルネットワークシステムをトレーニングすることと、前記第1の損失関数は、1つまたは複数の入力ビデオフレームと1つまたは複数のワーピングした再構成されたビデオフレームとの間の損失を決定し、

前記第1の損失関数を非活性化することと、

前記1つまたは複数のトレーニング反復の後に実施される1つまたは複数の後続のトレーニング反復について、第2の損失関数を使用して前記ニューラルネットワークシステムをトレーニングすることと、前記第2の損失関数は、少なくとも1つの入力ビデオフレームと少なくとも1つの再構成されたビデオフレームとの間の損失を決定し、

を行うように構成された、装置。

【請求項15】

請求項10乃至13のいずれか一項に記載の方法を実行するための手段をさらに備える、請求項14に記載の装置。

【発明の詳細な説明】

【技術分野】

【0001】

[0001]本開示は、概して、ビデオの符号化（または圧縮）と復号（復元）とを含むビデオコーディングに関する。たとえば、本開示の態様は、1つまたは複数の再帰ベースの機械学習システムを使用してビデオコンテンツをコーディング（たとえば、符号化および／または復号）することに関する。

【背景技術】

【0002】

[0002]多くのデバイスおよびシステムは、ビデオデータが消費のために処理および出力されることを可能にする。デジタルビデオデータは、消費者およびビデオプロバイダの需要を満たすための大量のデータを含む。たとえば、ビデオデータの消費者は、高い忠実度、解像度、フレームレートなどを含む高い品質のビデオを望む。その結果、これらの需要を満たすために必要とされる大量のビデオデータは、ビデオデータを処理および記憶する通信ネットワークおよびデバイスに負担をかける。

【0003】

[0003]ビデオコーディング技法が、ビデオデータを圧縮するために使用され得る。ビデオコーディングの目標は、ビデオ品質に対する劣化を回避するかまたは最小限に抑えなが

10

20

30

40

50

ら、より低いビットレートを使用する形式にビデオデータを圧縮することである。常に発展しつつあるビデオサービスが利用可能になるとともに、より良いコーディング効率をもつ符号化技法が必要とされる。

【発明の概要】

【0004】

[0004] 1つまたは複数の再帰ベースの機械学習システムを使用してビデオコンテンツをコーディング（たとえば、符号化および／または復号）するためのシステムおよび技法について説明される。1つの例示的な例では、ビデオデータを処理する方法が提供される。本方法は、ニューラルネットワークシステムのエンコーダ部分によって、ニューラルネットワークシステムの現在の時間ステップの動作のための入力ビデオフレームと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された動き推定データと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からのニューラルネットワークシステムのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと；ニューラルネットワークシステムのエンコーダ部分によって、ニューラルネットワークシステムの現在の時間ステップの動作のための出力データを生成することと、出力データは、入力ビデオフレームと、前の時間ステップの動作からの再構成された動き推定データと、前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて生成される；ニューラルネットワークシステムのデコーダ部分によって、現在の時間ステップの動作のための出力データと、前の時間ステップの動作からのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データと、を取得することと；出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、再構成されたビデオフレームは、入力ビデオフレームを表す、を含む。

【0005】

[0005] 別の例では、メモリと、メモリと結合された（たとえば、回路中に実装される）プロセッサを含む、ビデオデータを処理するための装置が提供される。いくつかの例では、2つ以上のプロセッサは、メモリに結合され得、動作のうちの1つまたは複数を実施するように使用され得る。プロセッサは、ニューラルネットワークシステムのエンコーダ部分を使用して、ニューラルネットワークシステムの現在の時間ステップの動作のための入力ビデオフレームと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された動き推定データと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からのニューラルネットワークシステムのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと；エンコーダ部分を使用して、ニューラルネットワークシステムの現在の時間ステップの動作のための出力データを生成することと、出力データは、入力ビデオフレームと、前の時間ステップの動作からの再構成された動き推定データと、前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて生成される；ニューラルネットワークシステムのデコーダ部分を使用して、現在の時間ステップの動作のための出力データと前の時間ステップの動作からのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと；出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、再構成されたビデオフレームは、入力ビデオフレームを表す、を行うように構成される。

【0006】

[0006] 別の例では、非一時的コンピュータ可読媒体であって、1つまたは複数のプロセッサによって実行されたとき、1つまたは複数のプロセッサに、ニューラルネットワークシステムのエンコーダ部分を使用して、ニューラルネットワークシステムの現在の時間ス

テップの動作のための入力ビデオフレームと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された動き推定データと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からのニューラルネットワークシステムのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと；ニューラルネットワークシステムのエンコーダ部分を使用して、ニューラルネットワークシステムの現在の時間ステップの動作のための出力データを生成することと、出力データは、入力ビデオフレームと、前の時間ステップの動作からの再構成された動き推定データと、前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて生成される；ニューラルネットワークシステムのデコーダ部分を使用して、現在の時間ステップの動作のための出力データと前の時間ステップの動作からのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと；出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、デコーダ部分を使用して、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、再構成されたビデオフレームは、入力ビデオフレームを表す、を行わせる命令を記憶した非一時的コンピュータ可読媒体が提供される。

【0007】

【0007】別の例では、ビデオデータを処理するための装置が提供される。本装置は、ニューラルネットワークシステムの現在の時間ステップの動作のための入力ビデオフレームと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された動き推定データと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からの少なくとも1つの再帰レイヤからの再帰状態データとを取得するための手段と；ニューラルネットワークシステムの現在の時間ステップの動作のための出力データを生成するための手段と、出力データは、入力ビデオフレームと、前の時間ステップの動作からの再構成された動き推定データと、前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて生成される；現在の時間ステップの動作のための出力データと、前の時間ステップの動作からのニューラルネットワークシステムのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得するための手段と；出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成するための手段と、再構成されたビデオフレームは、入力ビデオフレームを表す、を含む。

【0008】

【0008】いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための動き推定データを決定することと；前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと；現在の時間ステップのために決定された動き推定データを使用して、前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することによって、現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと、をさらに備える。

【0009】

【0009】いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成された残差データを決定することと；現在の時間ステップのために決定された再構成された残差データにワーピングした再構成されたビデオフレームを加算することによって、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、をさらに備える。

【0010】

[0010]いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成された残差データを決定することと、デコーダ部分のポストワーピングネットワークによって、ワーピングした再構成されたビデオフレームと現在の時間ステップのために決定された再構成された残差データとを処理することとをさらに備える。そのような態様では、再構成されたビデオフレームは、ポストワーピングネットワークによるワーピングした再構成されたビデオフレームと再構成された残差データとの処理に基づいて現在の時間ステップの動作のために生成される。

【0011】

[0011]いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、ニューラルネットワークシステムの動き推定ネットワークによって、入力ビデオフレームと前の時間ステップの動作からの前に再構成されたビデオフレームまたは前の時間ステップの動作からの前の入力ビデオフレームとのうちの少なくとも1つとに少なくとも部分的に基づいて、現在の時間ステップの動作のための動き推定データを決定することと；現在の時間ステップのために決定された動き推定データを使用して、前に再構成されたビデオフレームまたは前の入力ビデオフレームとのうちの少なくとも1つの1つまたは複数のピクセルを修正することによって、現在の時間ステップのためのワーピングした再構成されたビデオフレームを生成することと；ニューラルネットワークシステムのエンコーダ部分に、ワーピングした再構成されたビデオフレームと、現在の時間ステップの動作のために決定された動き推定データとを送ることとをさらに備える。そのような態様では、現在の時間ステップの動作のために生成された出力データは、ワーピングした再構成されたビデオフレームと現在の時間ステップの動作のために決定された動き推定データとに少なくとも部分的に基づいて、ニューラルネットワークシステムのエンコーダ部分によって生成される。

【0012】

[0012]いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、1つまたは複数のトレーニング反復について、第1の損失関数を使用してニューラルネットワークシステムをトレーニングすることと、第1の損失関数は、1つまたは複数の入力ビデオフレームと1つまたは複数のワーピングした再構成されたビデオフレームとの間の損失を決定する；第1の損失関数を非活性化することと；1つまたは複数のトレーニング反復の後に実施される1つまたは複数の後続のトレーニング反復について、第2の損失関数を使用してニューラルネットワークシステムをトレーニングすることと、第2の損失関数は、少なくとも1つの入力ビデオフレームと少なくとも1つの再構成されたビデオフレームとの間の損失を決定する、をさらに備える。

【0013】

[0013]いくつかの態様では、再帰状態データは、少なくとも1つの再帰レイヤのノードの1つまたは複数の値を含む。

【0014】

[0014]いくつかの態様では、ニューラルネットワークシステムのデコーダ部分は、複数の再帰レイヤを含み、複数の再帰レイヤの各再帰レイヤは、それぞれの再帰状態データを与える。

【0015】

[0015]いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、現在の時間ステップの動作または前の時間ステップの動作のために決定された動き推定データを使用して前の時間ステップの動作からの再帰状態データを修正することによって現在の時間ステップの動作のためのワーピングした再帰状態データを生成することとをさらに備える。そのような態様では、現在の時間ステップの動作のために生成された再構成されたビデオフレームは、ワーピングした再帰状態データに少なくとも部分的に基づいて生成される。

【0016】

[0016]いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、デコーダ部分によって、再構成されたマスクを生成することと、再構成されたマスクは、前の時間ステップの動作中に生成された前に再構成されたビデオフレーム中のデータの存在を示す1つまたは複数の値を含む；再構成されたマスクと、現在の時間ステップの動作のために生成されるワーピングした再構成されたビデオフレームと、現在の時間ステップの動作のために決定された再構成された残差データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、再構成されたマスクは、ワーピングした再構成されたビデオフレームと再構成された残差データとの寄与を重み付けする、をさらに備える。

10

【0017】

[0017]いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと；現在の時間ステップの動作中に、現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することによって、現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと；現在の時間ステップの動作中にデコーダ部分によって、現在の時間ステップの動作のための第1の再構成された残差データを決定することと；現在の時間ステップのために決定された再構成された残差データにワーピングした再構成されたビデオフレームを加算することによって、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと；現在の時間ステップの動作中にデコーダ部分によって、現在の時間ステップの動作中にエンコーダ部分によって処理された前の入力ビデオフレームに少なくとも部分的に基づいて、前の時間ステップの動作からの第2の再構成された残差データを決定することと；現在の時間ステップの動作中に、第2の再構成された残差データを使用して前に再構成されたビデオフレームを修正することと；現在の時間ステップの動作のための最終の出力ビデオフレームとして、修正された前に再構成されたビデオフレームを出力することと、をさらに備える。

20

【0018】

[0018]いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、ニューラルネットワークシステムのエンコーダ部分によって、1つまたは複数のひずみマップを取得することと；ニューラルネットワークシステムのエンコーダ部分によって、1つまたは複数のひずみマップに少なくとも部分的に基づいて、出力データを生成することと、をさらに備える。いくつかの態様では、1つまたは複数のひずみマップは、現在の時間ステップの動作のための入力ビデオフレームと前の時間ステップの動作からの前に再構成されたビデオフレームとの間のひずみを示す第1のひずみマップを含む。いくつかの態様では、1つまたは複数のひずみマップは、現在の時間ステップの動作のための入力ビデオフレームとワーピングした再構成されたビデオフレームとの間のひずみを示す第2のひずみマップを含む。そのような態様では、ワーピングした再構成されたビデオフレームは、前の時間ステップの動作からの前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することによって生成され得る。

30

40

【0019】

[0019]いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、ニューラルネットワークシステムのエンコーダ部分の第1のレイヤによって、入力データを処理することと；エンコーダ部分の第1のレイヤによって、エンコーダ部分の第1のレイヤのための複数の重み値を決定することと；複数の重み値にニューラルネットワークシステムのエンコーダ部分の第2のレイヤの重みを設定することと；エンコーダ部分の第2のレイヤによって、複数の重み値に設定された重みを使用して入力データを処理することと、をさらに備える。

【0020】

[0020]いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒

50

体は、記憶媒体中に出力データを記憶することをさらに備える。

【0021】

[0021]いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、記憶媒体中に再構成されたビデオフレームを記憶することをさらに備える。

【0022】

[0022]いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、少なくとも1つのデバイスに送信媒体を介して再構成されたビデオフレームを送ることをさらに備える。

【0023】

[0023]別の例示的な例では、ビデオデータを処理する方法が提供される。本方法は、ニューラルネットワークシステムのデコーダ部分によって、ニューラルネットワークシステムの現在の時間ステップの動作のための出力データとニューラルネットワークシステムの前の時間ステップの動作からのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと；ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための動き推定データを決定することと；前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと；現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することに少なくとも部分的によって、現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと；ワーピングした再構成されたビデオフレームに少なくとも部分的に基づいて現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、を含む。

【0024】

[0024]別の例では、ビデオデータを処理するための装置は、メモリと、メモリと結合された（たとえば、回路中に実装される）プロセッサとを含む。いくつかの例では、2つ以上のプロセッサは、メモリに結合され得、動作のうちの1つまたは複数を実施するように使用され得る。プロセッサは、ニューラルネットワークシステムのデコーダ部分を使用して、ニューラルネットワークシステムの現在の時間ステップの動作のための出力データとニューラルネットワークシステムの前の時間ステップの動作からのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと；ニューラルネットワークシステムのデコーダ部分を使用して、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて現在の時間ステップの動作のための動き推定データを決定することと；前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと；現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することに少なくとも部分的によって、現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと；ワーピングした再構成されたビデオフレームに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、を行うように構成される。

【0025】

[0025]別の例では、非一時的コンピュータ可読媒体であって、1つまたは複数のプロセッサによって実行されたとき、1つまたは複数のプロセッサに、ニューラルネットワークシステムのデコーダ部分を使用して、ニューラルネットワークシステムの現在の時間ステップの動作のための出力データとニューラルネットワークシステムの前の時間ステップの動作からのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと；ニューラルネットワークシステムのデコーダ部分を使用して、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて現在の時間ステップの動作のための動き推定データを決定することと；前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと；現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームの1つまたは

複数のピクセルを修正することに少なくとも部分的によって、現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと；ワーピングした再構成されたビデオフレームに少なくとも部分的に基づいて現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、を行わせる命令を記憶した非一時的コンピュータ可読媒体が提供される。

【0026】

[0026]別の例では、ビデオデータを処理するための装置が提供される。本装置は、ニューラルネットワークシステムの現在の時間ステップの動作のための出力データとニューラルネットワークシステムの前の時間ステップの動作からのニューラルネットワークシステムのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得するための手段と；出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための動き推定データを決定するための手段と；前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得するための手段と；現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することに少なくとも部分的によって、現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成するための手段と；ワーピングした再構成されたビデオフレームに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成するための手段と、を含む。

【0027】

[0027]いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成された残差データを決定することと；現在の時間ステップのために決定された再構成された残差データにワーピングした再構成されたビデオフレームを加算することに少なくとも部分的によって、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、をさらに備える。

【0028】

[0028]いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、現在の時間ステップの動作中にデコーダ部分によって、現在の時間ステップの動作中にニューラルネットワークシステムのエンコーダ部分によって処理された前の入力ビデオフレームに少なくとも部分的に基づいて、前の時間ステップの動作からの再構成された残差データを決定することと；現在の時間ステップの動作中に、前の時間ステップの動作からの再構成された残差データを使用して前に再構成されたビデオフレームを修正することと；現在の時間ステップの動作のための最終の出力ビデオフレームとして修正された前に再構成されたビデオフレームを出力することと、をさらに備える。

【0029】

[0029]いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成された残差データを決定することと；デコーダ部分のポストワーピングネットワークによって、ワーピングした再構成されたビデオフレームと現在の時間ステップのために決定された再構成された残差データとを処理することと；ポストワーピングネットワークによるワーピングした再構成されたビデオフレームと再構成された残差データとの処理に少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、をさらに備える。

【0030】

[0030]いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、現在の時間ステップの動作のために決定された動き推定データまたは前の時間ステップの動作のために決定された動き推定を使用して前の時間ステップの動作からの再帰状

10

20

30

40

50

態データを修正することに少なくとも部分的によって、現在の時間ステップの動作のためのワーピングした再帰状態データを生成することと；ワーピングした再帰状態データに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、をさらに備える。

【0031】

【0031】いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、デコーダ部分によって、再構成されたマスクを生成することと、再構成されたマスクは、前の時間ステップの動作中に生成された前に再構成されたビデオフレーム中のデータの存在を示す1つまたは複数の値を含む；再構成されたマスクと、現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームと、現在の時間ステップの動作のために決定された再構成された残差データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、再構成されたマスクは、ワーピングした再構成されたビデオフレームと再構成された残差データとの寄与を重み付けする、をさらに備える。

【0032】

【0032】いくつかの態様では、上記で説明された方法、装置、およびコンピュータ可読媒体は、ニューラルネットワークシステムのエンコーダ部分によって、現在の時間ステップの動作のための入力ビデオフレームと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された動き推定データと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からのニューラルネットワークシステムのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと；ニューラルネットワークシステムのエンコーダ部分によって、現在の時間ステップの動作のための出力データを生成することと、出力データは、入力ビデオフレームと、前の時間ステップの動作からの再構成された動き推定データと、前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて生成される、をさらに備える。

【0033】

【0033】いくつかの態様では、装置は、モバイルデバイス、ウェアラブルデバイス、エクステンデッドリアリティデバイス（たとえば、仮想現実（VR）デバイス、拡張現実（AR）デバイス、もしくは複合現実（MR）デバイス）、パーソナルコンピュータ、ラップトップコンピュータ、ビデオサーバ、テレビジョン、または他のデバイスを備える。いくつかの態様では、装置は、モバイルデバイス（たとえば、携帯電話またはいわゆる「スマートフォン」）を備える。いくつかの態様では、装置は、1つまたは複数の画像またはビデオフレームをキャプチャするための少なくとも1つのカメラをさらに含む。たとえば、装置は、1つもしくは複数の画像および／またはビデオフレームを含む1つもしくは複数のビデオをキャプチャするためのカメラ（たとえば、RGBカメラ）または複数のカメラを含むことができる。いくつかの態様では、装置は、1つまたは複数の画像、ビデオ、通知、または他の表示可能なデータを表示するためのディスプレイを含む。いくつかの態様では、装置は、少なくとも1つのデバイスに送信媒体を介して再構成されたビデオフレームを送信するように構成された送信機を含む。いくつかの態様では、プロセッサは、ニューラルプロセッシングユニット（NPU）、中央処理ユニット（CPU）、グラフィックス処理ユニット（GPU）、または他の処理デバイスもしくは構成要素を含む。

【0034】

【0034】本概要は、特許請求される主題の重要な特徴または必須の特徴を識別するように意図されておらず、特許請求される主題の範囲を決定するために別個に使用されるようにも意図されていない。主題は、本特許の明細書全体、いずれかまたはすべての図面、および各請求項の適切な部分を参照することによって理解されたい。

【0035】

【0035】前述の内容は、他の特徴および実施形態とともに、以下の明細書、特許請求の範囲、および添付の図面を参照すれば、より明らかになるであろう。

【0036】

【0036】本出願の例示的な実施形態は、以下の図を参照しながら以下で詳細に説明される。

【図面の簡単な説明】

【0037】

【図1】【0037】システムオンチップ（SOC）の例示的な実装形態を示す図。

【図2A】【0038】全結合ニューラルネットワークの一例を示す図。

【図2B】【0039】局所結合ニューラルネットワークの一例を示す図。

【図2C】【0040】畳み込みニューラルネットワークの一例を示す図。

【図2D】【0041】画像から視覚特徴を認識するように設計されたディープ畳み込みネットワーク（DCN）の詳細な例を示す図。

10

【図3】【0042】深層畳み込みネットワーク（DCN）を示すブロック図。

【図4】【0043】いくつかの例による、再帰ベースのニューラルネットワークを使用してビデオコーディングを実施するように動作可能なデバイスを含むシステムの一部を示す図。

【図5A】【0044】いくつかの例による、ビデオコーディングのために使用されるニューラルネットワークシステムの一部を示す図。

【図5B】【0045】いくつかの例による、ビデオコーディングのために使用される再帰ベースのニューラルネットワークの詳細な例を示す図。

【図6】【0046】いくつかの例による、動き推定および動き補償技法の例を示す図。

【図7】【0047】いくつかの例による、ビデオコーディングのために使用される再帰ベースのニューラルネットワークシステムの別の例を示す図。

20

【図8】【0048】いくつかの例による、ビデオコーディングのために使用される再帰ベースのニューラルネットワークシステムの別の例を示す図。

【図9】【0049】いくつかの例による、ビデオコーディングのために使用される再帰ベースのニューラルネットワークシステムの別の例を示す図。

【図10】【0050】いくつかの例による、単一レベルの再帰およびマルチレベルの再帰構成の一部を示す図。

【図11】【0051】いくつかの例による、ビデオコーディングのために使用される再帰ベースのニューラルネットワークシステムの別の例を示す図。

【図12】【0052】いくつかの例による、ビデオコーディングのために使用される再帰ベースのニューラルネットワークシステムの別の例を示す図。

30

【図13】【0053】いくつかの例による、ビデオコーディングのために使用される再帰ベースのニューラルネットワークシステムの別の例を示す図。

【図14】【0054】いくつかの例による、再帰ベースのニューラルネットワークシステムへの入力としてひずみマップを使用することの一例を示す図。

【図15】【0055】いくつかの例による、再帰ベースのニューラルネットワークシステムにおいて追加のニューラルネットワークレイヤを使用することの一例を示す図。

【図16】【0056】いくつかの例による、他の技法を使用して取得された結果に対する本明細書で説明される技法を使用して取得された結果の一例を示すグラフ。

【図17】【0057】いくつかの例による、他の技法を使用して取得された結果に対する本明細書で説明される技法を使用して取得された結果の一例を示すグラフ。

40

【図18】【0058】いくつかの例による、ビデオデータを処理するためのプロセスの一例を示すフローチャート。

【図19】【0059】いくつかの例による、ビデオデータを処理するためのプロセスの別の例を示すフローチャート。

【図20】【0060】本明細書で説明される様々な技法を実装することができる例示的なコンピューティングデバイスの例示的なコンピューティングデバイスアーキテクチャを示す図。

【発明を実施するための形態】

【0038】

【0061】本開示のいくつかの態様および実施形態が以下で提供される。当業者には明らか

50

なように、これらの態様および実施形態のうちのいくつかは独立して適用される場合があり、それらのうちのいくつかは組み合わせて適用される場合がある。以下の説明では、解説の目的で、本出願の実施形態の完全な理解を与えるために具体的な詳細が述べられる。ただし、様々な実施形態は、これらの具体的な詳細なしに実施され得ることが明らかであろう。図および説明は限定するものではない。

【0039】

[0062]以下の説明は、例示的な実施形態を与えるにすぎず、本開示の範囲、適用性、または構成を限定するものではない。むしろ、例示的な実施形態の次の説明は、例示的な実施形態を実装することを可能にする説明を当業者に提供するであろう。添付の特許請求の範囲に記載されるように、本出願の趣旨および範囲から逸脱することなく、要素の機能および構成において様々な変更が行われ得ることを理解されたい。

【0040】

[0063]デジタルビデオデータは、特に、高品質のビデオデータに対する需要が成長し続けるにつれて大量のデータを含むことができる。たとえば、ビデオデータの消費者は、一般に、高い忠実度、解像度、フレームレートなどをもつますます高い品質のビデオを望む。しかしながら、そのような需要を満たすために必要とされる大量のビデオデータは、通信ネットワークならびにビデオデータを処理し、記憶するデバイスに有意な負担をかける可能性がある。

【0041】

[0064]ビデオデータをコーディングするために、様々な技法が使用され得る。ビデオコーディングは、特定のビデオコーディング規格に従って実施され得る。例示的なビデオコーディング規格は、高効率ビデオコーディング（H E V C）、アドバンスドビデオコーディング（A V C）、ムービングピクチャエキスパートグループ（M P E G）コーディング、および汎用ビデオコーディング（V V C）を含む。ビデオコーディングは、しばしば、ビデオ画像またはシーケンス中に存在する冗長を利用するインター予測またはイントラ予測などの予測方法を使用する。ビデオコーディング技法の共通の目標は、ビデオ品質の劣化を回避するかまたは最小限に抑えながら、より低いビットレートを使用する形式にビデオデータを圧縮することである。ビデオサービスに対する需要が高まり、新しいビデオサービスが利用可能になるにつれて、より良いコーディング効率、パフォーマンス、およびレート制御をもつコーディング技法が必要とされる。

【0042】

[0065]本明細書では、1つまたは複数の機械学習（M L）システムを使用してビデオコーディングを実施するための技法について説明される。M Lは、擬似インテリジェンス（A I）のサブセットである。M Lシステムは、明示的な命令を使用せずにパターンおよび推測に依拠することによって様々なタスクを実施するためにコンピュータシステムが使用することができるアルゴリズムおよび統計モデルを含む。M Lシステムの一例は、擬似ニューロン（たとえば、神経細胞モデル）の相互接続されたグループから構成され得る（擬似ニューラルネットワークとも呼ばれる）ニューラルネットワークである。ニューラルネットワークは、特に、画像解析および／またはコンピュータビジョンアプリケーション、インターネットプロトコル（I P）カメラ、モノのインターネット（I o T）デバイス、自律車両、サービスロボットなどの様々な適用例および／またはデバイスのために使用され得る。

【0043】

[0066]ニューラルネットワーク中の個々のノードは、入力データを取り、データに対して単純な演算を実施することによって、生物学的ニューロンをエミュレートし得る。入力データに対して実施された単純な演算の結果は、他のニューロンに選択的に受け渡される。重み値が、各ベクトルとネットワーク中のノードとに関連付けられ、これらの値は、入力データがどのように出力データに関係するかを制約する。たとえば、各ノードの入力データは、対応する重み値によって乗算され得、この積は合計され得る。積の合計は、随意的バイアスによって調整され得、活性化関数が結果に適用され、ノードの出力信号または

(活性化マップまたは特徴マップと呼ばれることがある)「出力活性化」が生じ得る。重み値は、初期に、ネットワークを通してトレーニングデータの反復フローによって決定され得る(たとえば、重み値は、ネットワークが特定のクラスの典型的な入力データ特性によってそれらのクラスをどのように識別すべきかを学習するトレーニングフェーズ中に確定される)。

【0044】

[0067]特に、リカレントニューラルネットワーク(RNN)、多層パーセプトロン(MLP)ニューラルネットワーク、畳み込みニューラルネットワーク(CNN)など、様々なタイプのニューラルネットワークが存在する。RNNは、レイヤの出力を節約し、この出力を、レイヤの結果を予測する際に助けるために入力にフィードバックするという原理で動作する。MLPニューラルネットワークでは、データは入力レイヤに供給され得、1つまたは複数の隠れレイヤが、データに抽象化レベルを提供する。予測が、次いで、抽象化されたデータに基づいて出力レイヤ上で行われ得る。MLPは、入力にクラスまたはラベルが割り当てられる分類予測問題に特に好適であり得る。畳み込みニューラルネットワーク(CNN)は、フィードフォワード人工ニューラルネットワークのタイプである。畳み込みニューラルネットワークは、各々が受容野(たとえば、入力空間の空間的に局所化された領域)を有し、入力空間を集散的にタイリングする人工ニューロンの集合を含み得る。畳み込みニューラルネットワークは多数の適用例を有する。特に、CNNは、パターン認識および分類の領域内で広く使用されている。

【0045】

[0068](複数の隠れレイヤが存在するとき、ディープニューラルネットワークと呼ばれる)層状ニューラルネットワークアーキテクチャでは、人工ニューロンの第1のレイヤの出力は、人工ニューロンの第2のレイヤへの入力になり、人工ニューロンの第2のレイヤの出力は、人工ニューロンの第3のレイヤの入力になり、以下同様である。畳み込みニューラルネットワークは、特徴の階層を認識するようにトレーニングされ得る。畳み込みニューラルネットワークアーキテクチャにおける計算は、1つまたは複数の計算チェーンにおいて構成され得る処理ノードの集団にわたって分散され得る。これらの多層化アーキテクチャは、一度に1つのレイヤをトレーニングされ得、バックプロパゲーションを使用して微調整され得る。

【0046】

[0069]図1は、本明細書で説明される機能のうちの1つまたは複数を実施するように構成された中央処理ユニット(CPU)102またはマルチコアCPUを含み得るシステムオンチップ(SOC)100の例示的な実装形態を示す。様々な情報の中でも、パラメータまたは変数(たとえば、ニューラル信号およびシナプス荷重)、算出デバイスに関連するシステムパラメータ(たとえば、重みをもつニューラルネットワーク)、遅延、周波数ビン情報、タスク情報が、ニューラル処理ユニット(NPU)108に関連するメモリブロックに記憶され、CPU102に関連するメモリブロックに記憶され、グラフィックス処理ユニット(GPU)104に関連するメモリブロックに記憶され、デジタル信号プロセッサ(DSP)106に関連するメモリブロックに記憶され、メモリブロック118に記憶され得、および／または複数のブロックにわたって分散され得る。CPU102において実行される命令が、CPU102に関連するプログラムメモリからロードされ得るか、またはメモリブロック118からロードされ得る。

【0047】

[0070]SOC100はまた、GPU104、DSP106など、特定の機能に適合された追加の処理ブロックと、第5世代(5G)接続性、第4世代ロングタームエボリューション(4G LTE(登録商標))接続性、Wi-Fi(登録商標)接続性、USB接続性、Bluetooth(登録商標)接続性などを含み得る接続性ブロック110と、たとえば、ジェスチャーを検出および認識し得るマルチメディアプロセッサ112とを含み得る。一実装形態では、NPUは、CPU102、DSP106、および／またはGPU104において実装される。SOC100はまた、センサープロセッサ114、画像信号

プロセッサ（ISP）116、および／または全地球測位システムを含み得るナビゲーションモジュール120を含み得る。

【0048】

【0071】SOC100はARM命令セットに基づき得る。本開示の一態様では、CPU102にロードされる命令は、ルックアップテーブル（LUT）において、入力値とフィルタ重みとの乗算積に対応する記憶された乗算結果を探索するためのコードを備え得る。CPU102にロードされる命令はまた、乗算積のルックアップテーブルヒットが検出されたとき、乗算積の乗算演算中に乗算器を無効にするためのコードを備え得る。加えて、CPU102にロードされる命令は、乗算積のルックアップテーブルミスが検出されたとき、入力値とフィルタ重みとの計算された乗算積を記憶するためのコードを備え得る。

10

【0049】

【0072】SOC100および／またはその構成要素は、本明細書で説明される本開示の態様に従って機械学習技法を使用して（ビデオ符号化および／または復号とも呼ばれ、ビデオコーディングと総称される）ビデオ圧縮および／または復元を実施するように構成され得る。ビデオ圧縮および／または復元を実施するためにディープラーニングアーキテクチャを使用することによって、本開示の態様は、デバイス上でのビデオ圧縮および／または復元の効率を増加させることができる。たとえば、説明されるビデオコーディング技法を使用するデバイスは、機械学習ベースの技法を使用してより効率的にビデオを圧縮ことができ、別のデバイスに圧縮されたビデオを送信することができ、他のデバイスは、本明細書で説明される機械学習ベースの技法を使用してより効率的に圧縮されたビデオを解凍することができる。

20

【0050】

【0073】上記のように、ニューラルネットワークは、機械学習システムの一例であり、入力レイヤと、1つまたは複数の隠れレイヤと、出力レイヤとを含むことができる。データは、入力レイヤの入力ノードから与えられ、処理は、1つまたは複数の隠れレイヤの隠れノードによって実施され、出力は、出力レイヤの出力ノードを通して生成される。ディープラーニングネットワークは、一般に、複数の隠れレイヤを含む。ニューラルネットワークの各レイヤは、人工ニューロン（またはノード）を含むことができる特徴マップまたは活性化マップを含むことができる。特徴マップは、フィルタ、カーネルなどを含むことができる。ノードは、レイヤのうちの1つまたは複数のノードの重要度を示すために使用される1つまたは複数の重みを含むことができる。場合によっては、ディープラーニングネットワークは、一連の多くの隠れレイヤを有することができ、初期レイヤは、入力の単純で低レベルの特性を決定するために使用され、後のレイヤは、より複雑で抽象的な特性の階層を構築する。

30

【0051】

【0074】ディープラーニングアーキテクチャは、特徴の階層を学習し得る。たとえば、視覚データが提示された場合、第1のレイヤは、エッジなど、入力ストリーム中の比較的単純な特徴を認識するように学習し得る。別の例では、聴覚データが提示された場合、第1のレイヤは、特定の周波数におけるスペクトル電力を認識するように学習し得る。第1のレイヤの出力を入力として取る第2のレイヤは、視覚データの場合の単純な形状、または聴覚データの場合の音の組合せなど、特徴の組合せを認識するように学習し得る。たとえば、上位レイヤは、視覚データ中の複雑な形状、または聴覚データ中の単語を表現するように学習し得る。さらに上位のレイヤは、共通の視覚オブジェクトまたは発話フレーズを認識するように学習し得る。

40

【0052】

【0075】ディープラーニングアーキテクチャは、自然階層構造を有する問題に適用されたとき、特にうまく機能し得る。たとえば、原動機付き車両の分類は、ホイール、フロントガラス、および他の特徴を認識するための第1の学習から恩恵を受け得る。これらの特徴は、車、トラック、および飛行機を認識するために、異なる方法で、上位レイヤにおいて組み合わされ得る。

50

【0053】

[0076]ニューラルネットワークは、様々な結合性パターンを用いて設計され得る。フィードフォワードネットワークでは、情報が下位レイヤから上位レイヤに受け渡され、所与のレイヤにおける各ニューロンは、上位レイヤにおけるニューロンに通信する。上記で説明されたように、フィードフォワードネットワークの連続するレイヤにおいて、階層表現が蓄積され得る。ニューラルネットワークはまた、リカレントまたは（トップダウンとも呼ばれる）フィードバック結合を有し得る。リカレント結合では、所与のレイヤにおけるニューロンからの出力は、同じレイヤにおける別のニューロンに通信され得る。リカレントアーキテクチャは、ニューラルネットワークに順次送出される入力データチャンクのうちの2つ以上にわたるパターンを認識するのに役立ち得る。所与のレイヤにおけるニューロンから下位レイヤにおけるニューロンへの結合は、フィードバック（またはトップダウン）結合と呼ばれる。高レベルの概念の認識が、入力の特定の低レベルの特徴を弁別することを助け得るとき、多くのフィードバック結合をもつネットワークが役立ち得る。

10

【0054】

[0077]ニューラルネットワークのレイヤ間の結合は全結合または局所結合であり得る。図2Aは、全結合ニューラルネットワーク202の一例を示す。全結合ニューラルネットワーク202では、第1のレイヤにおけるニューロンは、第2のレイヤにおける各ニューロンが第1のレイヤにおけるあらゆるニューロンから入力を受信するように、その出力を第2のレイヤにおけるあらゆるニューロンに通信し得る。図2Bは、局所結合ニューラルネットワーク204の一例を示す。局所結合ニューラルネットワーク204では、第1のレイヤにおけるニューロンは、第2のレイヤにおける限られた数のニューロンに結合され得る。より一般的には、局所結合ニューラルネットワーク204の局所結合レイヤは、レイヤにおける各ニューロンが同じまたは同様の結合性パターンを有するように構成されるが、異なる値を有し得る結合強度で構成され得る（たとえば、210、212、214、および216）。局所結合の結合性パターンは、所与の領域中の上位レイヤニューロンが、ネットワークへの総入力のうちの制限された部分のプロパティにトレーニングを通して調節された入力を受信し得るので、上位レイヤにおいて空間的に別個の受容野を生じ得る。

20

【0055】

[0078]局所結合ニューラルネットワークの一例は、畳み込みニューラルネットワークである。図2Cは、畳み込みニューラルネットワーク206の一例を示す。畳み込みニューラルネットワーク206は、第2のレイヤにおける各ニューロンのための入力に関連する結合強度が共有されるように構成され得る（たとえば、208）。畳み込みニューラルネットワークは、入力の空間口ケーションが有意味である問題に好適であり得る。畳み込みニューラルネットワーク206は、本開示の態様による、ビデオ圧縮および／または復元の1つまたは複数の態様を実施するために使用され得る。

30

【0056】

[0079]1つのタイプの畳み込みニューラルネットワークは、深層畳み込みネットワーク(DCN)である。図2Dは、車載カメラなどの画像キャプチャデバイス230から入力された画像226から視覚特徴を認識するように設計されたDCN200の詳細な例を示す。本例のDCN200は、交通標識と、交通標識上で与えられた数とを識別するようにトレーニングされ得る。もちろん、DCN200は、車線マーキングを識別すること、または交通信号を識別することなど、他のタスクのためにトレーニングされ得る。

40

【0057】

[0080]DCN200は、教師あり学習を用いてトレーニングされ得る。トレーニング中に、DCN200は、速度制限標識の画像226など、画像を提示され得、次いで、出力222を生成するために、フォワードパスが算出され得る。DCN200は、特徴抽出セクションと分類セクションとを含み得る。画像226を受信すると、畳み込みレイヤ232が、特徴マップ218の第1のセットを生成するために、畳み込みカーネル（図示せず）を画像226に適用し得る。一例として、畳み込みレイヤ232のための畳み込みカー

50

ネルは、 28×28 特徴マップを生成する 5×5 カーネルであり得る。本例では、4つの異なる特徴マップが、特徴マップ 218 の第1のセットにおいて生成されるので、4つの異なる畳み込みカーネルが、畳み込みレイヤ 232 において画像 226 に適用された。畳み込みカーネルは、フィルタまたは畳み込みフィルタと呼ばれることもある。

【0058】

[0081] 特徴マップ 218 の第1のセットは、特徴マップ 220 の第2のセットを生成するために、最大プーリングレイヤ（図示せず）によってサブサンプリングされ得る。最大プーリングレイヤは、特徴マップ 218 の第1のセットのサイズを低減する。すなわち、 14×14 などの特徴マップ 220 の第2のセットのサイズが、 28×28 などの特徴マップ 218 の第1のセットのサイズよりも小さい。低減されたサイズは、メモリ消費を低減しながら、後続のレイヤに同様の情報を提供する。特徴マップ 220 の第2のセットは、特徴マップの1つまたは複数の後続のセット（図示せず）を生成するために、1つまたは複数の後続の畳み込みレイヤ（図示せず）を介して、さらに畳み込まれ得る。

【0059】

[0082] 図 2D の例では、特徴マップ 220 の第2のセットは、第1の特徴ベクトル 224 を生成するために畳み込まれる。さらに、第1の特徴ベクトル 224 は、第2の特徴ベクトル 228 を生成するために、さらに畳み込まれる。第2の特徴ベクトル 228 の各特徴は、「標識」、「60」、および「100」など、画像 226 の可能な特徴に対応する数を含み得る。ソフトマックス関数（図示せず）が、第2の特徴ベクトル 228 中の数を確率にコンバートし得る。したがって、DCN 200 の出力 222 は、画像 226 が1つまたは複数の特徴を含む確率である。

【0060】

[0083] 本例では、「標識」および「60」についての出力 222 における確率は、「30」、「40」、「50」、「70」、「80」、「90」、および「100」など、出力 222 の他のものの確率よりも高い。トレーニングの前に、DCN 200 によって生成される出力 222 は、不正確である可能性がある。したがって、誤差が、出力 222 とターゲット出力との間で計算され得る。ターゲット出力は、画像 226（たとえば、「標識」および「60」）のグラントゥールス（ground truth）である。次いで、DCN 200 の重みは、DCN 200 の出力 222 がターゲット出力とより密接に整合されるように調整され得る。

【0061】

[0084] 重みを調整するために、学習アルゴリズムは、重みのための勾配ベクトルを算出し得る。勾配（gradient）は、重みが調整された場合に、誤差が増加または減少する量を示し得る。最上レイヤにおいて、勾配は、最後から2番目のレイヤにおける活性化されたニューロンと出力レイヤにおけるニューロンとを結合する重みの値に直接対応し得る。下位レイヤでは、勾配は、重みの値と、上位レイヤの算出された誤差勾配とに依存し得る。次いで、重みは、誤差を低減するために調整され得る。重みを調整するこの様式は、それがニューラルネットワークを通して「バックワードパス（backward pass）」を伴うので、「バックプロパゲーション（back propagation）」と呼ばれることがある。

【0062】

[0085] 実際には、重みの誤差勾配は、計算された勾配が真の誤差勾配を近似するように、少数の例にわたって計算され得る。この近似方法は、確率的勾配降下（stochastic gradient descent）と呼ばれることがある。システム全体の達成可能な誤差レートが減少しなくなるまで、または誤差レートがターゲットレベルに達するまで、確率的勾配降下が繰り返され得る。学習の後に、DCN は新しい画像を提示され得、ネットワークを通したフォワードパスは、DCN の推論または予測と見なされ得る出力 222 をもたらし得る。

【0063】

[0086] 深層信念ネットワーク（DBN）は、隠れノードの複数のレイヤを備える確率モデルである。DBN は、トレーニングデータセットの階層表現を抽出するために使用され得る。DBN は、制限ボルツマンマシン（RBM）のレイヤを積層することによって取得

10

20

30

40

50

され得る。RBMは、入力のセットにわたる確率分布を学習することができる人工ニューラルネットワークのタイプである。RBMは、各入力にそれぞれにカテゴリ分類されるべきクラスに関する情報の不在下で確率分布を学習することができるので、RBMは、教師なし学習においてしばしば使用される。ハイブリッド教師なしおよび教師ありパラダイムを使用して、DBNの下部RBMは、教師なし様式でトレーニングされ得、特徴抽出器として働き得、上部RBMは、（前のレイヤからの入力とターゲットクラスとの同時分布上で）教師あり様式でトレーニングされ得、分類器として働き得る。

【0064】

【0087】深層畳み込みネットワーク（DCN）は、追加のプーリング（pooling）レイヤおよび正規化レイヤで構成された、畳み込みネットワークのネットワークである。DCNは、多くのタスクに関して最先端の性能を達成している。DCNは、入力と出力ターゲットの両方が、多くの標本について知られており、勾配降下方法の使用によってネットワークの重みを変更するために使用される、教師あり学習を使用してトレーニングされ得る。

【0065】

【0088】DCNは、フィードフォワードネットワークであり得る。加えて、上記で説明されたように、DCNの第1のレイヤにおけるニューロンから次の上位レイヤにおけるニューロンのグループへの結合は、第1のレイヤにおけるニューロンにわたって共有される。DCNのフィードフォワードおよび共有結合は、高速処理のために活用され得る。DCNの算出負担は、たとえば、リカレントまたはフィードバック結合を備える同様のサイズのニューラルネットワークのそれよりもはるかに少ないことがある。

【0066】

【0089】畳み込みネットワークの各レイヤの処理は、空間的に不変のテンプレートまたは基底投射と見なされ得る。入力が、カラー画像の赤色、緑色、および青色チャネルなど、複数のチャネルに最初に分解された場合、その入力に関してトレーニングされた畳み込みネットワークは、画像の軸に沿った2つの空間次元と、色情報をキャプチャする第3の次元とをもつ、3次元であると考えられ、畳み込み結合の出力は、後続のレイヤにおいて特徴マップを形成すると考えられ、特徴マップ（たとえば、220）の各要素が、前のレイヤ（たとえば、特徴マップ218）における様々なニューロンから、および複数のチャネルの各々から入力を受信し得る。特徴マップにおける値は、整流（rectification）、 $\max(0, x)$ など、非線形性を用いてさらに処理され得る。隣接するニューロンからの値は、さらにプールされ得、これは、ダウンサンプリングに対応し、さらなる局所不変性と次元削減とを与え得る。

【0067】

【0090】図3は、深層畳み込みネットワーク350の一例を示すブロック図である。深層畳み込みネットワーク350は、結合性および重み共有に基づく、複数の異なるタイプのレイヤを含み得る。図3に示されているように、深層畳み込みネットワーク350は、畳み込みブロック354A、354Bを含む。畳み込みブロック354A、354Bの各々は、畳み込み層（CONV）356と、正規化層（LNorm）358と、最大プーリング層（MAX POOL）360とで構成され得る。

【0068】

【0091】畳み込み層356は、1つまたは複数の畳み込みフィルタを含み得、これは、特徴マップを生成するために入力データ352に適用され得る。2つの畳み込みブロック354A、354Bのみが示されているが、本開示はそのように限定しておらず、代わりに、設計選好に従って、任意の数の畳み込みブロック（たとえば、ブロック354A、354B）が深層畳み込みネットワーク350中に含まれ得る。正規化レイヤ358は、畳み込みフィルタの出力を正規化し得る。たとえば、正規化レイヤ358は、白色化またはラテラル抑制を行い得る。最大プーリングレイヤ360は、局所不変性および次元削減のために、空間にわたってダウンサンプリングアグリゲーションを行い得る。

【0069】

【0092】たとえば、深層畳み込みネットワークの並列フィルタバンクは、高性能および低

10

20

30

40

50

電力消費を達成するために、SOC 100のCPU 102またはGPU 104にロードされ得る。代替実施形態では、並列フィルタバンクは、SOC 100のDSP 106またはISP 116にロードされ得る。さらに、深層畳み込みネットワーク 350は、それぞれ、センサーおよびナビゲーションに専用の、センサープロセッサ 114およびナビゲーションモジュール 120など、SOC 100上に存在し得る他の処理ブロックにアクセスし得る。

【0070】

[0093]深層畳み込みネットワーク 350はまた、(「FC1」と標示された)レイヤ 362Aおよび(「FC2」と標示された)レイヤ 362Bなどの1つまたは複数の全結合(full connected)レイヤを含み得る。深層畳み込みネットワーク 350は、ロジスティック回帰(LR: logistic regression)レイヤ 364をさらに含み得る。深層畳み込みネットワーク 350の各レイヤ 356、358、360、362、364の間には、更新されるべき重み(図示せず)がある。レイヤ(たとえば、356、358、360、362、364)の各々の出力は、畳み込みブロック 354Aのうちの第1のものにおいて供給された入力データ 352(たとえば、画像、オーディオ、ビデオ、センサーデータおよび/または他の入力データ)から階層特徴表現を学習するために、深層畳み込みネットワーク 350中のレイヤ(たとえば、356、358、360、362、364)のうちの後続のレイヤの入力として働き得る。深層畳み込みネットワーク 350の出力は、入力データ 352についての分類スコア 366である。分類スコア 366は、確率のセットであり得、ここで、各確率は、入力データが特徴のセットからの特徴を含む確率である。

【0071】

[0094]上記のように、デジタルビデオデータは、大量のデータを含むことができ、これは、通信ネットワークならびにビデオデータを処理し、記憶するデバイスに有意な負担をかける可能性がある。たとえば、圧縮されていないビデオコンテンツを記録することは、概して、記録されるビデオコンテンツの解像度が増加するにつれて大幅に増加する大きいファイルサイズを生じる。1つの例示的な例では、1080p/24(たとえば、24フレーム毎秒でキャプチャされる幅が1920ピクセルで高さが1080ピクセルの解像度)で記録される圧縮されていない16ビットパーチャネルビデオは、フレームごとに12.4メガバイトまたは毎秒297.6メガバイトを占有し得る。24フレーム毎秒で4K解像度で記録される圧縮されていない16ビットパーチャネルビデオは、フレームごとに49.8メガバイトまたは毎秒1195.2メガバイトを占有し得る。

【0072】

[0095]ネットワーク帯域幅は、大きいビデオファイルが問題になる可能性がある別の制約である。たとえば、ビデオコンテンツは、しばしば、ワイヤレスネットワークを介して(たとえば、LTE、LTEアドバンスド、新無線(NR)、Wi-Fi(登録商標)、Bluetooth、または他のワイヤレスネットワークを介して)送出され、消費者のインターネットトラフィックの大部分を構成することができる。ワイヤレスネットワークにおける利用可能な帯域幅の量の進歩にもかかわらず、これらのネットワークにおいてビデオコンテンツを送出するために使用される帯域幅の量を低減することが依然として望ましいことがある。

【0073】

[0096]圧縮されていないビデオコンテンツが、物理的なストレージのために大きいメモリを伴い、送信のためにかなりの帯域幅を伴い得る大きいファイルを生じる可能性があるため、そのようなビデオコンテンツを圧縮し、次いで、解凍するためにビデオコーディング技法が利用され得る。

【0074】

[0097]ビデオコンテンツのサイズ、したがって、ビデオコンテンツを記憶することに関与するストレージの量およびビデオコンテンツを送出することに関与する帯域幅の量を低減するために、様々なビデオコーディング技法が、特に、HEVC、AVC、MPEG、VVCなどの特定のビデオコーディング規格に従って実施され得る。ビデオコーディング

は、しばしば、ビデオ画像またはシーケンス中に存在する冗長を利用するインター予測またはイントラ予測などの予測方法を使用する。ビデオコーディング技法の共通の目標は、ビデオ品質の劣化を回避するかまたは最小限に抑えながら、より低いビットレートを使用する形式にビデオデータを圧縮することである。ビデオサービスに対する需要が高まり、新しいビデオサービスが利用可能になるにつれて、より良いコーディング効率、パフォーマンス、およびレート制御をもつコーディング技法が必要とされる。

【0075】

[0098]概して、符号化デバイスは、符号化されたビデオビットストリームを生成するためにビデオコーディング規格に従ってビデオデータを符号化する。いくつかの例では、符号化ビデオビットストリーム（または「ビデオビットストリーム」または「ビットストリーム」）は、一連の1つまたは複数のコード化ビデオシーケンスである。符号化デバイスは、各ピクチャを複数のスライスに区分することによってピクチャのコード化表現を生成することができる。スライスは、スライス中の情報が、同じピクチャ内の他のスライスからのデータへの依存性なしにコーディングされるように、他のスライスから独立している。スライスは、独立しているスライスセグメントを含む1つまたは複数のスライスセグメントと、存在する場合、前のスライスセグメントに依存する1つまたは複数の依存するスライスセグメントとを含む。HEVCでは、スライスは、ルーマサンプルおよびクロマサンプルのコーディングツリーブロック（CTB）に区分される。ルーマサンプルのCTBおよびクロマサンプルの1つまたは複数のCTBは、サンプルのためのシンタックスとともに、コーディングツリーユニット（CTU）と呼ばれる。CTUは「ツリーブロック」または「最大コーディングユニット」（LCU）と呼ばれることもある。CTUは、HEVC符号化のための基本処理ユニットである。CTUは、様々なサイズの複数のコーディングユニット（CU）に分割され得る。CUは、コーディングブロック（CB）と呼ばれるルーマおよびクロマサンプルアレイを含んでいる。

【0076】

[0099]ルーマおよびクロマCBは、予測ブロック（PB）にさらに分割され得る。PBは、（利用可能なとき、または使用のために有効にされたとき）インター予測またはイントラブロックコピー（IBC）予測のために同じ動きパラメータを使用するルーマ成分またはクロマ成分のサンプルのブロックである。ルーマPBおよび1つまたは複数のクロマPBは、関連するシンタックスとともに、予測ユニット（PU）を形成する。インター予測の場合、動きパラメータのセット（たとえば、1つまたは複数の動きベクトル、参照インデックスなど）は、PUごとにビットストリーム中でシグナリングされ、ルーマPBおよび1つまたは複数のクロマPBのインター予測のために使用される。動きパラメータは動き情報と呼ばれることもある。CBはまた、1つまたは複数の変換ブロック（TB）に区分され得る。TBは、予測残差信号をコーディングするために残差変換（たとえば、場合によっては、同じ2次元変換）が適用される色成分のサンプルの正方形ブロックを表現する。変換ユニット（TU）は、ルーマおよびクロマサンプルのTBと、対応するシンタックス要素とを表現する。変換コーディングについては、以下でより詳細に説明する。

【0077】

[0100]HEVC規格によれば、変換は、TUを使用して実施され得る。TUは、所与のCU内のPUのサイズに基づいてサイズ決定され得る。TUは、PUと同じサイズであるか、またはPUよりも小さいことがある。いくつかの例では、CUに対応する残差サンプルは、残差4分木（RQT）として知られる4分木構造を使用して、より小さいユニットに再分割され得る。RQTのリーフノードはTUに対応し得る。TUに関連するピクセル差分値は、変換係数を生成するために変換され得る。変換係数は、次いで、符号化デバイスによって量子化され得る。

【0078】

[0101]ビデオデータのピクチャがCUに区分されると、符号化デバイスは、予測モードを使用して各PUを予測する。予測ユニットまたは予測ブロックは、次いで、（以下で説明される）残差を得るために元のビデオデータから減算される。CUごとに、シンタック

ステータを使用してビットストリーム内で予測モードがシグナリングされ得る。予測モードは、イントラ予測（もしくはピクチャ内予測）またはインター予測（もしくはピクチャ間予測）を含み得る。イントラ予測は、ピクチャ内の空間的に隣接するサンプル間の相関を利用する。たとえば、イントラ予測を使用して、各PUは、たとえば、PUに関する平均値を見つけるためのDC予測、平坦面をPUに適合させるための平面予測、隣接データから外挿するための方向予測、または任意の他の好適なタイプの予測を使用して、同じピクチャ中の隣接画像データから予測される。インター予測は、画像サンプルのブロックについての動き補償予測を導出するためにピクチャ間の時間相関を使用する。たとえば、インター予測を使用して、各PUは、（出力順序において現在ピクチャの前または後の）1つまたは複数の参照ピクチャ中の画像データからの動き補償予測を使用して予測される。ピクチャ間予測を使用してピクチャエリアをコーディングすべきなのか、ピクチャ内予測を使用してピクチャエリアをコーディングすべきなのかの決定は、たとえば、CUレベルにおいて行われ得る。

【0079】

[0102]いくつかの例では、ピクチャの1つまたは複数のスライスは、スライスタイプを割り当てられる。スライスタイプは、Iスライスと、Pスライスと、Bスライスとを含む。Iスライス（フレーム内、独立して復号可能）は、イントラ予測によってのみコーディングされるピクチャのスライスであり、したがって、Iスライスは、スライスの任意の予測ユニットまたは予測ブロックを予測するためにフレーム内のデータのみを必要とするので、独立して復号可能である。Pスライス（単方向予測されるフレーム）は、イントラ予測を用いて、および単方向インター予測を用いてコーディングされ得るピクチャのスライスである。Pスライス内の各予測ユニットまたは予測ブロックは、イントラ予測またはインター予測のいずれかを用いてコーディングされる。インター予測が適用されるとき、予測ユニットまたは予測ブロックは、1つの参照ピクチャによってのみ予測され、したがって、参照サンプルは、1つのフレームの1つの参照領域からのみのものである。Bスライス（双方向予測フレーム）は、イントラ予測を用いて、およびインター予測（たとえば、双予測または単予測のいずれか）を用いてコーディングされ得るピクチャのスライスである。Bスライスの予測ユニットまたは予測ブロックは、2つの参照ピクチャから双方向予測され得、ここで、各ピクチャが1つの参照領域を導き、双方向予測されたブロックの予測信号を生成するために、2つの参照領域のサンプルセットが（たとえば、等しい重みを用いて、または異なる重みを用いて）重み付けされる。上記で説明されたように、1つのピクチャのスライスが独立してコーディングされる。いくつかの場合には、ピクチャは、ただ1つのスライスとしてコーディングされ得る。

【0080】

[0103]イントラ予測および／またはインター予測を使用して予測を実行した後に、符号化デバイスは、変換および量子化を実施することができる。たとえば、予測の後に、符号化デバイスは、PUに対応する残差値を計算し得る。残差値は、コーディングされているピクセルの現在ブロック（PU）と、現在ブロックを予測するために使用される予測ブロック（たとえば、現在ブロックの予測されたバージョン）との間のピクセル差分値を備え得る。たとえば、予測ブロックを生成した（たとえば、インター予測またはイントラ予測を出した）後に、符号化デバイスは、現在ブロックから、予測ユニットによって生成された予測ブロックを減算することによって、残差ブロックを生成することができる。残差ブロックは、現在ブロックのピクセル値と予測ブロックのピクセル値との間の差分を定量化するピクセル差分値のセットを含む。いくつかの例では、残差ブロックは、2次元ブロックフォーマット（たとえば、ピクセル値の2次元行列またはアレイ）で表され得る。そのような例では、残差ブロックはピクセル値の2次元表現である。

【0081】

[0104]予測が実施された後に残存し得るいずれの残差データも、離散コサイン変換、離散サイン変換、整数変換、ウェーブレット変換、他の好適な変換関数、またはそれらの任意の組合せに基づき得るブロック変換を使用して変換される。いくつかの場合には、1つ

10

20

30

40

50

または複数のブロック変換（たとえば、サイズ 32×32 、 16×16 、 8×8 、 4×4 、または他の好適なサイズ）が各CUにおける残差データに適用され得る。いくつかの実施形態では、符号化デバイスによって実装される変換および量子化プロセスのためにTUが使用され得る。1つまたは複数のPUを有する所与のCUは、1つまたは複数のTUをも含み得る。以下でさらに詳細に説明されるように、残差値は、ブロック変換を使用して変換係数に変換され得、次いで、エントロピーコーディングのためのシリアル化変換係数を生成するために、TUを使用して量子化および走査され得る。

【0082】

[0105]符号化デバイスは、変換係数の量子化を実施し得る。量子化は、係数を表すために使用されるデータの量を低減するために変換係数を量子化することによって、さらなる圧縮を与える。たとえば、量子化は、係数の一部または全部に関連するビット深度を低減し得る。一例では、 n ビット値をもつ係数は、量子化中に m ビット値に切り捨てられ得、 n は m よりも大きい。

【0083】

[0106]量子化が実施されると、コード化ビデオビットストリームは、量子化変換係数と、予測情報（たとえば、予測モード、動きベクトル、ブロックベクトルなど）と、区分情報と、他のシンタックスデータなどの任意の他の好適なデータとを含む。コード化ビデオビットストリームの異なる要素は、次いで、符号化デバイスによってエントロピー符号化され得る。いくつかの例では、符号化デバイスは、エントロピー符号化され得るシリアル化ベクトルを作り出すために、量子化変換係数を走査するためにあらかじめ定義された走査順序を利用し得る。いくつかの例では、符号化デバイスは適応型走査を実施し得る。ベクトル（たとえば、1次元ベクトル）を形成するために量子化変換係数を走査した後に、符号化デバイスは、ベクトルをエントロピー符号化し得る。たとえば、符号化デバイスは、コンテキスト適応型可変長コーディング、コンテキスト適応型バイナリ算術コーディング、シンタックスベースコンテキスト適応型バイナリ算術コーディング、確率間隔区分エントロピーコーディング、または別の好適なエントロピー符号化技法を使用し得る。

【0084】

[0107]符号化デバイスは、符号化されたビデオビットストリームを記憶することができ、および／または復号デバイスを含むことができる受信デバイスに通信リンクを介して符号化されたビデオビットストリームデータを送ることができる。符号化デバイスは、（たとえば、エントロピーデコーダを使用して）エントロピー復号し、符号化ビデオデータを構成する1つまたは複数のコード化ビデオシーケンスの要素を抽出することによって、符号化ビデオビットストリームデータを復号し得る。符号化デバイスは、次いで、符号化ビデオビットストリームデータを再スケーリングし、符号化ビデオビットストリームデータに対して逆変換を実施し得る。残差データが、次いで、符号化デバイスの予測段階にパスされる。復号デバイスは、次いで、イントラ予測、インター予測、IBC、および／または他のタイプの予測を使用してピクセルのブロック（たとえば、PU）を予測する。いくつかの例では、逆変換の出力（残差データ）に予測が加算される。復号デバイスは、復号ビデオをビデオ宛先デバイスに出力し得、ビデオ宛先デバイスは、復号ビデオデータをコンテンツの消費者に表示するためのディスプレイまたは他の出力デバイスを含み得る。

【0085】

[0108]様々なビデオコーディング規格（たとえば、上記で説明されたHEVCビデオコーディング技法）によって定義されたビデオコーディングシステムおよび技法は、未加工のビデオコンテンツ中の情報の大部分を保持することが可能であり得、信号処理および情報理論概念に基づいてアプライオリに定義され得る。しかしながら、あらかじめ定義された圧縮アルゴリズムが、概して（たとえば、任意のタイプのビデオコンテンツに）適用可能であり得るが、これらのあらかじめ定義されたビデオコーディング技法は、ネットワークベースの環境または他のタイプの環境において有益であるいくつかの特性が欠如していることがある。たとえば、ビデオコーディングシステムは、低レイテンシの動作を実装することが可能でなければならない。しかしながら、既存のビデオコーディング規格に従って

10

20

30

40

50

動作するビデオコーディングシステムおよび関連する技法は、ビデオを符号化し、それを（ピクチャグループまたはG o Pと呼ばれることがある）いくつかのフレームのバッチで送信し、したがって、送信されるバッチ（たとえば、G o P）のフレームの累算を待つ必要がある。そのようなバッチベースのビデオコーディングは、フレームごとのコーディングおよび送信を使用して達成されるであろうよりも高いレイテンシを生じる。さらに、そのようなビデオコーディングシステムおよび技法は、ビデオの特性に応じてランタイムにG o Pサイズ（すなわち、残差情報だけではなくフルフレームを送信する頻度）を調整する能力が欠如しており、レートひずみ空間においてあまり最適ではない動作点につながる。

【0086】

【0109】本明細書では、エンコーダ段階および／またはデコーダ段階にフィードバックを与える再帰型ニューラルネットワークアーキテクチャを使用して、ビデオコーディング（ビデオ圧縮／符号化および／またはビデオ復元／復号）を実施する（プロセスとも呼ばれる）方法、システム、装置、およびコンピュータ可読媒体（「システムおよび技法」と総称される）について説明される。いくつかの例では、システムおよび技法は、ビデオフレームまたは画像を圧縮し、解凍するために複数のニューラルネットワークシステムを利用することができる。たとえば、第1のニューラルネットワークシステムは、（Iフレームとも呼ばれる）イントラ予測フレームと同様に1つまたは複数の前のフレームに基づいて動き補償を使用せずにフルフレームを圧縮する画像エンコーダ（たとえば、画像オートエンコーダ）として働くことができる。第2のニューラルネットワークシステム（たとえば、再帰オートエンコーダ）は、1つまたは複数の前のフレームに基づいて動き補償を使用して、および（Pフレームとも呼ばれる）単一の参照フレームを使用する単予測フレームまたは（Bフレームと呼ばれる）2つの参照フレームを使用する双予測フレームなどのインター予測されたフレームと同様の）残差情報を使用してフレームを符号化および／または復号するために使用され得る。

【0087】

【0110】第2のニューラルネットワークシステムによって実施される動き補償は、残差情報とともに前に再構成されたフレームからのピクセルが現在のフレーム中の新しい位置にどのように移動したのかについて説明することによって現在のフレームのデータを決定するために使用され得る。第2のニューラルネットワークシステムは、将来の時間ステップにおいてニューラルネットワークのいくつかのレイヤにフィードバックされ得る状態情報を記憶する（たとえば、第2のニューラルネットワークのデコーダ部分中に）1つまたは複数の再帰レイヤを含むことができる。たとえば、第2のニューラルネットワークシステムは、現在の時間ステップ t においてフレーム x_t を処理することができ、第2のニューラルネットワークシステムが次の時間ステップ $t+1$ において後続のフレーム x_{t+1} を処理するために1つまたは複数の再帰レイヤからフレーム x_t を処理することに関連する状態情報をフィードバックすることができる。1つまたは複数の再帰レイヤは、1つもしくは複数の再帰ニューラルネットワーク（RNN）、1つもしくは複数のゲート付き再帰型ユニット（GRU）、1つもしくは複数の長短期記憶（LSTM）ネットワーク、1つもしくは複数のConvRNN、1つもしくは複数のConvGRU、1つもしくは複数のConvLSTM、それらの任意の組合せ、および／または他のタイプのニューラルネットワークアーキテクチャの部分であり得る。

【0088】

【0111】本明細書では、そのようなニューラルネットワークシステムに関係する様々な態様について説明される。たとえば、いくつかの実装形態では、第2のニューラルネットワークは、（たとえば、従来の規格ベースのビデオコーデックにおいて使用されるG o P構造などの）逆時間方向フレームへの依存なしにフレームごとに動作することができる（前のフレームのみを参照する）完全に自己回帰的な条件構造を有することができる。場合によっては、第2のニューラルネットワークの出力は、動き補償、残差、再構成されたフレーム、動き推定（たとえば、オプティカルフロー、ブロック動きベクトル、または適応型畳み込みカーネル）、それらの任意の組合せ、および／または他の出力によってワーピン

10

20

30

40

50

グされた前の再構成されたフレームを含むことができる。第1の時間ステップ（たとえば、時間ステップ t ）からのフィードバック情報は、デコーダからエンコーダにおよび／または後の時間ステップにおいて（たとえば、時間ステップ $t+1$ において）デコーダに与えられ得る。フィードバックは、復号段階において導入されたエラーを表すために使用され得、これは、それらのエラーを補償するためにエンコーダによって使用され得る。前の時間ステップのデコーダから次の時間ステップにおいてエンコーダに与えられる情報は、1つもしくは複数の前に再構成されたフレーム、1つもしくは複数の前に再構成された動き推定、1つもしくは複数の前に再構成された残差、（たとえば、デコーダの1つまたは複数の再帰レイヤからの）デコーダの1つまたは複数の前の再帰状態、それらの任意の組合せ、および／または他の出力を含むことができる。再帰レイヤの再帰状態を、本明細書では隠れ状態と呼ぶこともある。

10

【0089】

[0112]本明細書で説明される1つまたは複数の再帰ベースのニューラルネットワークシステムおよび技法を使用してビデオコンテンツをコーディングすること（たとえば、コーディング／圧縮すること、および復号／解凍すること）によって、本開示の態様は、従来の、アプライオリに定義されたビデオ圧縮アルゴリズム（たとえば、特に、H E V C、V V C、M P E G - 4）のものに迫るかまたはそれを越えたビデオビットレートおよび／または品質を有する圧縮ビデオコンテンツを生成するビデオ符号化および復号機構を提供することができる。場合によっては、本明細書で説明される再帰ベースのニューラルネットワークシステムおよび技法を使用するビデオ符号化および復号機構は、ニューラルネットワークを再トレーニングすることによって様々な使用事例に適応可能であり得る。さらに、本開示の態様は、大量のシンタックスをビットストリームとともに送信することを必要とすることなしにビデオエンコーダおよびデコーダの自律生成を提供し得る。

20

【0090】

[0113]図4は、ビデオの符号化および復号を実施するように構成されたデバイス402を含むシステム400を示す。デバイス402は、カメラ407と記憶媒体414（たとえば、データストレージデバイス）とに結合される。いくつかの実装形態では、カメラ407は、第1のニューラルネットワークシステム410および／または第2のニューラルネットワークシステム411によって符号化するためにプロセッサ404に画像データ408（たとえば、ビデオデータストリーム）を与えるように構成される。いくつかの実装形態では、デバイス402は、複数のカメラ（たとえば、デュアルカメラシステム、3つのカメラ、または他の数のカメラ）に結合され得、および／またはそれらを含むことができる。場合によっては、デバイス402は、マイクロフォンおよび／または他の入力デバイス（たとえば、キーボード、マウス、タッチスクリーンおよび／もしくはタッチパッドなどのタッチ入力デバイス、ならびに／または他の入力デバイス）に結合され得る。いくつかの例では、カメラ407、記憶媒体414、マイクロフォン、および／または他の入力デバイスは、デバイス402の部分であり得る。

30

【0091】

[0114]デバイス402はまた、1つまたは複数のワイヤレスネットワーク、1つまたは複数のワイヤードネットワーク、またはそれらの組合せなどの送信媒体418を介して第2のデバイス490に結合される。たとえば、送信媒体418は、ワイヤレスネットワーク、ワイヤードネットワーク、またはワイヤードネットワークとワイヤレスネットワークとの組合せによって与えられるチャネルを含むことができる。送信媒体418は、ローカルエリアネットワーク、ワイドエリアネットワーク、またはインターネットなどのグローバルネットワークなど、パケットベースネットワークの一部を形成し得る。送信媒体418は、ルータ、スイッチ、基地局、またはソースデバイスから受信デバイスへの通信を可能にするために有用であり得る任意の他の機器を含み得る。ワイヤレスネットワークは、任意のワイヤレスインターフェースまたはワイヤレスインターフェースの組合せを含み得、任意の好適なワイヤレスネットワーク（たとえば、インターネットまたは他のワイドエリアネットワーク、パケットベースネットワーク、W i F i、無線周波数（R F）、U W

40

50

B、Wi-Fi-Direct、セルラー、ロングタームエボリューション（LTE）、WiMax（登録商標）など）を含み得る。ワイヤードネットワークは、任意のワイヤードインターフェース（たとえば、ファイバー、イーサネット（登録商標）、電力線イーサネット、同軸ケーブルを介したイーサネット、デジタル信号線（DSL）など）を含み得る。ワイヤードおよび／またはワイヤレスネットワークは、基地局、ルータ、アクセスポイント、ブリッジ、ゲートウェイ、スイッチなど、様々な機器を使用して実装され得る。符号化ビデオビットストリームデータは、ワイヤレス通信プロトコルなどの通信規格に従って変調され、受信デバイスに送信され得る。

【0092】

[0115]デバイス402は、メモリ406と、第1のインターフェース（「I/F 1」）412と、第2のインターフェース（「I/F 2」）416とに結合された（本明細書では「プロセッサ」と呼ぶ）1つまたは複数のプロセッサ404を含む。プロセッサ404は、カメラ407から、メモリ406から、および／または記憶媒体414から画像データ408を受信するように構成される。プロセッサ404は、第1のインターフェース412を介して（たとえば、メモリバスを介して）記憶媒体414に結合され、第2のインターフェース416（たとえば、ネットワークインターフェースデバイス、ワイヤレストランシーバおよびアンテナ、1つもしくは複数の他のネットワークインターフェースデバイス、またはそれらの組合せ）を介して送信媒体418に結合される。

【0093】

[0116]プロセッサ404は、エンコーダ部分462とデコーダ部分466とを含む第1のニューラルネットワークシステム410と、エンコーダ部分432とデコーダ部分436とを含む第2のニューラルネットワークシステム411とを含む。いくつかの実装形態では、ニューラルネットワークシステム410および／またはニューラルネットワークシステム411は、1つまたは複数のオートエンコーダを含むことができる。エンコーダ部分462は、入力データ470を受信し、入力データ470に少なくとも部分的に基づいて出力データ474を生成するために入力データ470を処理するように構成される。同様に、エンコーダ部分432は、入力データ440を受信し、入力データ440に少なくとも部分的に基づいて出力データ444を生成するために入力データ440を処理するように構成される。場合によっては、ニューラルネットワークシステム411のエンコーダ部分432は、以下でさらに説明されるように、入力データ440に基づき、さらに、第1の状態データ450に基づいて出力データ444を生成するように構成される。場合によっては、エンコーダ部分432は、以下でさらに説明されるように、入力データ440と第1の状態データ450とに基づき、さらに、前の時間ステップ $t-1$ において（たとえば、前の時間ステップ $t-1$ において入力データ470を処理するとき）ニューラルネットワークシステム410のデコーダ部分466によって生成された表現476に基づいて、現在の時間ステップ t において出力データ444を生成するように構成される。場合によっては、エンコーダ部分432は、以下でさらに説明するように、入力データ440と第1の状態データ450とに基づき、さらに、前の時間ステップ $t-1$ において（たとえば、前の時間ステップ $t-1$ において入力データ440を処理するとき）ニューラルネットワークシステム411のデコーダ部分436によって生成された表現446に基づいて現在の時間ステップ t において出力データ444を生成するように構成される。

【0094】

[0117]いくつかの実装形態では、ニューラルネットワークシステム410のエンコーダ部分462は、出力データ474を生成するために入力データ470の不可逆的圧縮を実施するように構成され、したがって、出力データ474は、入力データ470よりも少数のビットを有する。エンコーダ部分462は、任意の前の表現（たとえば、1つまたは複数の前に再構成されたフレーム）に基づいて動き補償を使用せずに入力データ470（たとえば、ビデオフレーム）を圧縮するためにトレーニングされ得る。たとえば、エンコーダ部分462は、ビデオフレームを、そのビデオフレームからのビデオデータのみを使用し、前に再構成されたフレームのいかなるデータをも使用せずに圧縮することができる。

本明細書では、エンコーダ部分 4 6 2 によって処理されるビデオフレームは、イントラ予測フレーム（Iフレーム）と呼ばれることがある。いくつかの例では、Iフレームは、従来のビデオコーディング技法を使用して（たとえば、H E V C、V V C、M P E G - 4、または他のビデオコーディング規格に従って）生成され得る。そのような例では、プロセッサ 4 0 4 は、H E V C 規格に関して上記で説明されたものなど、ブロックベースのイントラ予測を実施するように構成されたビデオコーディングデバイス（たとえば、符号化デバイス）を含み得るか、またはそれに結合され得る。そのような例では、ニューラルネットワークシステム 4 1 0 は、プロセッサ 4 0 4 から除外され得る。

【0095】

[0118]いくつかの実装形態では、ニューラルネットワークシステム 4 1 1 のエンコーダ部分 4 3 2 は、出力データ 4 4 4 を生成するために入力データ 4 4 0 の不可逆的圧縮を実施するように構成され、入力データ 4 4 0 よりも少数のビットを有する出力データ 4 4 4 を生じる。エンコーダ部分 4 3 2 は、前の表現（たとえば、1つまたは複数の前に再構成されたフレーム）に基づいて動き補償を使用して入力データ 4 4 0（たとえば、ビデオフレーム）を圧縮するためにトレーニングされ得る。たとえば、エンコーダ部分 4 3 2 は、ビデオフレームを、そのビデオフレームからのビデオデータを使用し、前に再構成されたフレームのデータを使用して圧縮することができる。本明細書では、エンコーダ部分 4 6 2 によって処理されるビデオフレームは、イントラ予測フレーム（Pフレーム）と呼ばれることがある。動き補償は、残差情報とともに前に再構成されたフレームからのピクセルが現在のフレーム中の新しい位置にどのように移動したのかについて説明することによって現在のフレームのデータを決定するために使用され得る。動き補償技法の例について、図 6 に関して以下で説明される。

【0096】

[0119]特定の実装形態では、ニューラルネットワークシステム 4 1 0 のエンコーダ部分 4 6 2 は、ニューラルネットワーク 4 6 3 と量子化器 4 6 4 とを含む。ニューラルネットワーク 4 6 3 は、1つもしくは複数の畳み込みニューラルネットワーク（CNN）、1つもしくは複数の全結合ニューラルネットワーク、1つもしくは複数のゲート付き再帰型ユニット（GRU）、それらの任意の組合せ、および／または中間データ 4 7 2 を生成する他のタイプのニューラルネットワークアーキテクチャを含むことができる。中間データ 4 7 2 は量子化器 4 6 4 への入力である。ニューラルネットワークシステム 4 1 1 のエンコーダ部分 4 3 2 は、ニューラルネットワーク 4 3 3 と量子化器 4 3 4 とを含む。ニューラルネットワーク 4 3 3 は、1つもしくは複数の CNN、1つもしくは複数の全結合ニューラルネットワーク、1つもしくは複数の GRU、それらの任意の組合せ、および／または次いで量子化器 4 3 4 に入力される中間データ 4 4 2 を生成する他のタイプのニューラルネットワークアーキテクチャを含む。エンコーダ部分 4 6 2 およびエンコーダ部分 4 3 2 中に含まれ得る構成要素の例は、図 5 A～図 1 5 に示されている。

【0097】

[0120]量子化器 4 6 4 は、出力データ 4 7 4 を生成するために中間データ 4 7 2 の量子化、場合によっては、エントロピーコーディングを実施するように構成される。出力データ 4 7 4 は、量子化（場合によっては、エントロピーコード化）データを含むことができる。同様に、量子化器 4 3 4 は、出力データ 4 4 4 を生成するために中間データ 4 4 2 の量子化、場合によっては、エントロピーコーディングを実施するように構成される。出力データ 4 4 4 は、量子化（場合によっては、エントロピーコード化）データを含む。量子化器 4 6 4 および量子化器 4 3 4 によって実施される量子化動作は、同じ動作であり得、これは、中間データ 4 7 2 と中間データ 4 4 2 とから量子化コード（またはニューラルネットワークシステム 4 1 0 および／もしくはニューラルネットワークシステム 4 1 1 によって生成された量子化コードを表すデータ）の生成を生じることができる。量子化コード（または量子化コードを表すデータ）は、ラテントコード（latent codes）またはラテント（latent）と呼ばれることもある。本明細書では、ラテントに適用されるエントロピーモデルは、「プライア（prior）」と呼ばれることがある。いくつかの例では、量子化お

10

20

30

40

50

よびエントロピーコーディング動作は、既存のビデオコーディング規格に従ってビデオデータを符号化および／または復号するときに実施される既存の量子化およびエントロピーコーディング動作を使用して実施され得る。いくつかの例では、ニューラルネットワークシステム 410 および／またはニューラルネットワークシステム 411 によって行われ得る。たとえば、ニューラルネットワークシステム 410 および／またはニューラルネットワークシステム 411 は、教師つきトレーニングを使用してトレーニングされ得、トレーニング中に残差データが入力として使用され、量子化コードとエントロピーコードとが知られている出力（ラベル）として使用される。

【0098】

[0121]ニューラルネットワークシステム 410 のデコーダ部分 466 は、（たとえば、直接量子化器 464 からおよび／または記憶媒体 414 から）出力データ 474 を受信することと、出力データ 474 に少なくとも部分的に基づいて入力データ 470 の表現 476 を生成するために出力データ 474 を処理することとを行うように構成される。同様に、ニューラルネットワークシステム 411 のデコーダ部分 436 は、（たとえば、直接量子化器 434 からおよび／または記憶媒体 414 から）出力データ 444 を受信することと、出力データ 444 に少なくとも部分的に基づいて入力データ 440 の表現 446 を生成するために出力データ 444 を処理することとを行うように構成される。場合によっては、デコーダ部分 436 は、以下でさらに説明するように、出力データ 444 と第 2 の状態データ 452 とに基づいて表現 446 を生成するように構成される。場合によっては、デコーダ部分 436 は、以下でさらに説明するように、出力データ 444 と第 2 の状態データ 452 とに基づき、さらに、前の時間ステップ $t-1$ において（たとえば、前の時間ステップ $t-1$ において入力データ 470 を処理するとき）ニューラルネットワークシステム 410 のデコーダ部分 466 によって生成された表現 476 に基づいて現在の時間ステップ t において表現 446 を生成するように構成される。場合によっては、デコーダ部分 436 は、以下でさらに説明するように、出力データ 444 と第 2 の状態データ 452 とに基づき、さらに、前の時間ステップ $t-1$ において（たとえば、前の時間ステップ $t-1$ において入力データ 440 を処理するとき）デコーダ部分 436 によって生成された表現 446 に基づいて現在の時間ステップ t において出力データ 446 を生成するように構成される。

【0099】

[0122]特定の実装形態では、ニューラルネットワークシステム 410 のデコーダ部分 466 は、1つもしくは複数の CNN、1つもしくは複数の全結合ニューラルネットワーク、1つもしくは複数の GRU、それらの任意の組合せ、および／または他のタイプのニューラルネットワークアーキテクチャを含み得るニューラルネットワーク 468 を含む。ニューラルネットワークシステム 411 のデコーダ部分 436 は、1つもしくは複数の CNN、1つもしくは複数の全結合ニューラルネットワーク、1つもしくは複数の GRU、1つもしくは複数の長短期記憶 (LSTM) ネットワーク、1つもしくは複数の ConvRNN、1つもしくは複数の ConvGRU、1つもしくは複数の ConvLSTM、それらの任意の組合せ、および／または他のタイプのニューラルネットワークアーキテクチャを含み得るニューラルネットワーク 438 を含む。デコーダ部分 436 中に含まれ得る構成要素の例は、図 5A～図 15 に示されている。

【0100】

[0123]デコーダ部分 436 のニューラルネットワーク 438 は、1つもしくは複数の RNN、GRU、LSTM、ConvRNN、ConvGRU、ConvLSTM、それらの任意の組合せ、および／または1つもしくは複数の再帰レイヤをもつ他のニューラルネットワークアーキテクチャを含むことができる再帰構成要素を有する。再帰構成要素（たとえば、1つまたは複数の再帰レイヤ）は、第 1 の状態データ 450 と第 2 の状態データ 452 とを与えることができる。第 1 の状態データ 450 と第 2 の状態データ 452 とは、入力データ 440 の 1 つのセットのための表現 446 の生成から生じたニューラルネットワーク 438 の状態に対応する。（たとえば、現在のビデオフレームの処理中などの時

10

20

30

40

50

間ステップ t に関連する) 第 1 の状態データ 450 と第 2 の状態データ 452 とは、(たとえば、現在のビデオフレームの後に処理される次のビデオフレームの処理中などの次の時間ステップ $t+1$ における) 入力データ 440 の次の連続セットのための、それぞれ、エンコーダ部分 432 およびデコーダ部分 436 への入力として与えられる。例示的な例では、状態データ 450、452 中に表されるデコーダ部分 436 の状態は、ニューラルネットワーク 438 内のノードの 1 つまたは複数の値を含む。したがって、ニューラルネットワークシステム 411 は、入力データ 440 とトレーニングセット (たとえば、入力画像と出力画像とのトレーニングセット) を介した入力データ 440 の表現 446 との間の差を低減する (たとえば、最小化する) ためにトレーニングされ得る再帰ベースのニューラルネットワークシステムとして動作する。いくつかの実装形態では、第 1 の状態データ 450 は、第 2 の状態データ 452 に一致する (たとえば、第 1 の状態データ 450 は、第 2 の状態データ 452 と同じである)。他の実装形態では、第 1 の状態データ 450 は、第 2 の状態データ 452 とは異なることができる。

10

【0101】

[0124] 図 4 に示されているように、フィードバック情報は、ニューラルネットワークシステム 411 のデコーダ部分 436 からニューラルネットワークシステム 411 のエンコーダ部分 432 に与えられる。フィードバックは、復号段階において導入されたエラーを表すために使用され得、これは、それらのエラーを補償するためにエンコーダ部分 432 によって使用され得る。次の時間ステップ t においてデコーダ部分 436 からエンコーダに与えられ得る前の時間ステップ $t-1$ からの情報は、デコーダ部分 436 のニューラルネットワーク 438 からの (

20

【0102】

【数 1】

$$\hat{x}_t$$

【0103】

として示される) 前に再構成されたフレーム、前に再構成された動き推定、前に再構成された残差、および／または前の再帰状態のうちの 1 つまたは複数の値を含むことができる。たとえば、時間相関を示し得るビデオフレームのシーケンスの場合、ビデオのフレームに関連するデコーダ部分 436 の前の再帰状態は、ビデオの次のフレームを処理するときにエンコーダ部分 432 とデコーダ部分 436 との両方への入力として与えられ得る。他のフィードバック情報に加えてそのような再帰状態を使用することは、圧縮比 (たとえば、出力データ 444 中のビット数と比較した入力データ 440 中のビット数) を増加させることができ、したがって、デコーダ部分 436 が遭遇するエラーをエンコーダ部分 432 が考慮することを可能にすることによってコーディング効率を増加させることができる。例示的な例として、状態データ 450、452 は、エンコーダ部分 432 とデコーダ部分 436 とが一連のフレームの次のフレームのための予測を生成することを可能にする入力として機能することができ、エンコーダ部分 432 は、デコーダ部分 436 の履歴を考慮に入れることなしに出力データ 444 を符号化することと比較して低減されたビットレートにおいて出力データ 444 として残差 (たとえば、次のフレームと次のフレームの予測との間の差) を符号化することができる。したがって、符号化データは、低減された量のストレージスペースを使用して記憶され、低減された量のネットワーク帯域幅を使用してもしくはより高い送信レートで送信され、またはそれらの任意の組合せで行われ得る。

30

40

【0104】

[0125] プロセッサ 404 は、送信媒体 418 または記憶媒体 414 のうちの少なくとも 1 つに出力データ 444 を送るよう構成される。たとえば、出力データ 444 は、出力データ 444 を生成するために符号化／圧縮されているビデオデータの再生のためになど、再構成されたデータとして入力データ 440 の表現 446 を生成するためにデコーダ部分 436 による後での取出しおよび復元のために記憶媒体 414 において記憶され得る。いくつかの実装形態では、出力データ 444 は、再構成されたデータとして入力データ 4

50

40の表現446を生成するために（たとえば、デバイス402中のまたは別のデバイス中の）デコーダ部分436に一致する別のデコーダデバイスにおいて復号され得る。別の例として、第2のデバイス490は、デコーダ部分436に一致する（または実質的に一致する）デコーダを含み得、出力データ444は、第2のデバイス490において再構成されたデータとして入力データ440の表現446を生成するために送信媒体418を介して送信され得る。

【0105】

[0126]システム400の構成要素は、電子回路もしくは他の電子ハードウェアを含むことができ、および／またはこれらを使用して実装されることが可能であり、これは、1つもしくは複数のプログラム可能な電子回路（たとえば、マイクロプロセッサ、グラフィック処理ユニット（GPU）、デジタル信号プロセッサ（DSP）、中央処理装置（CPU）、および／または他の適切な電子回路）を含むことができ、ならびに／または、本明細書において説明される様々な動作を実施するためのコンピュータソフトウェア、ファームウェア、もしくはこれらの任意の組合せを含むことができ、および／もしくは、これらを使用して実装され得る。

【0106】

[0127]システム400がいくつかの構成要素を含むように示されているが、システム400が図4に示されている構成要素よりも多いまたは少ない構成要素を含むことができることを、当業者は諒解されよう。たとえば、システム400はまた、入力デバイスと出力デバイス（図示せず）とを含むコンピューティングデバイスを含むことができるか、またはその部分であり得る。いくつかの実装形態では、システム400はまた、1つもしくは複数のメモリデバイス（たとえば、1つもしくは複数のランダムアクセスメモリ（RAM）構成要素、読取り専用メモリ（ROM）構成要素、キャッシュメモリ構成要素、バッファ構成要素、データベース構成要素、および／もしくは他のメモリデバイス）、1つもしくは複数のメモリデバイスと通信しているおよび／またはそれに電氣的に接続された1つもしくは複数の処理デバイス（たとえば、1つもしくは複数のCPU、GPU、および／もしくは他の処理デバイス）、ワイヤレス通信を実施するための（たとえば、各ワイヤレスインターフェースのための1つもしくは複数のトランシーバとベースバンドプロセッサとを含む）1つもしくは複数のワイヤレスインターフェース、1つもしくは複数の有線接続を介した通信を実施するための1つもしくは複数のワイヤードインターフェース（たとえば、ユニバーサルシリアルバス（USB）入力などのシリアルインターフェース、ライトニングコネクタ、および／もしくは他のワイヤードインターフェース）、ならびに／または図4に示されていない他の構成要素を含むコンピューティングデバイスを含み得るか、またはその部分であり得る。

【0107】

[0128]いくつかの実装形態では、システム300は、コンピューティングデバイスによって局所的に実装され、および／またはその中に含まれ得る。たとえば、コンピューティングデバイスは、モバイルデバイス、パーソナルコンピュータ、タブレットコンピュータ、仮想現実（VR）デバイス（たとえば、ヘッドマウントディスプレイ（HMD）もしくは他のVRデバイス）、拡張現実（AR）デバイス（たとえば、HMD、AR眼鏡、もしくは他のARデバイス）、ウェアラブルデバイス、（たとえば、サービスとしてのソフトウェア（SaaS）システムもしくは他のサーバベースのシステム中の）サーバ、テレビジョン、および／または本明細書で説明される技法を実施するリソース能力をもつ任意の他のコンピューティングデバイスを含むことができる。

【0108】

[0129]一例では、ニューラルネットワークシステム411（および、場合によっては、ニューラルネットワークシステム410）は、プロセッサ404に結合され、プロセッサ404によって実行可能な命令を記憶するように構成されたメモリ406と、アンテナとプロセッサ404とに結合され、リモートデバイスに出力データ444を送信するように動作可能なワイヤレストランシーバとを含むポータブル電子デバイスに組み込まれ得る。

【0109】

【0130】図5Aは、（図4におけるニューラルネットワークシステム410の一例として）ニューラルネットワークシステム510と（ニューラルネットワークシステム411の一例として）ニューラルネットワークシステム511とを含むシステムを示す図である。ニューラルネットワークシステム510のIフレームエンコーダ563は、前に再構成されたフレームのデータに基づいて動き補償を実施することなしにビデオフレームを、そのビデオフレームのみからのビデオデータを使用して符号化または圧縮するためにトレーニングされる（たとえば、1つまたは複数のCNNを含む）ニューラルネットワークを含むことができる。たとえば、図5Aに示されているように、Iフレームエンコーダ563は、入力としてビデオフレーム $x_{I-frame}$ を受信し、ビデオフレーム $x_{I-frame}$ に含まれるデータのみに使用してビデオフレーム $x_{I-frame}$ を符号化または圧縮する。Iフレームエンコーダ563は、中間データ（たとえば、中間データ472）を生成するためにビデオフレーム $x_{I-frame}$ を符号化または圧縮することができる。場合によっては、中間データは、残差を表すデータを含むことができ、これは、ビデオフレーム $x_{I-frame}$ の予測と実際の入力ビデオフレーム $x_{I-frame}$ との間の差を表す。

10

【0110】

【0131】ニューラルネットワークシステム510の量子化器564は、図4に示されている量子化器464の一例である。量子化器564は、Iフレームエンコーダ563からの中間データ出力からラテントを生成する。場合によっては、ラテントは、圧縮ビデオフレーム $x_{I-frame}$ の量子化バージョン（たとえば、残差の量子化バージョン）を表す量子化コードを含む。場合によっては、ラテントは、量子化コードを表すニューラルネットワークデータ（たとえば、ノードの活性化マップまたは特徴マップ）を含むことができる。量子化器564はまた、Iフレームプライアを含む。Iフレームプライアは、エントロピーコードを表すエントロピーコードまたはニューラルネットワークデータ（たとえば、ノードの活性化マップまたは特徴マップ）を生成するためにラテントに適用されるエントロピーモデルを含むか、またはそれを表すことができる。量子化器564は、後での取出しのためにIフレームデコーダ568および／または記憶装置（たとえば、記憶媒体414）に出力される出力データ（たとえば、出力データ474）を生成する。たとえば、量子化器564からの出力データは、送信媒体418を介してデバイス490に送信され得る。

20

【0111】

【0132】ニューラルネットワークシステム510のIフレームデコーダ568は、再構成されたフレーム

30

【0112】

【数2】

$$\hat{x}_{I-frame}$$

【0113】

を生成するために量子化器564からの出力データを処理する。いくつかの実装形態では、Iフレームエンコーダ563と、量子化器564と、Iフレームデコーダ568とは、徹底的に一緒にトレーニングされる単一のニューラルネットワークまたは複数のニューラルネットワークを使用して実装され得る。図5Aに示されている損失関数

40

【0114】

【数3】

$$Loss0 = \sum_t distortion(x_t, \hat{x}_t)$$

【0115】

が、入力画像と再構成された出力画像とを含むトレーニングセットを用いて教師つき学習技法を使用して単一のニューラルネットワークをトレーニングするために使用され得る。ニューラルネットワークのトレーニングに関するさらなる詳細について、ニューラルネットワークシステム511に関して以下で説明される。いくつかの実装形態では、Iフレー

50

ムエンコーダ５６３と、Ｉフレームデコーダ５６８と、量子化器５６４とのニューラルネットワークは、別々にトレーニングされ得る。

【０１１６】

【０１３３】上記のように、いくつかの実装形態では、Ｉフレームエンコーダ５６３と、Ｉフレームデコーダ５６８と、量子化器５６４とは、Ｉフレームを生成するために（たとえば、ＨＥＶＣ、ＶＶＣ、ＭＰＥＧ－４、または他のビデオコーディング規格に従って）従来のビデオコーディング技法を実施することができる。たとえば、（ビデオ符号化および復号を実施することができる）ビデオコーディングデバイスは、ＨＥＶＣ規格または他のビデオコーディング規格に従ってＩフレームを生成するためにブロックベースのイントラ予測を実施するように構成され得る。

10

【０１１７】

【０１３４】ニューラルネットワークシステム５１１は、場合によっては、ビデオコーディング（たとえば、符号化および／または復号）を実施するためにニューラルネットワークシステム５１１が動き補償技法を使用するのでＰフレームオートエンコーダと呼ばれることがある。ニューラルネットワークシステム５１１のエンコーダ５３３は、１つまたは複数の前に再構成されたフレームのデータに基づいて動き補償を使用してビデオフレームを符号化または圧縮するためにトレーニングされる（たとえば、１つまたは複数のＣＮＮを含む）ニューラルネットワークを含むことができる。ニューラルネットワークシステム５１１は、 t と $t+1$ との２つの時間ステップにおいて図５Ａに示されている。図５Ａに示されている時間ステップ t は、Ｉフレームが生成された（たとえば、

20

【０１１８】

【数４】

$$\hat{x}_{I-frame}$$

【０１１９】

）後のニューラルネットワークシステム５１１の第１の時間ステップの動作を示す。時間ステップ t において、エンコーダ５３３は、入力としてビデオフレーム x_t と前に再構成されたフレーム $x_{I-frame}$ とを受信し、ビデオフレーム x_t とビデオフレーム $x_{I-frame}$ との間の動きの量を決定するために動き補償を実施することによってビデオフレーム x_t を符号化する。

30

【０１２０】

【０１３５】様々なタイプの動き補償がエンコーダ５３３によって実施され得る。たとえば、エンコーダ５３３は、（特に、ＨＥＶＣ、ＶＶＣ、ＭＰＥＧ－４などのビデオコーディング規格によって行われるものと同様の）ブロックベースの動き補償、オプティカルフローベースの動き補償、動的な（もしくは適応型の）畳み込みベースの動き補償、それらの任意の組合せ、および／または他のタイプの動き補償を実施することができる。様々な動き補償のオプションについて、図６に関して以下で説明される。

【０１２１】

【０１３６】ビデオフレーム x_t を符号化することの結果として、エンコーダ５３３は、中間データ（たとえば、中間データ４４２）を生成することができる。場合によっては、中間データは、残差を表すデータと、ビデオフレーム x_t とビデオフレーム $x_{I-frame}$ との間の動きの量を表す動きデータとを含むことができる。残差は、ビデオフレーム x_t の予測と実際の入力ビデオフレーム x_t との間の差を表すことができる。ビデオフレーム x_t とビデオフレーム $x_{I-frame}$ との間の動きの量を表す動きデータは、従来のビデオコーディング技法（たとえば、特に、ＨＥＶＣ、ＶＶＣ）における動きベクトルに類似するものとして概念化され得る。

40

【０１２２】

【０１３７】ニューラルネットワークシステム５１１の量子化器５３４は、図４に示されている量子化器４３４の一例である。量子化器５３４は、エンコーダ５３３からの中間データ出力からラテントを生成する。場合によっては、ラテントは、符号化されたビデオフレー

50

ム x_t の量子化バージョン（たとえば、残差の量子化バージョンおよび動きデータの量子化バージョン）を表す量子化コードを含む。場合によっては、ラテントは、量子化コードを表すニューラルネットワークデータ（たとえば、ノードの活性化マップまたは特徴マップ）を含むことができる。量子化器 534 はまた、エントロピーコードを表すエントロピーコードまたはニューラルネットワークデータ（たとえば、ノードの活性化マップまたは特徴マップ）を生成するためにラテントに適用されるエントロピーモデルを含むか、またはそれを含むことができるプライアを含む。量子化器 534 は、後での取出しのためにデコーダ 538 および／または記憶装置（たとえば、記憶媒体 414）に出力される出力データ（たとえば、出力データ 444）を生成する。たとえば、量子化器 534 からの出力データは、送信媒体 418 を介してデバイス 490 に送信され得る。

10

【0123】

[0138]ニューラルネットワークシステム 511 のデコーダ 538 は、再構成されたフレーム

【0124】

【数5】

$$\hat{x}_t$$

【0125】

を生成するために出力データを処理する。デコーダ 538 は、量子化器 534 から、記憶装置（たとえば、記憶媒体 414）から、または送信媒体（たとえば、送信媒体 418）を介して別のデバイスから出力データを受信することができる。図 5A に示されているように、デコーダ 538 は、1 つまたは複数の再帰前デコーダレイヤ 570 と、1 つまたは複数の再帰レイヤ 571 と、1 つまたは複数の再帰後レイヤ 572 とを含む再帰レイヤまたは構成要素を含む。1 つまたは複数の再帰レイヤ 571 は、状態情報を記憶することができ、後続の時間ステップにおいて（たとえば、図 5A に示されているように、時間ステップ $t+1$ において）デコーダ 538 および／またはエンコーダ 533 のいくつかのレイヤに状態情報をフィードバックすることができる。後の時間ステップにおいてニューラルネットワークシステム 511 中の前のレイヤに与えられたフィードバックは、メモリのタイプをニューラルネットワークシステム 511 に与える。1 つまたは複数の再帰前デコーダレイヤ 570 と 1 つまたは複数の再帰後レイヤ 572 とは、一方向にネットワークの他のレイヤを通して情報を前方にパスする順方向レイヤを含むことができるが、1 つまたは複数の再帰レイヤは、上記のように、フィードバックを通して順方向におよび／または逆方向に情報をパスすることができる。

20

【0126】

[0139]デコーダ 538 の再帰状態は、 h として図 5A に示されている。ニューラルネットワークシステム 510 が I フレームを生成した時間ステップ $t-1$ に対応する図 5A に示されている再帰状態 h_{t-1} は、（たとえば、ニューラルネットワークシステム 510 がいかなる再帰レイヤも備えていないので）1 つまたは複数の 0 値で開始され得る。図 5A の再帰状態 h_t は、タイムスタンプ t におけるデコーダ 538 の再帰状態を表す。再帰状態 h_t は、1 つまたは複数の再帰レイヤ 571 からのニューラルネットワーク情報を含むことができる。たとえば、ニューラルネットワーク情報は、（たとえば、デコーダ 538 を通してパスされるデータによって乗算された 1 つまたは複数の再帰レイヤ 571 の重みの乗算に基づいて生成される）活性化マップまたは特徴マップを含むことができる。図 5A に示されているように、再帰状態 h_t は、次の時間ステップ $t+1$ において 1 つまたは複数の再帰レイヤ 571 とエンコーダ 533 とにフィードバックとしてパスされる。

30

40

【0127】

[0140]出力データの処理に基づいて、デコーダ 538 の 1 つまたは複数の再帰後レイヤ 572 は、再構成された動き推定データ

【0128】

【数6】

50

$\hat{f}_t$

【0 1 2 9】

と再構成された残差データ

【0 1 3 0】

【数 7】

 $\hat{f}_t$

【0 1 3 1】

とを出力する。場合によっては、再構成された動き推定データ

【0 1 3 2】

【数 8】

 $\hat{f}_t$

【0 1 3 3】

は、図 6 に関して以下で説明されるように、オプティカルフローデータ（たとえば、変位ベクトル）、動的畳み込みデータ（たとえば、データ畳み込みのための行列もしくはカーネル）、またはブロックベースの動きデータ（たとえば、動きベクトル）を含むことができる。再構成された動き推定データ

【0 1 3 4】

【数 9】

 $\hat{f}_t$

【0 1 3 5】

は、前に再構成されたフレーム

【0 1 3 6】

【数 10】

 $\hat{x}_{I-frame}$

【0 1 3 7】

を（たとえば、動き補償を実施することによって）ワーピングする（warp）ためにデコーダ 5 3 8 のワーピングエンジン 5 6 7 によって使用され、ワーピングしたフレーム（warped frame）の生成を生じる。たとえば、前に再構成されたフレーム

【0 1 3 8】

【数 11】

 $\hat{x}_{I-frame}$

【0 1 3 9】

のピクセルは、新しいロケーションに移動されるか、あるいは再構成された動き推定データ

【0 1 4 0】

【数 12】

 $\hat{f}_t$

【0 1 4 1】

に従って修正され得る。ワーピングエンジン 5 6 7 によって出力されたワーピングしたフレームは、加算器エンジン 5 6 9 に出力される。加算器エンジン 5 6 9 は、現在の時間ステップ t のための再構成されたフレーム

【0 1 4 2】

【数 13】

10

20

30

40

50

$\hat{x}_t$

【0143】

を生成するために再構成された残差データ

【0144】

【数14】

 $\hat{r}_t$

【0145】

にワーピングしたフレームを加算する（または組み合わせる）。

10

【0146】

【0141】再帰状態情報（たとえば、再帰状態 h_t ）に加えて、再構成されたフレーム

【0147】

【数15】

 $\hat{x}_t$

【0148】

はまた、次の時間ステップ $t+1$ の動作のためにニューラルネットワークシステム 511 にフィードバックされ得る。たとえば、図 5 A に示されているように、再構成されたフレーム

20

【0149】

【数16】

 $\hat{x}_t$

【0150】

と再帰状態 h_t とは、エンコーダ 533 への入力として与えられる。場合によっては、図 5 A に示されているように、前に再構成された動き推定データ

【0151】

【数17】

 $\hat{r}_t$

30

【0152】

と前に再構成された残差データ

【0153】

【数18】

 $\hat{r}_t$

【0154】

とはまた、次の時間ステップ $t+1$ の動作においてエンコーダ 533 への入力としてフィードバックされ得る。時間ステップ $t+1$ において、エンコーダ 533 は、現在の入力ビデオフレーム

40

【0155】

【数19】

 $\hat{x}_{t+1}$

【0156】

のコーディングされたバージョンを生成するために、現在の入力ビデオフレーム

【0157】

【数20】

 $\hat{x}_{t+1}$

50

【0158】

と、前に再構成されたフレーム

【0159】

【数21】

$$\hat{x}_t$$

【0160】

と、再帰状態 h_t と、前に再構成された動き推定データ

【0161】

【数22】

$$\hat{f}_t$$

10

【0162】

と、前に再構成された残差データ

【0163】

【数23】

$$\hat{r}_t$$

【0164】

とを処理することができる。量子化器534は、ラテントを生成するためにコーディングされたビデオデータを量子化することができ、出力データを生成するためにプライアを使用してラテントに対して（プライアを使用して）エントロピーコーディングを実施することができる。

20

[0142]デコーダ538の1つまたは複数の再帰前デコーダレイヤ570と、1つまたは複数の再帰レイヤ571と、1つまたは複数の再帰後レイヤ572とは、現在の時間ステップ $t+1$ のための出力データを処理することができ、再構成された動き推定データ

【0165】

【数24】

$$\hat{f}_{t+1}$$

30

【0166】

と再構成された残差データ

【0167】

【数25】

$$\hat{r}_{t+1}$$

【0168】

とを出力することができる。デコーダ538のワーピングエンジン567は、現在の時間ステップ $t+1$ のためのワーピングしたフレームを生成するために、前に再構成されたフレーム

40

【0169】

【数26】

$$\hat{x}_t$$

【0170】

を（たとえば、動き補償を実施することによって）ワーピングすることができる。加算器エンジン569は、現在の時間ステップ $t+1$ のための再構成されたフレーム

【0171】

【数27】

50

$$\hat{x}_{t+1}$$

【0172】

を生成するために、再構成された残差データ

【0173】

【数28】

$$\hat{r}_{t+1}$$

【0174】

にワーピングしたフレームを加算することができる。

【0175】

[0143]再構成されたフレーム

【0176】

【数29】

$$\hat{x}_{t+1}$$

【0177】

と、再帰状態 h_{t+1} と、再構成された動き推定データ

【0178】

【数30】

$$\hat{f}_{t+1}$$

【0179】

と、再構成された残差データ

【0180】

【数31】

$$\hat{r}_{t+1}$$

【0181】

とは、次の時間ステップ $t+2$ の動作においてエンコーダ533への入力としてフィードバックされ得る。ニューラルネットワークシステム511は、（たとえば、蓄積し得るあらゆる動きエラーをリセットするなどのためのビデオ中のある時点へのランダムアクセスのために）次のIフレームが必要とされるまでフレームごとにフレームのビデオシーケンスのための動き推定ベースのコーディングプロセスを続けることができる。Iフレームが必要とされるとき、ビデオシーケンス中のビデオの次のフレームが、ニューラルネットワークシステム510によって符号化され得る。

【0182】

[0144]いくつかの実装形態では、エンコーダ533と、量子化器534と、デコーダ538とは、徹底的に一緒にトレーニングされる単一のニューラルネットワークまたは複数のニューラルネットワークを使用して実装され得る。図5Bは、ニューラルネットワークシステム511のために使用され得るニューラルネットワークアーキテクチャの一例を示す図である。いくつかの実装形態では、図5Bに示されているものと同様のニューラルネットワークアーキテクチャが、ニューラルネットワークシステム510のために使用され得るが、再帰レイヤ（たとえば、ConvGRUレイヤ）がない。

【0183】

[0145]図5Bに示されているように、エンコーダ533への入力は、エンコーダ533によって符号化されることになる現在のフレーム

【0184】

【数32】

10

20

30

40

50

$$\hat{x}_t$$

【0185】

と、（たとえば、ニューラルネットワークシステム510もしくはニューラルネットワークシステム511からの）前に再構成されたフレーム

【0186】

【数33】

$$\hat{x}_{t-1}$$

【0187】

と、前の再帰状態 h_{t+1} とを含む。エンコーダに与えられ得る他の入力は、前に再構成された動き推定データ

【0188】

【数34】

$$\hat{f}_{t-1}$$

【0189】

と前に再構成された残差データ

【0190】

【数35】

$$\hat{r}_{t-1}$$

【0191】

とを含む。

【0192】

【0146】入力は、（Conv c:64 k:5 s:2として示され、ここで、cは、チャンネル（たとえば、カーネルまたはフィルタ）の数であり、kは、カーネルサイズであり、sは、所与の畳み込みレイヤに関連するストライドである）第1の畳み込みレイヤに与えられる。畳み込みレイヤは、受容フィールド（receptive field）と呼ばれるノード（たとえば、現在のフレーム

【0193】

【数36】

$$\hat{x}_t$$

【0194】

のピクセル）の領域を処理することなどによって畳み込み方式で入力データを処理する。畳み込みレイヤは、（フィルタとも呼ばれる）1つまたは複数のカーネルから構成され得、ここで、各カーネルは、異なる活性化または特徴マップに対応することができ、カーネルの各畳み込み反復は、畳み込みレイヤのノードまたはニューロンである。たとえば、カーネルが各畳み込み反復においてカバーする現在のフレーム

【0195】

【数37】

$$\hat{x}_t$$

【0196】

の領域は、カーネルのための受容フィールドであろう。1つの例示的な例では、フレーム

【0197】

【数38】

$$\hat{x}_t$$

【0198】

が 28×28 のアレイを含み、各カーネル（と対応する受容フィールドと）は 5×5 のアレイ（ここで、 $k = 5$ ）である場合、畳み込みレイヤ中に 24×24 個のノードがあることになる。ビデオフレームが、一般に、 $1080p$ ビデオのための 1920×1080 など、 28×28 よりも大きい解像度を有することに留意されたい。ノードとそのノードのための受容フィールドとの間のそれぞれ接続は重みを学習し、場合によっては、各ノードが入力中のその特定のローカル受容フィールドを分析することを学習するように全体的なバイアスを学習する。畳み込みレイヤの各ノードは、同じ重みを有することができ、いくつかの事例では、（共有重みおよび共有バイアスと呼ばれる）バイアスを有することができる。たとえば、カーネルは、入力として重み（数）のアレイと同じ深度とを有する。一例では、カーネルは、フレーム

10

【0199】

【数39】

$$\hat{x}_t$$

【0200】

の3つの色構成要素（たとえば、赤（R）、緑（G）、および青（B）の色構成要素）に従って現在のフレーム

【0201】

【数40】

$$\hat{x}_t$$

20

【0202】

のために3の深度を有することができる。

【0203】

【0147】畳み込みレイヤの畳み込み性質は、畳み込みレイヤの各ノードがその対応する受容フィールドに適用されることに起因する。たとえば、畳み込みレイヤのカーネルは、現在のフレーム

【0204】

【数41】

$$\hat{x}_t$$

30

【0205】

のアレイの左上隅で開始することができ、フレーム

【0206】

【数42】

$$\hat{x}_t$$

【0207】

の周りに畳み込むことができる。上記のように、カーネルの各畳み込み反復は、畳み込みレイヤのノードまたはニューロンと見なされ得る。各畳み込み反復において、カーネルの値は、画像の元のピクセル値の対応する数で乗算される（たとえば、 5×5 のカーネルアレイは、フレーム

40

【0208】

【数43】

$$\hat{x}_t$$

【0209】

アレイの左上隅の入力ピクセル値の 5×5 のアレイによって乗算される）。各畳み込み反復からの乗算は、その反復またはノードのための総和を取得するために一緒に加算され得

50

る。プロセスは、次に、畳み込みレイヤ中の次のノードの受容フィールドに従ってフレーム
【0210】
【数44】

$$\hat{x}_t$$

【0211】

中の次のロケーションにおいて続けられる。たとえば、カーネルは、次の受容フィールド
に向かって（図5Bにおいてsによって示される）ストライド量だけ移動され得る。スト
ライド量は、1、2、または他の好適な量に設定され得る。たとえば、ステップ量が1に
設定される場合、カーネルは、畳み込み反復ごとに1ピクセルだけ右に移動されることにな
る。入力量のそれぞれの一意のロケーションにおいてカーネルを処理することは、その
ロケーションのためのカーネル結果を表す数を生成し、畳み込みレイヤのノードごとに総
和値が決定されることを生じる。

10

【0212】

【0148】入力レイヤから畳み込み隠れレイヤ622aへのマッピングは、活性化マップ（
または特徴マップ）と呼ばれる。活性化マップは、入力量の各ロケーションにおけるカー
ネル結果を表す各ノードの値を含む。活性化マップは、入力量に対するカーネルの各反復
から生じる様々な総和値を含むアレイ（たとえば、現在のフレーム

【0213】

【数45】

$$\hat{x}_t$$

20

【0214】

）を含むことができる。たとえば、5×5のカーネルが28×28の入力フレームの各ピ
クセル（1のステップ量）に適用される場合、活性化マップは、24×24のアレイを含
むことになる。畳み込みレイヤは、入力の複数の特徴を識別するためにいくつかの活性化
マップを含むことができる。たとえば、3つの活性化マップを使用して、畳み込みレイヤ
は、各特徴が入力全体にわたって検出可能である状態で3つの異なる種類の特徴を検出す
ることができる。

【0215】

30

【0149】正規化レイヤと非線形レイヤとは、第1の畳み込みレイヤの後に適用され得る。
正規化レイヤは、隠れレイヤ（たとえば、畳み込みレイヤ）中の入力を非線形に正規化す
るために（たとえば、トレーニング中に）使用され得る。正規化レイヤの1つの例示的な
例は、図5Bのアーキテクチャで使用されているBatchNormを含む。1つの例示
的な例では、BatchNormは、次のように隠れレイヤ z^k を正規化することができる
。

【0216】

【数46】

$$z^{n,k} = \frac{z^k - E[z^k]}{\sqrt{V[z^k]}} \quad \text{式(1)}$$

40

【0217】

【0150】ここで、 $E[x]$ は、 x の第1のモーメント（たとえば、平均または期待値）で
あり、 $V[x]$ は、 x の第2のモーメント（たとえば、分散）である。

【0218】

【0151】非線形レイヤは、線形演算を計算していたシステムに非線形性を導入するために
使用され得る。非線形レイヤの1つの例示的な例は、図5Bのアーキテクチャで使用され
ているように整流線形ユニット（ReLU: rectified linear unit）である。ReLUレ
イヤは、入力量中の値のすべてに関数 $f(x) = \max(0, x)$ を適用することができ

50

、これは、すべての負の活性化を0に変更する。したがって、ReLUは、畳み込みレイヤの受容フィールドに影響を及ぼすことなしに、ニューラルネットワークの非線形プロパティを増加させることができる。

【0219】

【0152】第1のReLUの出力は、畳み込みレイヤと、BatchNormレイヤと、ReLUレイヤとの第2のセットに与えられる。第2のReLUの出力は、第1の残差レイヤ(ResBlock B)に与えられ、これは、その出力を第2の残差レイヤ(ResBlock A)に与える。ResBlock AとResBlock Bとの詳細も図5Bに示されている。ResBlock Aの出力は、第2のReLUの出力と組み合わせられ(またはそれに加算され)、組み合わせられた出力は、第3の畳み込みレイヤに与えられる。第3の畳み込みレイヤの出力は、第3のBatchNormレイヤに与えられ、これは、(ソフト/ハード量子化として示される)量子化レイヤにデータを出力する。ソフト/ハード量子化レイヤは、1つまたは複数の量子化された(場合によっては、エン트로ピーコード化された)係数 z_t を出力する。1つまたは複数の係数 z_t が、場合によっては、記憶され得、場合によっては、デコーダ538に与えられ得る。

【0220】

【0153】デコーダ538への入力は、1つまたは複数の係数 z_t である。1つまたは複数の係数 z_t は、(Tconv c:128 k:3 s:2として示される)第1の畳み込みレイヤと、それに続く第1のBatchNormレイヤおよび第1のReLUレイヤとによって処理される。第1のReLUの出力は、第1の残差レイヤ(ResBlock B)に与えられ、これは、その出力を第2の残差レイヤ(ResBlock A)に与える。ResBlock Aの出力は、第1のReLUレイヤの出力と組み合わせられ(またはそれに加算され)、組み合わせられた出力は、(Tconv c:64 k:5 s:2として示される)第2の畳み込みレイヤに与えられる。第2の畳み込みレイヤの後に、第2のBatchNormレイヤと第2のReLUレイヤとが続き、これは、(Tconv c:C2 k:5 s:2として示される)第3の畳み込みレイヤと、それに続く第3のBatchNormレイヤとにデータを出力する。第3のBatchNormレイヤの出力は、(ConvGRU c:C3 k:3 s:1として示される)再帰レイヤに与えられる。図示のように、再帰レイヤへの別の入力、前の時間ステップ $t-1$ のための前の再帰状態 h_{t-1} であり、再帰レイヤの出力は、現在の時間ステップ t のための再帰状態を含む。再帰レイヤの出力は、(Conv c:5 k:5 s:1として示される)最後の畳み込みレイヤに与えられ、これは、再構成された動き推定データ

【0221】

【数47】

$$\hat{f}_t$$

【0222】

と再構成された残差データ

【0223】

【数48】

$$\hat{r}_t$$

【0224】

とを出力する。

【0225】

【0154】上記のように、デコーダ538のワーピングエンジン567は、現在の時間ステップ t のためのワーピングしたフレームを生成するために、前に再構成されたフレーム

【0226】

【数49】

$$\hat{x}_{t-1}$$

【0227】

を（たとえば、動き補償を実施することによって）ワーピングすることができ、加算器エンジン569は、現在の時間ステップ t のための再構成されたフレーム

【0228】

【数50】

$\hat{x}_t$

【0229】

を生成するために、再構成された残差データ

【0230】

【数51】

$\hat{r}_t$

【0231】

にワーピングしたフレームを加算することができる。いくつかの実装形態では、ワーピングエンジン567および／または加算器エンジン569は、図7に示されているエンコーダの構成などで、（たとえば、1つもしくは複数の畳み込みレイヤ、1つもしくは複数の正規化レイヤ、1つもしくは複数の非線形レイヤ、1つもしくは複数の残差ブロック、または他のニューラルネットワークレイヤを使用して実装される）ニューラルネットワークシステム511の部分であり得る。いくつかの実装形態では、ワーピングエンジン567および／または加算器エンジン569は、ニューラルネットワークシステム511の外部にある構成要素であり得る。たとえば、そのような実装形態では、ワーピングおよび加算は、非学習演算子であり、その場合、これらの演算は、ニューラルネットワークまたは他の機械学習システムを使用して行われたい。

【0232】

[0155]ニューラルネットワークシステム511は、任意の好適な機械学習トレーニング技法を使用してトレーニングされ得る。1つの例示的な例では、ニューラルネットワークシステム511をトレーニングするために教師付き学習技法が使用され得る。たとえば、ニューラルネットワークのノードの重み（場合によっては、バイアスなどの他のパラメータ）を調整するためにバックプロパゲーショントレーニングプロセスが使用され得る。バックプロパゲーションは、フォワードパスと、損失関数と、バックワードパスと、重み更新とを含むことができる。フォワードパスと、損失関数と、バックワードパスと、パラメータ更新とは、1つのトレーニング反復のために実施される。プロセスは、ニューラルネットワークシステム511のパラメータの重みが正確にチューニングされるまでトレーニングデータの各セットのために一定数の反復の間繰り返され得る。

【0233】

[0156]1つの例示的な例では、ニューラルネットワークシステム511をトレーニングするために使用されるトレーニングデータは、入力として（符号化／圧縮されない）ビデオフレーム x_t と、知られている出力として再構成されたビデオフレーム

【0234】

【数52】

$\hat{x}_t$

【0235】

（入力ビデオフレーム x_t の再構成されたバージョン）と、を含むことができる。フォワードパスは、ニューラルネットワークを通して入力ビデオフレームをパスすることを含むことができる。重みは、最初に、ニューラルネットワークがトレーニングされる前にランダム化され得る。ニューラルネットワークシステムのための第1のトレーニング反復について、重みがまだ較正されていないので、出力は、いかなる特定の出力にも選好を与えない値を含み得る。たとえば、出力は、再構成されたビデオフレーム

【0236】

【数53】

$$\hat{x}_t$$

【0237】

を表す値をもつデータ表現（たとえば、ベクトル、テンソルなど）を含むことができる。
最初の重みを使用した第1のトレーニング反復の後に、再構成されたビデオフレーム

【0238】

【数54】

$$\hat{x}_t$$

10

【0239】

は、おそらく対応する入力フレーム x_t に似ていないものになる。

【0240】

[0157]出力中のエラーを分析するために損失関数を使用され得る。入力として圧縮されていないビデオフレームを使用し、知られている出力として再構成されたビデオフレーム

【0241】

【数55】

$$\hat{x}_t$$

20

【0242】

を使用する例では、ニューラルネットワークシステム511をトレーニングするために図5Aに示されている損失関数

【0243】

【数56】

$$Loss0 = \sum_t distortion(x_t, \hat{x}_t)$$

【0244】

が使用され得る。他のトレーニングデータが使用されるとき、他の損失関数を使用され得る。別の損失関数の一例は、

【0245】

【数57】

$$E_{total} = \sum_2 \frac{1}{2} (target - output)^2$$

30

【0246】

として定義される平均2乗誤差（MSE）を含む。MSEは、実際の応答と予測された（出力）応答との差を2乗したものの和の1/2倍を計算する。

【0247】

[0158]実際の出力値（ネットワークによって出力された再構成されたフレーム）が予測された出力（入力フレームの知られている再構成されたバージョン）とは大きく異なり得るので、損失（またはエラー）は、第1のトレーニングビデオフレームについて高くなり得る。トレーニングの目的は、予測された出力のための損失の量を最小化することである。ニューラルネットワークは、どの入力（重み）がニューラルネットワークの損失に最も寄与したのかを決定することによってバックワードパスを実施することができ、損失が減少し、最終的に最小化されるように重みを調整することができる。ニューラルネットワークの損失に最も寄与した重みを決定するために（Wは、特定のレイヤにおける重みである dL/dW として示される）重みに対する損失の導関数が計算され得る。たとえば、重みは、それらが勾配の反対方向に変化するように更新され得る。重み更新は、

【0248】

40

50

【数 5 8】

$$w = w_i - \eta \frac{dL}{dw}$$

【0 2 4 9】

として示され得、ここで、 w は、重みを示し、 w_i は、最初の重みを示し、 η は、学習速度を示す。学習速度は、任意の好適な値に設定され得、高い学習速度はより大きい重み更新を含み、より低い値はより小さい重み更新を示す。ニューラルネットワークシステム 5 1 1は、所望の出力が達成されるまでそのような方式でトレーニングされ続けることができる。場合によっては、ニューラルネットワークシステム 5 1 0は、同様の方式でトレーニングされ得る。

10

【0 2 5 0】

[0159]図 6は、動き補償ベースのニューラルネットワークシステム（たとえば、ニューラルネットワークシステム 4 1 1および／またはニューラルネットワークシステム 5 1 1）によって実施され得る異なるタイプの動き推定を示す図である。図 6では、項 x は、（現在のフレームが符号化されているときの前のフレームに対応する）動きが推定され得るフレームを示し、項 f は、動き推定を示し、項 y は、動き補償フレームを示し、これは $y = f(x)$ として計算され得る。

【0 2 5 1】

[0160]実施され得る 1つのタイプの動き推定は、ブロックベースの動き推定技法 6 0 2である。ブロックベースの動き推定は、ブロックごとに実施され得る。たとえば、（動き補償されることになる）現在のフレーム y 中のブロックごとに、動き推定 f が前のフレーム x 中の対応するブロックのロケーションを定義する。1つの例示的な例では、動き推定 f は、前のフレーム x 中の対応するブロックに対する現在のフレーム y 中のブロックの変位（たとえば、水平変位および垂直変位）を示す動きベクトルを含むことができる。前のフレーム x からのブロックは、ブロック間の類似性（たとえば、ピクセル値の類似性）を決定することによって現在のフレーム y 中のブロックに対応するように決定され得る。

20

【0 2 5 2】

[0161]実施され得る別のタイプの動き推定は、オプティカルフロー動き推定技法 6 0 4である。オプティカルフロー動き推定は、ピクセルごとに実施され得る。たとえば、現在のフレーム y 中のピクセルごとに、動き推定 f は、前のフレーム x 中の対応するピクセルのロケーションを定義する。ピクセルごとの動き推定 f は、フレーム間のピクセルの移動を示すベクトルを含むことができる。場合によっては、（動きベクトルマップとも呼ばれる）オプティカルフローマップは、フレーム間のオプティカルフローベクトルの計算に基づいて生成され得る。オプティカルフローマップは、フレーム中のピクセルごとにオプティカルフローベクトルを含むことができ、ここで、各ベクトルは、フレーム間のピクセルの移動を示す。1つの例示的な例では、ピクセルのためのオプティカルフローベクトルは、第 1のフレームから第 2のフレームへのピクセルの移動を示す（たとえば、 x 変位および y 変位などの水平変位および垂直変位を示す）変位ベクトルであり得る。動き補償ベースのニューラルネットワークシステムのエンコーダ（たとえば、エンコーダ 5 3 3）は、現在のフレーム y 中のピクセルごとにオプティカルフローベクトルを符号化もしくは圧縮することができるか、またはオプティカルフローマップが使用されるときにオプティカルフローマップを符号化もしくは圧縮することができる。動き補償ベースのニューラルネットワークシステムのデコーダ（たとえば、デコーダ 5 3 8）によって再構成される動き推定情報

30

40

【0 2 5 3】

【数 5 9】

(f)

【0 2 5 4】

50

は、再構成されたオプティカルフローベクトルのグループ（たとえば、フレーム中のピクセルごとに1つ）または再構成されたオプティカルフローマップを含むことができる。

【0255】

【0162】場合によっては、オプティカルフローマップは、フレーム中の一部のピクセルのためのベクトルを含むことができる。たとえば、高密度オプティカルフローは、高密度オプティカルフローマップ中に含まれ得るフレーム中のピクセルごとにオプティカルフローベクトルを生成するために隣接するフレーム間で計算され得る。別の例では、Lucas-Kanadeオプティカルフローは、オプティカルフローマップ中に含まれ得るフレーム中の一部または全部のピクセルのためのオプティカルフローベクトルを生成するために隣接するフレーム間で計算され得る。フレーム間のオプティカルフローを決定するために任意の他の好適なタイプのオプティカルフロー技法またはアルゴリズムが使用され得る。各オプティカルフローマップは、2Dベクトルフィールドを含むことができ、各ベクトルは、第1のフレームから第2のフレームへのポイントの移動を示す変位ベクトルである。

【0256】

【0163】上記のように、オプティカルフローベクトルまたはオプティカルフローマップは、一連のフレームの隣接するフレーム間で（たとえば、隣接するフレーム x_t および x_{t-1} のセットの間で）計算され得る。2つの隣接するフレームは、連続的にキャプチャされたフレームである2つの直接隣接するフレームまたは一連のフレームの一定の距離だけ離れている（たとえば、互いの2フレーム内にあるか、互いの3フレーム内にあるか、もしくは他の好適な距離にある）2つのフレームを含むことができる。フレーム x_{t-1} からフレーム x_t へのオプティカルフローは、 $O_{x_{t-1}}$, $x_t = \text{dof}(x_{t-1}, x_t)$ によって与えられ得、ここで、 dof は、高密度オプティカルフローである。オプティカルフローマップを生成するために任意の好適なオプティカルフロープロセスが使用され得る。1つの例示的な例では、フレーム x_{t-1} 中のピクセル $I(x, y, t)$ は、次のフレーム x_t 中で距離 $(\Delta x, \Delta y)$ だけ移動することができる。ピクセルが同じであり、強度がフレーム x_{t-1} と次のフレーム x_t との間で変化しないと仮定すると、以下の式が仮定され得る。

【0257】

【数60】

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \quad \text{式(2)}$$

【0258】

【0164】上記の式(2)の右辺のテイラー級数近似を取り、次いで、共通項を除去し、 Δt で除算することによって、オプティカルフロー式が次のように導出され得る。

【0259】

【数61】

$$f_x u + f_y v + f_t = 0, \quad \text{式(3)}$$

【0260】

【0165】ここで、

【0261】

【数62】

10

20

30

40

50

$$f_x = \frac{df}{dx},$$

$$f_y = \frac{df}{dy},$$

$$f_t = \frac{df}{dt},$$

$$u = \frac{\Delta x}{\Delta t}, \text{ および}$$

$$v = \frac{\Delta y}{\Delta t}.$$

10

【0262】

【0166】オプティカルフロー式（3）を使用して、画像勾配 f_x および f_y は、（ f_t とし
て示される）時間に沿った勾配とともに発見され得る。項 u および v は、 $I(x, y, t)$
の速度またはオプティカルフローの x および y 成分であり、未知である。推定技法は、
場合によっては、オプティカルフロー式が2つの未知の変数で解かれないことがあるとき
に必要とされ得る。オプティカルフローを推定するために任意の好適な推定技法が使用さ
れ得る。そのような推定技法の例は、差分方法（たとえば、Lucas-Kanade 推
定、Horn-Schunck 推定、Buxton-Buxton 推定、もしくは他の好
適な差分方法）、位相相関、ブロックベースの方法、または他の好適な推定技法を含む。
たとえば、Lucas-Kanade は、ピクセル I の局所近傍においてオプティカルフ
ロー（画像ピクセルの変位）が小さく、ほぼ一定であると仮定し、最小2乗法を使用して
その近傍中のすべてのピクセルのための基本オプティカルフロー式を解く。

20

【0263】

【0167】実施され得る別のタイプの動き推定は、（適応型畳み込みとも呼ばれる）動的畳
み込み動き推定技法606である。動的畳み込み動き推定は、ピクセルごとに実施され得
る。たとえば、現在のフレーム y 中のピクセルごとに、動き推定 f は、前のフレーム x 中
のピクセルの近傍の線形結合を定義する。図6に示されているように、現在のフレーム y
中のピクセルロケーションにおけるピクセルについて、前のフレーム x 中で同じピクセル
ロケーションにおけるピクセルを囲むピクセルの近傍の周りにカーネルまたは行列が定義
される。図6の例示的な動的畳み込み動き推定は、現在のフレーム y 中の（（行番号、列
番号）に対応する）ロケーション（2, 7）におけるピクセルのために実施され、ここで
、現在のフレーム y 中の左上のピクセルロケーションは、ロケーション（0, 0）にある
。前のフレーム x 中のロケーション（2, 7）におけるピクセルの周りにカーネルが定義
される。たとえば、図示のように、サイズ 5×5 のカーネルが、前のフレーム x 中のロケ
ーション（2, 7）におけるピクセルを囲むピクセルの近傍に適用される。

30

【0264】

【0168】カーネルまたは行列中の各エントリは、数（図6の例では 5×5 の数のセット）
を含むことができる。カーネル中で数のセットは、新たに再構成されたフレーム（現在の
フレーム y ）中に再構成されているピクセルの同じピクセルロケーションに対応する前の
フレーム x 中のピクセルロケーションにおけるピクセルを囲むピクセルの近傍からのピク
セルをどのように組み合わせるのかについて説明することができる。動き補償ベースのニ
ューラルネットワークシステムのエンコーダ（たとえば、エンコーダ533）は、現在の
フレーム y 中のピクセルごとにカーネルからの数（たとえば、図6からの 5×5 のカー
ネルからの25の数）の表現を圧縮することができる。動き補償ベースのニューラルネッ
トワークシステムのデコーダ（たとえば、デコーダ538）によって再構成される動き推定
情報

40

【0265】

【数63】

50

【0266】

は、前に再構成されたフレームをワーピングするためにデータ畳み込みを実施するための使用され得るピクセルごとの再構成されたカーネルを含むことができる。

【0267】

[0169]図7は、ポストワーピング拡張を実施するための追加のポストワーピングネットワーク574をもつニューラルネットワークシステム511の一例を示す図である。場合によっては、ワーピングエンジン567がワーピングを実施し、加算器エンジン569が残差を追加した後、再構成されたフレームを改善する余地が依然としてあり得る。たとえば、再構成されたフレーム

【0268】

【数64】

$$\hat{x}_t$$

【0269】

は、（ワーピングエンジン567による）ワーピングと（加算器エンジン569による）残差の追加との後に、元のフレーム

【0270】

【数65】

$$\hat{x}$$

【0271】

中に存在しなかった視覚的アーティファクトを有し得る。ポストワーピングネットワーク574は、（たとえば、アーティファクトを除去することによって）再構成されたフレームをさらに向上または改良するために使用されるニューラルネットワークの構成要素（たとえば、1つまたは複数の畳み込みレイヤ）を含むことができる。

【0272】

[0170]ポストワーピングネットワーク574は、ワーピングエンジン567が前に再構成されたフレームをワーピングした後に適用され得る。たとえば、図7に示されているように、時間ステップtにおいて、（ワーピングが適用された後の）ワーピングしたフレームと再構成された残差

【0273】

【数66】

$$\hat{r}_t$$

【0274】

とがポストワーピングネットワーク574に入力される。ポストワーピングネットワークは、あらゆる視覚的不純物（たとえば、視覚的アーティファクト）が除去された再構成されたフレーム

【0275】

【数67】

$$\hat{x}_t$$

【0276】

を生成することができる。いくつかの例では、ポストワーピングネットワーク574は、残りのニューラルネットワークシステム511のニューラルネットワークとは無関係にトレーニングされる。いくつかの例では、ポストワーピングネットワーク574は、残りのニューラルネットワークシステム511とは無関係にトレーニングされず、その場合、ニューラルネットワークシステム511全体は、エンドツーエンドでトレーニングされる。たとえば、クリーンなまたは圧縮されていないフレームが、ニューラルネットワークシステム511のニューラルネットワークに入力され得、ネットワークは、再構成されたフレ

10

20

30

40

50

ームを出力する。(入力フレームと比較して)再構成されたフレームに関連するエラーまたは損失を査定するためにバックプロパゲーションが使用され得、ニューラルネットワークシステム511のパラメータは、上記で説明されたように、チューニングされ得る。

【0277】

[0171]図8および図9は、追加の動き推定ネットワーク575をもつニューラルネットワークシステム511の一例を示す図である。動き推定ネットワーク575は、動き推定の拡張学習を与えることができる。図8および図9に示されているように、(動き推定ネットワーク575に入力された)前の入力フレーム x_t と、動き推定ネットワーク575によってエンコーダ側において決定された動き推定

【0278】

【数68】

$$f_{t+1}^{enc}$$

【0279】

と、(ワーピングした再構成されたフレーム

【0280】

【数69】

$$f_{t+1}^{enc}(\hat{x}_t)$$

【0281】

として示される)動き推定

【0282】

【数70】

$$f_{t+1}^{enc}$$

【0283】

によってワーピングされた前の再構成されたフレームとを含む追加の入力変数が、時間ステップ $t+1$ においてニューラルネットワークシステム511のエンコーダ側に与えられる。

【0284】

[0172]上記のように、いくつかの実装形態では、ニューラルネットワークシステム511は、エンドツーエンド方式でトレーニングされる。場合によっては、ニューラルネットワークシステム511は、いくつかのパラメータを計算しようと試みる時にトレーニング中に問題を有する可能性がある。たとえば、ニューラルネットワークが、動き推定パラメータ(たとえば、オプティカルフロー)を決定し、それらの動き推定パラメータを使用して前に再構成されたフレームをワーピングしようと試みる時に問題が起こる可能性がある。別の例では、残差を推定し、残差にワーピングしたフレームを加算する時に、問題が起こる可能性がある。そのような問題は、再構成がどのように取得されるのかにかかわらずニューラルネットワークシステム511が可能な最良の再構成を取得しようと試みることにより起こる可能性がある。たとえば、できるだけ元のフレームに近い再構成されたフレームを生成しようと試みる際に、ニューラルネットワークシステム511は、動き推定パラメータ(たとえば、オプティカルフロー)を適切に学習しないことがあり、再構成されたフレームを生成するために残差だけに依拠することがある。場合によっては、再構成されたオプティカルフローの出力がブランクテンソルまたはデータの無い他のデータ構造であることすらあり得る。しかしながら、ビデオ中の冗長の大部分は、そのような動きデータから(たとえば、フレーム間のオプティカルフローから)来ることがあり、その場合、動き推定パラメータを学習することが重要になり得る。

【0285】

[0173]動き推定ネットワーク575は、ニューラルネットワークシステム511のエン

10

20

30

40

50

コーダ側で動き推定ニューラルネットワークを与えることによってニューラルネットワークシステム511が動き推定パラメータを学習するのを助けることができる。たとえば、動き推定ネットワーク575は、動き推定パラメータ（たとえば、オプティカルフローベクトル、オプティカルフロマップ、ブロックベースの動き情報、データ畳み込みを実施するために使用されるピクセルごとのカーネルなど）を推定し、エンコーダ533にそれらの動き推定パラメータを与える別個のニューラルネットワークを含むことができる。ニューラルネットワークシステム511のエンコーダ側の動き推定パラメータを決定することによって、動き推定パラメータの実際のセットがエンコーダ533に与えられ、エンコーダ533に動き推定（たとえば、オプティカルフロー）を検討させる。

【0286】

[0174]いくつかの例では、動き推定ネットワーク575は、ニューラルネットワークシステム511の残りとは無関係にトレーニングされ得るか、または（動き推定ネットワーク575を含む）ニューラルネットワークシステム511全体は、エンドツーエンド方式と一緒にトレーニングされ得る。動き推定ネットワーク575を考慮に入れる様々な損失項は、以下を含むことができる。

【0287】

【数71】

$$\text{Loss1} = \sum_t \text{distortion}(f_t^{\text{enc}}(x_{t-1}), x_t)$$

$$\text{Loss2} = \sum_t \text{distortion}(f_t^{\text{enc}}(\hat{x}_{t-1}), x_t)$$

$$\text{Loss3} = \sum_t \text{distortion}(\hat{f}_t(x_{t-1}), x_t)$$

$$\text{Loss4} = \sum_t \text{distortion}(\hat{f}_t(\hat{x}_{t-1}), x_t)$$

【0288】

[0175]いくつかの実装形態では、動き推定ネットワーク575は、（グラウンドトゥールース出力を使用せずに）教師なし学習を使用してトレーニングされ得る。Loss1およびLoss2項は、エンコーダ側で使用され得、Loss3およびLoss4項は、デコーダ側で使用され得る。動き推定の一例としてオプティカルフローを使用する1つの例示的な例では、2つの連続するフレーム x_t および x_{t+1} を仮定すれば、動き推定ネットワーク575の目的は、2つのフレーム x_t および x_{t+1} の間のオプティカルフローを推定することである。動き推定ネットワーク575は、推定されたオプティカルフローパラメータを出力することになり、これは、 x_t の再構成されたバージョン

【0289】

【数72】

$$\hat{x}_t$$

【0290】

をワーピングするために使用されることになる。損失（たとえば、上記のLoss2）は、ワーピングした

【0291】

【数73】

$$\hat{x}_t$$

【0292】

と元のフレーム x_t との間で決定され得る。損失は、次いで、上記で説明されたバックプロパゲーション技法を使用して重み（および、場合によっては、他のパラメータ）をチューニングするために使用され得る。場合によっては、Loss2ではなく、損失項Loss

10

20

30

40

50

1 が、エンコーダ側で使用され得る。

【0293】

【0176】エンコーダ側で推定された動き推定パラメータ（たとえば、オプティカルフロー）を復元するためにデコーダ538をトレーニングするために、同じ概念がLoss3またはLoss4を使用してデコーダ側で使用され得る。たとえば、デコーダ538は、現在のフレーム x_t と前のフレーム x_{t-1} （Loss3が使用されるとき）または前に再構成されたフレーム

【0294】

【数74】

$$\hat{x}_{t-1}$$

10

【0295】

（Loss4が使用されるとき）のいずれかとを含む2つの連続するフレーム間の動き（たとえば、オプティカルフロー）を計算または推定することができる。このようにして、動き推定パラメータを推定するために圧縮されていないフレームが使用され得る。得られた動き推定パラメータは、フレーム x_t に一致しようとする試みにおいて前のフレーム（ x_{t-1} または

【0296】

【数75】

$$\hat{x}_{t-1}$$

20

【0297】

）をワーピングするために使用され得る。ワーピングの出力が現在の時間ステップにおける入力フレーム x_t に一致することが望まれる。

【0298】

【0177】ワーピングしたフレームと入力フレーム x_t との間の損失は、次いで、Loss3またはLoss4のいずれかを使用して計算され得る。上記のように、Loss3は、実際の前のフレーム x_{t-1} の使用を含み、Loss4は、前に再構成されたフレーム

【0299】

【数76】

$$\hat{x}_{t-1}$$

30

【0300】

（前のフレーム x_{t-1} の解凍または復号されたバージョン）の使用を含む。時間ステップ $t-1$ からの実際の前のフレーム x_{t-1} が次の時間ステップ t においてニューラルネットワークシステム511にとって利用可能でない状況においてデコーダ側ではLoss4（エンコーダ側ではLoss2）が使用され得る。損失を査定し、ニューラルネットワークシステム511のパラメータを調整またはチューニングするために、バックプロパゲーションが使用され得る。そのようなトレーニングは、デコーダ538が動き推定パラメータを学習するのを支援する。

【0301】

【0178】いくつかの例では、（実際の前のフレーム x_{t-1} が時間ステップ t において利用可能であるのかどうかに応じて）損失項Loss1およびLoss3または損失項Loss2およびLoss4がトレーニングの異なる段階において活性化される。たとえば、エンコーダ側でLoss2とデコーダ側でLoss4とが活性化され、ニューラルネットワークシステム511のトレーニングが最初に始まり、ある時点まで使用され得、その後、ニューラルネットワークシステム511は、（図5Aおよび上記に示されているように）唯一の損失項としてLoss0を使用してトレーニングされる（損失項Loss1～Loss4は非活性化される）。トレーニング中に損失項Loss1およびLoss3またはLoss2およびLoss4を最初に使用することによって、ニューラルネットワークシ

40

50

システム511は、動き推定パラメータ（たとえば、オプティカルフローベクトルまたはマップ）を学習させられる。ニューラルネットワークシステム511は、動き推定パラメータを十分に学習すると、その情報を保持することができ、さらなるトレーニング反復におよび推測中に（トレーニングされたネットワークが実際のビデオデータ上での動作のために展開されるときに）動き推定パラメータをどのように推定するのかを維持することになる。したがって、追加の損失項 $Loss1 \sim Loss4$ は、動き推定パラメータがニューラルネットワークシステム511によって学習されると非活性化され得、ニューラルネットワークシステム511をさらにトレーニングするために $Loss0$ を使用し続けることができる。

【0302】

[0179]図10は、単一レベルの再帰構成とマルチレベルの再帰構成とを含むニューラルネットワークシステム511の異なる再帰レイヤ構成の一例を示す図である。エンコーダ533とデコーダ538とは、ボトルネック1034によって分離された異なるレイヤとして示されている。ボトルネック1034は、量子化器534を使用して量子化されたラテント表現に対応する。単一レベルの再帰構成では、エンコーダ533は、順方向レイヤを含み、デコーダは、4つの順方向レイヤと単一の再帰レイヤ1071とを含む。エンコーダ533の順方向レイヤは、1つまたは複数の畳み込みレイヤ、正規化レイヤ、非線形レイヤ、残差ブロック、または他のニューラルネットワークレイヤなど、図5Bに示されているニューラルネットワークレイヤのうちの1つまたは複数を含むことができる。順方向レイヤ1070aと、順方向レイヤ1070bと、順方向レイヤ1070cとは、図5Aに示されている1つまたは複数の再帰前デコーダレイヤ570に対応する。1つの例示的な例では、順方向レイヤ1070a、1070b、および1070cは、(ConvGRU $c:C_3$ $k:3$ $s:1$ として図5B中で示されている)再帰レイヤより前に発生する図5Bのデコーダ538の(レイヤ579として図5Bに示されている)すべてのレイヤを含むことができる。図10の再帰レイヤ1071は、図5Aに示されている1つまたは複数の再帰レイヤ571に対応する。再帰レイヤ1071の一例は、図5Bからの(設定 $c:C_3$ $k:3$ $s:1$ をもつ)ConvGRU再帰レイヤである。図示のように、時間ステップ t からの再帰レイヤ1071の再帰状態は、再帰レイヤ1071に、および次の時間ステップ $t+1$ においてエンコーダ533に与えられる。順方向レイヤ1073は、図5Aに示されている1つまたは複数の再帰後レイヤ572に対応する。1つの例示的な例では、順方向レイヤ1073は、図5Bのデコーダ538の(Conv $c:5$ $k:5$ $s:1$ として示されている)畳み込みレイヤを持続することができる。

【0303】

[0180]マルチレベルの再帰構成は、単一の再帰構成と同じ順方向レイヤを含むが、デコーダ538に複数のレベルの再帰レイヤを含む。異なる再帰レイヤの各々は、(隠れ状態とも呼ばれる)別個の再帰状態を維持し、各再帰状態は、ニューラルネットワークシステム511中のその時点までの異なるレベルの特徴をキャプチャする。図10に示されているように、デコーダ538の再帰レイヤにフィードバックされることに加えて、再帰状態は、デコーダ538の再帰レイヤからエンコーダ533中のミラーロケーションにフィードバックされる。複数の再帰状態により、エンコーダ533とデコーダ538とは、より効率的なビデオコーディングを実施するために追加情報を利用することが可能になり得る。

【0304】

[0181]図11は、追加のワーピングエンジン576をもつニューラルネットワークシステム511の一例を示す図である。ニューラルネットワークシステム511の様々な構成要素は、説明のために図11から省略されている(たとえば、量子化器534、1つまたは複数の再帰前デコーダレイヤ570など)が、それらの構成要素がニューラルネットワークシステム511中に存在することを当業者ならば理解されよう。ワーピングエンジン576は、現在の時間ステップ(たとえば、時間ステップ $t+1$)において再帰状態を使用する前に1つまたは複数の再帰レイヤ571の前の時間ステップ(たとえば、時間ステップ t)からの前の再帰状態(または隠れ状態)をワーピングするために使用され得る。

ワーピングエンジン 576 は、現在のフレーム中のピクセルのロケーションに再帰状態を一致させるために再帰状態をワーピングするためにエンコーダ 533 とデコーダ 538 とによって利用可能な動き推定パラメータを使用することができる。

【0305】

[0182] 現在の時間ステップにおいてそれらを使用する前の前の再帰状態のワーピングは、ビデオのために有用であり得る。たとえば、オブジェクトは、ビデオフレームにわたって時間的に移動することができ、そのような動きは、（たとえば、オプティカルフロー、ブロックベースの動き推定などを使用して）上記で説明された動き推定技法によってモデル化され得る。デコーダ 538 の 1 つまたは複数の再帰レイヤ 571 の再帰状態が後の時間ステップにおいてデコーダ 538 とエンコーダ 533 とに直接フィードバックされるとき、オブジェクトの動きは、（たとえば、図 5A に示されているように、動き補償が再帰レイヤの後に実施されるので）隠れ状態内でモデル化されない。再帰状態は、入力フレームを表し（たとえば、それは、入力画像の特徴表現を含み）、その場合、入力画像中のピクセルの移動は、再帰状態の特徴空間中の同様の移動を生じなければならない。したがって、動き推定に基づいて再帰状態をワーピングすることによって、入力フレームから推定される各移動は、再帰状態中の対応する移動に移され、再帰状態を現在のフレーム中のピクセルのロケーションに一致させる。

【0306】

[0183] 図 11 を参照すると、現在の時間ステップにおけるワーピングエンジン 576 は、前の時間ステップからのデコーダ 538 の再帰レイヤの前の再帰または隠れ状態 h をワーピングするために使用される。前の隠れ状態は、現在の時間ステップにおいて再帰レイヤ中で再使用される前に現在の時間ステップにおいて推定される動き推定パラメータ（たとえば、オプティカルフロー）を使用してワーピングされる。たとえば、ワーピングエンジン 576 は、時間ステップ $t+1$ において、時間ステップ $t+1$ において推定される動き推定パラメータ

【0307】

【数 77】

$$\hat{f}_{t+1}$$

【0308】

を使用して時間ステップ t からの前の再帰状態 h_t をワーピングすることができる。ワーピングした再帰状態は、次いで、現在の時間ステップ t において再帰レイヤ中で再使用され得る。

【0309】

[0184] いくつかの実装形態では、追加の再帰レイヤが、ワーピングエンジン 576 の前に与えられ得る。追加の再帰レイヤは、現在の時間ステップ $t+1$ の動作のための動き推定パラメータ

【0310】

【数 78】

$$\hat{f}_{t+1}$$

【0311】

（たとえば、オプティカルフロー）を再構成するために使用され得、これは、時間ステップ t からの前の再帰状態 h_t をワーピングするためにワーピングエンジン 576 によって使用され得る。ワーピングした前の再帰状態

【0312】

【数 79】

$$\hat{f}_t(\hat{h}_t)$$

【0313】

は、次いで、現在の時間ステップ t において再帰レイヤに入力され得る。いくつかの実装形態では、反復動作は、最初に、現在の時間ステップ $t + 1$ における動き推定パラメータ（たとえば、オプティカルフロー）を決定するために実施され得、これは、時間ステップ t からの前の再帰状態 h_t をワーピングするためにワーピングエンジン 576 によって使用され得る。ワーピングした前の再帰状態

【0314】

【数80】

$$\hat{f}_t(\hat{h}_t)$$

【0315】

は、次いで、現在の時間ステップ t において再帰レイヤに入力され得る。たとえば、フォワードパスは、動き推定パラメータを復元または再構成するために実施され得、次いで、動作は、前の再帰状態をワーピングするために再帰レイヤに戻ることができる。いくつかの実装形態では、現在の時間ステップ (t) の再帰レイヤへのフィードバックとしての（時間ステップ $t - 1$ からの）ワーピングした前に再構成されたフレーム。

【0316】

【0185】図12は、追加のマスキング機能をもつニューラルネットワークシステム511の一例を示す図である。図12に示されているように、1つまたは複数の再帰後レイヤ573は、再構成された動き推定データ

【0317】

【数81】

$$\hat{f}_t$$

【0318】

および再構成された残差データ

【0319】

【数82】

$$\hat{f}_t$$

【0320】

に加えて要素ごとのマスク

【0321】

【数83】

$$\hat{m}_t \in [0, 1]$$

【0322】

を出力することができる。ワーピングエンジン567は、図5Aに関して上記で説明されたことと同様に、前に再構成されたフレーム

【0323】

【数84】

$$\hat{x}_{t-1}$$

【0324】

をワーピングすることによってワーピングした前の再構成されたフレーム

【0325】

【数85】

$$\hat{f}_t(\hat{x}_{t-1})$$

【0326】

を生成することができる。マスク

10

20

30

40

50

【0327】

【数86】

$$\hat{m}_t$$

【0328】

は、ビデオフレーム中でピクセルごとに値を有するピクセルごとのマスクであり得る。マスク

【0329】

【数87】

$$\hat{m}_t$$

10

【0330】

は、再構成されたフレーム

【0331】

【数88】

$$\hat{x}_t$$

【0332】

を再構成するときに、ワーピングした前の再構成されたフレーム

【0333】

【数89】

$$\hat{f}_t(\hat{x}_{t-1})$$

20

【0334】

と残差

【0335】

【数90】

$$\hat{r}_t$$

【0336】

との寄与をマスクングする（または重み付けする）ために使用される。たとえば、いくつかの実装形態では、マスク

【0337】

【数91】

$$\hat{m}_t$$

30

【0338】

は、以下の式に従って適用され得る。

【0339】

【数92】

$$\hat{x}_t = \hat{m}_t \cdot \hat{f}_t(x_{t-1}) + (1 - \hat{m}_t) \cdot \hat{r}_t \quad \text{式(4)}$$

40

【0340】

[0186]式(4)によれば、マスク

【0341】

【数93】

$$\hat{m}_t$$

【0342】

の値が増加するにつれて、再構成されたフレーム

50

【 0 3 4 3 】

【 数 9 4 】

$$\hat{x}_t$$

【 0 3 4 4 】

上のワーピングした前の再構成されたフレーム

【 0 3 4 5 】

【 数 9 5 】

$$\hat{f}_t(\hat{x}_{t-1})$$

10

【 0 3 4 6 】

の効果が増加し、再構成されたフレーム

【 0 3 4 7 】

【 数 9 6 】

$$\hat{x}_t$$

【 0 3 4 8 】

上の残差データ

【 0 3 4 9 】

【 数 9 7 】

$$\hat{r}_t$$

20

【 0 3 5 0 】

の効果が低減される。一方、マスク

【 0 3 5 1 】

【 数 9 8 】

$$\hat{m}_t$$

【 0 3 5 2 】

の値が減少するにつれて、ワーピングした前の再構成されたフレーム

【 0 3 5 3 】

【 数 9 9 】

$$\hat{f}_t(\hat{x}_{t-1})$$

30

【 0 3 5 4 】

の効果は低減され、残差データ

【 0 3 5 5 】

【 数 1 0 0 】

$$\hat{r}_t$$

40

【 0 3 5 6 】

の効果は増加する。いくつかの実装形態では、

【 0 3 5 7 】

【 数 1 0 1 】

$$\hat{m}_t$$

【 0 3 5 8 】

は、0と1との間の値の範囲に制限され得る（両端値を含む [0, 1] ）。1つの例示的な例では、マスクがピクセルのために0の値を有する場合、そのピクセルのための次のフ

50

レーム中の値は、完全に残差から取られることになる。

【0359】

[0187]いくつかの例では、マスク

【0360】

【数102】

$$\hat{m}_t$$

【0361】

は、前のビデオフレーム中に存在しなかった現在のビデオフレームの部分をマスキングして除くオクルージョンマスクであり得る。そのようなシナリオは、たとえば、シーンの部分が一時的に遮断される時、オブジェクトがシーンに入るときなどに発生することができる。フレームのそのような利用不可能な部分のマスキングは、それらの部分のための動き情報（したがって、ワーピング）が利用不可能であるので実施され得る。たとえば、ピクセルが前のフレーム中で利用不可能である場合、現在のフレームと前のフレームとの間の動き推定は現在のフレームのピクセルのために実施されないことがある。そのような例では、それらのピクセルを含むビデオフレームの部分は、ワーピングされないことがある。そのような場合、現在のフレームの再構成は、オクルージョンの量に応じて、残差データ

【0362】

【数103】

$$\hat{r}_t$$

【0363】

により重度に（または場合によっては完全に）依拠することによって決定され得る。

【0364】

[0188]図13は、追加の再構成改良機能をもつニューラルネットワークシステム511の一例を示す図である。図13に示されている構成を使用して、ニューラルネットワークシステム511は、ニューラルネットワークシステム511の2つの時間ステップの動作において各ビデオフレームを再構成することができる。たとえば、現在の時間ステップ（たとえば、時間ステップ $t-1$ ）において決定された各再構成されたビデオフレーム（たとえば、

【0365】

【数104】

$$\hat{x}_{t-1}$$

【0366】

）は、後続の時間ステップ（たとえば、時間ステップ t ）においてさらに精製され得る。そのような場合、前に再構成されたビデオフレーム

【0367】

【数105】

$$\hat{x}$$

【0368】

は、2つの役割をサービスすることができる。たとえば、現在の時間ステップ t において、（図5Aに示されている構成においても行われているように）現在の時間ステップの再構成

【0369】

【数106】

$$\hat{x}_t$$

【0370】

を形成するために、前に再構成されたビデオフレーム

【 0 3 7 1】

【数 1 0 7】

$$\hat{x}_{t-1}$$

【 0 3 7 2】

がワーピングされ、残差

【 0 3 7 3】

【数 1 0 8】

$$\hat{r}_t$$

10

【 0 3 7 4】

が加算される。さらに、残差

【 0 3 7 5】

【数 1 0 9】

$$\hat{r}_{t-1}$$

【 0 3 7 6】

も出力され、前に再構成されたビデオフレーム

20

【 0 3 7 7】

【数 1 1 0】

$$\hat{x}_{t-1}$$

【 0 3 7 8】

に加算され、前に再構成されたビデオフレーム

【 0 3 7 9】

【数 1 1 1】

$$\hat{x}_{t-1}$$

30

【 0 3 8 0】

と比較して改善された品質を有する改良された再構成されたフレーム

【 0 3 8 1】

【数 1 1 2】

$$\hat{x}_{t-1}$$

【 0 3 8 2】

を生じる。

【 0 3 8 3】

[0189]図 1 3 に示されているように、各時間ステップにおいて、前のグラウンドトゥルースの（圧縮されていない）フレーム x_t が、前の再構成されたフレーム

40

【 0 3 8 4】

【数 1 1 3】

$$\hat{x}$$

【 0 3 8 5】

の改良を行うことを可能にするために、エンコーダ 5 3 3 への入力として加算される。たとえば、時間ステップ t において、前のグラウンドトゥルースのフレーム x_{t-1} が、前の再構成されたフレーム

【 0 3 8 6】

50

【数 1 1 4】

$$\hat{x}_{t-1}$$

【0 3 8 7】

の改良を行うことを可能にするためにエンコーダ 5 3 3 への入力として加算される。図示のように、時間ステップ t において、1 つまたは複数の再帰後レイヤ 5 7 3 は、残差

【0 3 8 8】

【数 1 1 5】

$$\hat{r}_t$$

10

【0 3 8 9】

に加えて、出力残差

【0 3 9 0】

【数 1 1 6】

$$\hat{r}'_{t-1}$$

【0 3 9 1】

を与える。残差

【0 3 9 2】

【数 1 1 7】

$$\hat{r}'_{t-1}$$

20

【0 3 9 3】

は、前のグラウンドトゥルースのフレーム x_{t-1} に基づいて生成され、改良された再構成されたフレーム

【0 3 9 4】

【数 1 1 8】

$$\hat{x}'_{t-1}$$

【0 3 9 5】

を生成するために、前の再構成

【0 3 9 6】

【数 1 1 9】

$$\hat{x}_{t-1}$$

30

【0 3 9 7】

を改良するために使用される。たとえば、改良された再構成は、次のように生成され得る。

【0 3 9 8】

【数 1 2 0】

$$\hat{x}'_{t-1} = \hat{x}_{t-1} + \hat{r}'_{t-1} \quad \text{式(5)}$$

40

【0 3 9 9】

[0190] 改良された再構成されたフレーム

【0 4 0 0】

【数 1 2 1】

$$\hat{x}'_{t-1}$$

【0 4 0 1】

は、時間ステップ $t-1$ における前のフレームのための最後の再構成として使用され得る。そのような場合、システム 5 0 0 は、ニューラルネットワークシステム 5 1 1 の1つの

50

時間ステップの動作によって出力ビデオシーケンスを遅延させる。ニューラルネットワークシステム 511 が、第 1 の最初の改良されていない再構成において実施されたエラーを知ることによって次の時間ステップにおいて再構成されたフレームを改良することが可能になることはコーディング効率（たとえば、レートひずみ）の点で有益であり得る。そのような構成は、前の時間ステップにおいて発生したあらゆる誤りを次の時間ステップにおいて補償することを可能にし、これは、レートひずみ最適化に利益を与えることができる。

【0402】

[0191] 図 14 は、ニューラルネットワークシステム 511 のエンコーダ 533 への入力として 1 つまたは複数のひずみ (distortion) マップを使用することの一例を示す図である。ひずみマップ 577 とひずみマップ 578 とを含む 2 つの例示的なひずみマップが図 14 に示されている。ひずみマップ 577 は、現在の時間ステップ t からの現在のソースフレーム x_t と前の時間ステップ $t-1$ からの前の再構成

【0403】

【数 122】

$$\hat{x}_{t-1}$$

【0404】

との間のひずみを含む。ひずみマップ 578 は、現在のソースフレーム x_t とワーピングした前に再構成されたビデオフレーム

【0405】

【数 123】

$$f_t^{enc}(\hat{x}_{t-1})$$

【0406】

との間のひずみを含む。上記のように、前に再構成されたビデオフレーム

【0407】

【数 124】

$$(\hat{x}_{t-1})$$

【0408】

は、オプティカルフローなどの動き推定に従ってワーピングされ得る。そのような場合、オプティカルフローは、フローネットを使用するなどしてエンコーダ側で推定され得る。フローネットは、畳み込みネットワークを用いてオプティカルフローを学習するニューラルネットワークアーキテクチャである。

【0409】

[0192] ひずみマップ中に含まれるひずみは、ニューラルネットワークシステム 511 をトレーニングするための損失関数（たとえば、上記で説明された L_{oss0} 、 L_{oss1} 、 L_{oss2} 、 L_{oss3} 、および／または L_{oss4} ）として使用されるひずみ（またはエラーもしくは損失）であり得る。ひずみは、特に、マルチスケール構造的類似性指数 (MS-SSIM)、ピクセルごとの平均 2 乗誤差 (MSE) などの任意のタイプのひずみに基づき得る。ひずみマップは、知覚ひずみマップと呼ばれることもある。

【0410】

[0193] (ニューラルネットワークシステム 511 の外の) ひずみマップを明示的に計算し、ひずみマップをエンコーダ 533 への入力として供給することによって、エンコーダ 533 は、画像のどの部分が知覚ひずみに寄与しているのか、したがって、どこにより多くの情報を割り振る必要があるのかに関して案内され得る。たとえば、エンコーダ 533 は、ひずみに最も寄与する画像の領域上のビットレートに焦点を当てることができる。そのような利益は、ニューラルネットワークによって学習するために重要である知覚ひずみマップ（たとえば、MS-SSIM）などのいくつかのひずみマップのための拡張であり得る。

10

20

30

40

50

【0411】

[0194]図15は、エンコーダ533中での乗法的交互作用を可能にする追加の線形ニューラルネットワークレイヤを用いるニューラルネットワークシステム511の一例を示す図である。デフォルトの方法で適用される畳み込みニューラルネットワークアーキテクチャは、乗法的交互作用を可能にしない。乗法的交互作用は、入力に対して実施されている計算が入力自体に依存するときである。たとえば、従来の畳み込みニューラルネットワークは、トレーニング中に重みを学習する。重みは、同じであり、ニューラルネットワークへの入力である入力値にかかわらずあらゆる入力に適用される。

【0412】

[0195]図15を参照すると、乗法的交互作用は、入力の各空間位置において独立して（図15の線形レイヤ（1）に示されている）エンコーダ533の線形レイヤ581を適用することによってエンコーダ533中で使用可能になる。線形レイヤ581の重みは、（図15の線形レイヤ（2）に示されている）線形レイヤ582によって出力された重み583によって決定される。たとえば、重み583は、線形レイヤ581のための重みとして使用され得る。同じ入力580が、線形レイヤ581と線形レイヤ582とに適用される。線形レイヤ581の重みを定義するために線形レイヤ582を使用することは、線形レイヤ581によって実施される計算を入力580自体の値に依存することを可能にする（したがって、入力580に適用される計算は入力自体に依存する）。たとえば、ニューラルネットワークシステム511は、前の再構成されたフレームと現在のフレームとの間のひずみが有意である領域と比較してそのひずみが無視できる領域において異なる計算を実施するように学習することができる。

【0413】

[0196]いくつかの実装形態では、図15の構成は、図16に示されている構成と組み合わせ使用され得る。たとえば、1つまたは複数のひずみマップは、入力580に加えて線形レイヤ582に投入され得、これは、重み583の計算を向上させるのを助けることができる。一例では、異なる空間的ロケーションにおける重みは、ひずみマップによって示されるように、異なる空間的ロケーションにおけるひずみの値に従って別様に設定され得る。

【0414】

[0197]再帰ベースのニューラルネットワークシステム（たとえば、システム400および／またはシステム500）ならびにビデオ圧縮について本明細書で説明される関連する技法は、高品質のピクセル当たりのビット（b p p）値の条件で、最先端のレベルで実施することができるビデオコーディングシステムを与える。たとえば、図16および図17に与えられているグラフに示されているように、そのような技法は、（実際的な関心のb p p領域である）1080pの解像度の場合に >0.14 b p pの範囲でHEVCおよびAVCを含む現在の業界標準のコーデックよりも優れている（たとえば、Netflixは、約0.09～0.12 b p pを使用して1080pのビデオをストリームする）。図16および図17における「Recurrent AE」は、本明細書で説明される再帰ベースのニューラルネットワークシステムを使用した結果を指し、図16および図17における「RDAE」は、別の機械学習ベースのビデオ圧縮技法を指す。

【0415】

[0198]本明細書で説明される再帰ベースのニューラルネットワークシステム（たとえば、システム400および／またはシステム500）は、低レイテンシの動作を与える。たとえば、システムは、いくつかのフレームのバッチで（たとえば、ピクチャグループ（G o P）で）ではなくフレームごとのビデオの生成および送信を可能にし、したがって、送信されるフレームの累算を待つ必要がない。さらに、システムは、ビデオの特性に応じてランタイムにG o Pサイズ（すなわち、残差情報だけではなくフルフレームを送信する頻度）を調整する能力を与え、したがって、レートひずみ空間においてより良い動作点につながる。様々な他の利益も、本明細書で説明される再帰ベースのニューラルネットワークシステムによって与えられる。

【0416】

【0199】図18は、本明細書で説明される再帰ベースの機械学習技法のうちの1つまたは複数を使用してビデオを処理するプロセス1800の一例を示すフローチャートである。ブロック1802において、処理1800は、ニューラルネットワークシステムのエンコーダ部分によって、ニューラルネットワークシステムの現在の時間ステップの動作のための入力ビデオフレームと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された動き推定データと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からのニューラルネットワークシステムのデコーダ部分の少なくとも1つの再帰レイヤからの少なくとも1つの再帰状態データとを取得することを含む。場合によっては、再帰状態データは、少なくとも1つの再帰レイヤのノードの1つまたは複数の値を含む。いくつかの例では、ニューラルネットワークシステムの前の時間ステップの動作によって生成された前に再構成されたビデオフレームなどの他の入力データが、図5Aに示されているように、エンコーダ部分によって取得され得る。

10

【0417】

【0200】ブロック1804において、プロセス1800は、ニューラルネットワークシステムのエンコーダ部分によって、ニューラルネットワークシステムの現在の時間ステップの動作のための出力データを生成することを含む。出力データは、入力ビデオフレームと、前の時間ステップの動作からの再構成された動き推定データと、前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からの少なくとも1つの再帰状態データとに少なくとも部分的に基づいて生成される。出力データは、図5Aに示されているように、ニューラルネットワークシステムの前の時間ステップの動作によって生成された前に再構成されたビデオフレームなど、場合によっては、他の入力データに基づいて生成され得る。

20

【0418】

【0201】1806において、プロセス1800は、ニューラルネットワークシステムのデコーダ部分によって、現在の時間ステップの動作のための出力データと前の時間ステップの動作からのデコーダ部分の少なくとも1つの再帰レイヤからの少なくとも1つの再帰状態データとを取得することを含む。たとえば、図5Aに示されているように、現在の時間ステップ $t+1$ の動作について、デコーダ538は、量子化器534から出力データを取得することができ、デコーダ538の再帰レイヤは、前の時間ステップ t の動作からそれ自体に再帰状態データ h_t をフィードバックすることができる。場合によっては、ニューラルネットワークシステムのデコーダ部分は、複数の再帰レイヤを含み、複数の再帰レイヤの各再帰レイヤは、それぞれの再帰状態データを与える。

30

【0419】

【0202】ブロック1808において、プロセス1800は、出力データと前の時間ステップの動作からの少なくとも1つの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成することを含む。再構成されたビデオフレームは、入力ビデオフレームを表す。たとえば、再構成されたビデオフレームは、入力ビデオフレームの再構成されたバージョンである。再構成されたビデオフレームは、表示、記憶、および／または送信のための出力であり得る。たとえば、いくつかの例では、プロセス1800は、記憶媒体中に出力データを記憶することを含む。いくつかの例では、プロセス1800は、記憶媒体中に再構成されたビデオフレームを記憶することを含む。いくつかの例では、プロセス1800は、少なくとも1つのデバイスに送信媒体を介して再構成されたビデオフレームを送ることを含む。場合によっては、再構成されたビデオフレームは、表示、記憶、および／または送信のために出力される前に（たとえば、再構成プロセスによるものである視覚的アーティファクトを除去するために、または他の処理のために）さらに処理され得る。

40

【0420】

【0203】いくつかの例では、プロセス1800は、ニューラルネットワークシステムのデ

50

コーダ部分によって、出力データと前の時間ステップの動作からの少なくとも1つの再帰状態データとに少なくとも部分的に基づいて現在の時間ステップの動作のための動き推定データを決定することを含む。そのような例では、プロセス1800は、前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと、現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することによって現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することを含むことができる。

【0421】

【0204】いくつかの例では、プロセス1800は、ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの少なくとも1つの再帰状態データとに少なくとも部分的に基づいて現在の時間ステップの動作のための再構成された残差データを決定することを含む。そのような例では、プロセス1800は、現在の時間ステップのために決定された再構成された残差データにワーピングした再構成されたビデオフレームを加算することによって現在の時間ステップの動作のための再構成されたビデオフレームを生成することを含むことができる。

【0422】

【0205】いくつかの例では、プロセス1800は、ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの少なくとも1つの再帰状態データとに少なくとも部分的に基づいて現在の時間ステップの動作のための再構成された残差データを決定することを含む。そのような例では、プロセス1800は、デコーダ部分のポストワーピングネットワークによって、ワーピングした再構成されたビデオフレームと現在の時間ステップのために決定された再構成された残差データとを処理することを含むことができる。場合によっては、再構成されたビデオフレームは、ポストワーピングネットワークによるワーピングした再構成されたビデオフレームと再構成された残差データとの処理に基づいて現在の時間ステップの動作のために生成される。

【0423】

【0206】いくつかの例では、プロセス1800は、ニューラルネットワークシステムの動き推定ネットワークによって、入力ビデオフレームと前の時間ステップの動作からの前に再構成されたビデオフレームまたは前の時間ステップの動作からの前の入力ビデオフレームとのうちの少なくとも1つとに少なくとも部分的に基づいて現在の時間ステップの動作のための動き推定データを決定することを含む。そのような例では、プロセス1800は、現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームまたは前の入力ビデオフレームとのうちの少なくとも1つの1つまたは複数のピクセルを修正することによって現在の時間ステップのためのワーピングした再構成されたビデオフレームを生成することを含むことができる。プロセス1800は、ニューラルネットワークシステムのエンコーダ部分にワーピングした再構成されたビデオフレームと現在の時間ステップの動作のために決定された動き推定データとを送ることを含むことができる。場合によっては、現在の時間ステップの動作のために生成された出力データは、ワーピングした再構成されたビデオフレームと現在の時間ステップの動作のために決定された動き推定データとに少なくとも部分的に基づいてニューラルネットワークシステムのエンコーダ部分によって生成される。

【0424】

【0207】いくつかの例では、プロセス1800は、1つまたは複数のトレーニング反復のために、第1の損失関数を使用してニューラルネットワークシステムをトレーニングすることを含む。第1の損失関数は、1つまたは複数の入力ビデオフレームと1つまたは複数のワーピングした再構成されたビデオフレームとの間の損失を決定する。たとえば、第1の損失関数は、上記で説明された損失関数Loss1、Loss2、Loss3、および／またはLoss4のいずれかを含むことができる。1つの例示的な例では、エンコーダ533をトレーニングするためにLoss1が使用され得、デコーダ538をトレーニングするためにLoss3が使用され得る。別の例示的な例では、エンコーダ533をトレ

10

20

30

40

50

ーニングするために Loss 2 が使用され得、デコーダ 538 をトレーニングするために Loss 4 が使用され得る。そのような例では、プロセス 1800 は、第 1 の損失関数を非活性化すること（たとえば、Loss 1、Loss 2、Loss 3、および／または Loss 4 を非活性化すること）を含むことができる。たとえば、第 1 の損失関数は、図 8 および図 9 に関して上記で説明されたように、ニューラルネットワークシステムが動き推定パラメータを学習するためにトレーニングされると決定することに応答して非活性化され得る。そのような例では、プロセス 1800 は、1 つまたは複数のトレーニング反復の後に実施される 1 つまたは複数の後続のトレーニング反復について、第 2 の損失関数を使用してニューラルネットワークシステムをトレーニングすることを含むことができる。第 2 の損失関数は、少なくとも 1 つの入力ビデオフレームと少なくとも 1 つの再構成されたビデオフレームとの間の損失を決定する。たとえば、第 2 の損失関数は、上記で説明された Loss 0 を含むことができる。

【0425】

[0208]いくつかの例では、プロセス 1800 は、現在の時間ステップの動作または前の時間ステップの動作のために決定された動き推定データを使用して前の時間ステップの動作からの少なくとも 1 つの再帰状態データを修正することによって現在の時間ステップの動作のためのワーピングした再帰状態データを生成することを含む。そのような例では、現在の時間ステップの動作のために生成された再構成されたビデオフレームは、ワーピングした再帰状態データに少なくとも部分的に基づいて生成される。

【0426】

[0209]いくつかの例では、プロセス 1800 は、デコーダ部分によって、再構成されたマスクを生成することを含む。再構成されたマスクは、前の時間ステップの動作中に生成された前に再構成されたビデオフレーム中のデータの存在を示す 1 つまたは複数の値を含む。そのような例では、プロセス 1800 は、再構成されたマスクと、現在の時間ステップの動作のために生成されるワーピングした再構成されたビデオフレームと、現在の時間ステップの動作のために決定された再構成された残差データとに少なくとも部分的に基づいて現在の時間ステップの動作のための再構成されたビデオフレームを生成することを含む。再構成されたマスクは、ワーピングした再構成されたビデオフレームと再構成された残差データとの寄与を重み付けする。1 つの例示的な例では、図 12 に関して上記で説明されたように、マスクは、式 (4)

【0427】

【数 125】

$$(\hat{x}_t = \hat{m}_t \cdot \hat{f}_t(x_{t-1}) + (1 - \hat{m}_t) \cdot \hat{r}_t)$$

【0428】

を使用して実装され得る。

【0429】

[0210]いくつかの例では、プロセス 1800 は、前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することを含む。そのような例では、プロセス 1800 は、現在の時間ステップの動作中に、現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームの 1 つまたは複数のピクセルを修正することによって現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することを含む。そのような例では、プロセス 1800 は、現在の時間ステップの動作中にデコーダ部分によって、現在の時間ステップの動作のための第 1 の再構成された残差データを決定することを含むことができる。そのような例では、プロセス 1800 は、現在の時間ステップのために決定された再構成された残差データにワーピングした再構成されたビデオフレームを加算することによって現在の時間ステップの動作のための再構成されたビデオフレームを生成することを含むことができる。プロセス 1800 は、現在の時間ステップの動作中にデコーダ部分によって、現在の時間ステップの動作中にエンコーダ部分によって処理された前の入力ビデオフレームに少なくとも部分的に

基づいて前の時間ステップの動作からの第2の再構成された残差データを決定することを含むことができる。そのような例では、プロセス1800は、現在の時間ステップの動作中に、第2の再構成された残差データを使用して前に再構成されたビデオフレームを修正することと、現在の時間ステップの動作のための最終の出力ビデオフレームとして修正された前に再構成されたビデオフレームを出力することを含むことができる。たとえば、図13に関して上記で説明されたように、修正された前に再構成されたビデオフレーム

【0430】

【数126】

$$\hat{x}_t$$

10

【0431】

は、時間ステップ $t+1$ において再構成されたビデオフレーム

【0432】

【数127】

$$\hat{x}_{t+1}$$

【0433】

を出力する代わりに時間ステップ $t+1$ において（たとえば、表示、記憶、送信、および／または他の動作のために）出力され得る。そのような例では、再構成されたビデオフレーム

20

【0434】

【数128】

$$\hat{x}_{t+1}$$

【0435】

は、次の時間ステップ $t+2$ における修正のために出力され得る。

【0436】

【0211】いくつかの例では、プロセス1800は、ニューラルネットワークシステムのエンコーダ部分によって、1つまたは複数のひずみマップを取得することと、ニューラルネットワークシステムのエンコーダ部分によって、1つまたは複数のひずみマップに少なくとも部分的に基づいて出力データを生成することを含む。場合によっては、図14に関して上記で説明されたように、1つまたは複数のひずみマップは、現在の時間ステップの動作のための入力ビデオフレームと前の時間ステップの動作からの前に再構成されたビデオフレームとの間のひずみを示す第1のひずみマップを含むことができる。場合によっては、図14に関して上記で説明されたように、1つまたは複数のひずみマップは、現在の時間ステップの動作のための入力ビデオフレームとワーピングした再構成されたビデオフレームとの間のひずみを示す第2のひずみマップを含むことができる（ここで、ワーピングした再構成されたビデオフレームは、上記で説明されたように、前の時間ステップの動作からの前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することによって生成され得る）。

30

40

【0437】

【0212】いくつかの例では、プロセス1800は、ニューラルネットワークシステムのエンコーダ部分の第1のレイヤによって、入力データを処理することを含む。そのような例では、プロセス1800は、エンコーダ部分の第1のレイヤによって、エンコーダ部分の第1のレイヤのための複数の重み値を決定することを含む。そのような例では、プロセス1800は、複数の重み値にニューラルネットワークシステムのエンコーダ部分の第2のレイヤの重みを設定することと、エンコーダ部分の第2のレイヤによって、複数の重み値に設定された重みを使用して入力データを処理することを含む。たとえば、図15に関して上記で説明されたように、線形レイヤ581の重みは、線形レイヤ582によって出力された重み583によって決定され得、これは、線形レイヤ581によって実施される

50

計算を入力 580 自体の値に依存することを可能にすることができる。

【0438】

[0213]図 19 は、本明細書で説明される再帰ベースの機械学習技法のうちの 1 つまたは複数をを使用してビデオを処理するプロセス 1900 の別の例を示すフローチャートである。ブロック 1902 において、プロセス 1900 は、ニューラルネットワークシステムのデコーダ部分によって、ニューラルネットワークシステムの現在の時間ステップの動作のための出力データとニューラルネットワークシステムの前の時間ステップの動作からのデコーダ部分の少なくとも 1 つの再帰レイヤからの再帰状態データとを取得することを含む。説明のために図 5A の例を使用して、現在の時間ステップ $t+1$ の動作について、デコーダ 538 は、量子化器 534 から出力データを取得することができ、デコーダ 538 の再帰レイヤは、前の時間ステップ t の動作からそれ自体に再帰状態データ h_t をフィードバックすることができる。場合によっては、ニューラルネットワークシステムのデコーダ部分は、複数の再帰レイヤを含み、複数の再帰レイヤの各再帰レイヤは、それぞれの再帰状態データを与える。

10

【0439】

[0214]ブロック 1904 において、プロセス 1900 は、ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて現在の時間ステップの動作のための動き推定データを決定することを含む。ブロック 1906 において、プロセス 1900 は、前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することを含む。ブロック 1908 において、プロセス 1900 は、現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームの 1 つまたは複数のピクセルを修正することに少なくとも部分的によって現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することを含む。

20

【0440】

[0215]ブロック 1910 において、プロセス 1900 は、ワーピングした再構成されたビデオフレームに少なくとも部分的に基づいて動作の現在の時間ステップの動作のための再構成されたビデオフレームを生成することを含む。再構成されたビデオフレームは、ニューラルネットワークシステムのエンコーダ部分または別のニューラルネットワークシステムのエンコーダによって処理される入力ビデオフレームを表す。たとえば、再構成されたビデオフレームは、入力ビデオフレームの再構成されたバージョンである。再構成されたビデオフレームは、表示、記憶、および／または送信のための出力であり得る。たとえば、いくつかの例では、プロセス 1800 は、記憶媒体中に出力データを記憶することを含む。いくつかの例では、プロセス 1800 は、記憶媒体中に再構成されたビデオフレームを記憶することを含む。いくつかの例では、プロセス 1800 は、少なくとも 1 つのデバイスに送信媒体を介して再構成されたビデオフレームを送ることを含む。場合によっては、再構成されたビデオフレームは、表示、記憶、および／または送信のために出力される前に（たとえば、再構成プロセスによるものである視覚的アーティファクトを除去するために、または他の処理のために）さらに処理され得る。

30

【0441】

[0216]いくつかの例では、プロセス 1900 は、ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて現在の時間ステップの動作のための再構成された残差データを決定することを含む。そのような例では、プロセス 1900 は、現在の時間ステップのために決定された再構成された残差データにワーピングした再構成されたビデオフレームを加算することに少なくとも部分的によって現在の時間ステップの動作のための再構成されたビデオフレームを生成することを含むことができる。

40

【0442】

[0217]いくつかの例では、プロセス 1900 は、現在の時間ステップの動作中にデコーダ部分によって、現在の時間ステップの動作中にニューラルネットワークシステムのエン

50

コーダ部分によって処理された前の入力ビデオフレームに少なくとも部分的に基づいて前の時間ステップの動作からの再構成された残差データを決定することを含む。そのような例では、プロセス1900は、現在の時間ステップの動作中に、前の時間ステップの動作からの再構成された残差データを使用して前に再構成されたビデオフレームを修正することを含むことができる。プロセス1900は、現在の時間ステップの動作のための最終の出力ビデオフレームとして修正された前に再構成されたビデオフレームを出力することを含むことができる。

【0443】

【0218】いくつかの例では、プロセス1900は、ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて現在の時間ステップの動作のための再構成された残差データを決定することを含む。そのような例では、プロセス1900は、デコーダ部分のポストワーピングネットワークによって、ワーピングした再構成されたビデオフレームと現在の時間ステップのために決定された再構成された残差データとを処理することを含むことができる。プロセス1900は、ポストワーピングネットワークによるワーピングした再構成されたビデオフレームと再構成された残差データとの処理に少なくとも部分的に基づいて現在の時間ステップの動作のための再構成されたビデオフレームを生成することを含むことができる。

【0444】

【0219】いくつかの例では、プロセス1900は、現在の時間ステップの動作のために決定された前記動き推定データまたは前の時間ステップの動作のために決定された動き推定を使用して前の時間ステップの動作からの再帰状態データを修正することに少なくとも部分的によって現在の時間ステップの動作のためのワーピングした再帰状態データを生成することを含む。そのような例では、プロセス1900は、ワーピングした再帰状態データに少なくとも部分的に基づいて現在の時間ステップの動作のための再構成されたビデオフレームを生成することを含むことができる。

【0445】

【0220】いくつかの例では、プロセス1900は、デコーダ部分によって、再構成されたマスクを生成することを含む。再構成されたマスクは、前の時間ステップの動作中に生成された前に再構成されたビデオフレーム中のデータの存在を示す1つまたは複数の値を含む。そのような例では、プロセス1900は、再構成されたマスクと、現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームと、現在の時間ステップの動作のために決定された再構成された残差データとに少なくとも部分的に基づいて現在の時間ステップの動作のための再構成されたビデオフレームを生成することを含むことができる。再構成されたマスクは、ワーピングした再構成されたビデオフレームと再構成された残差データとの寄与を重み付けする。1つの例示的な例では、図12に関して上記で説明されたように、マスクは、式(4)

【0446】

【数129】

$$(\hat{x}_t = \hat{m}_t \cdot \hat{f}_t(x_{t-1}) + (1 - \hat{m}_t) \cdot \hat{r}_t)$$

【0447】

を使用して実装され得る。

【0448】

【0221】いくつかの例では、プロセス1900は、ニューラルネットワークシステムのエンコーダ部分によって、現在の時間ステップの動作のための入力ビデオフレームと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された動き推定データと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からのニューラルネットワークシステムのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することを含む。場合

によっては、再帰状態データは、少なくとも1つの再帰レイヤのノードの1つまたは複数の値を含む。いくつかの例では、ニューラルネットワークシステムの前の時間ステップの動作によって生成された前に再構成されたビデオフレームなどの他の入力データが、図5Aに示されているように、エンコーダ部分によって取得され得る。プロセス1900は、ニューラルネットワークシステムのエンコーダ部分によって、現在の時間ステップの動作のための出力データを生成すること、出力データは、入力ビデオフレームと、前の時間ステップの動作からの再構成された動き推定データと、前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて生成される、を含むことができる。出力データは、図5Aに示されているように、ニューラルネットワークシステムの前の時間ステップの動作によって生成された前に再構成されたビデオフレームなど、場合によっては、他の入力データに基づいて生成され得る。

【0449】

[0222]いくつかの例では、プロセス1900は、1つまたは複数のトレーニング反復のために、第1の損失関数を使用してニューラルネットワークシステムをトレーニングすることを含む。第1の損失関数は、1つまたは複数の入力ビデオフレームと1つまたは複数のワーピングした再構成されたビデオフレームとの間の損失を決定する。たとえば、第1の損失関数は、上記で説明された損失関数Loss1、Loss2、Loss3、および／またはLoss4のいずれかを含むことができる。1つの例示的な例では、エンコーダ533をトレーニングするためにLoss1が使用され得、デコーダ538をトレーニングするためにLoss3が使用され得る。別の例示的な例では、エンコーダ533をトレーニングするためにLoss2が使用され得、デコーダ538をトレーニングするためにLoss4が使用され得る。そのような例では、プロセス1900は、第1の損失関数を非活性化すること（たとえば、Loss1、Loss2、Loss3、および／またはLoss4を非活性化すること）を含むことができる。たとえば、第1の損失関数は、図8および図9に関して上記で説明されたように、ニューラルネットワークシステムが動き推定パラメータを学習するためにトレーニングされると決定することに応答して非活性化され得る。そのような例では、プロセス1900は、1つまたは複数のトレーニング反復の後に実施される1つまたは複数の後続のトレーニング反復について、第2の損失関数を使用してニューラルネットワークシステムをトレーニングすることを含むことができる。第2の損失関数は、少なくとも1つの入力ビデオフレームと少なくとも1つの再構成されたビデオフレームとの間の損失を決定する。たとえば、第2の損失関数は、上記で説明されたLoss0を含むことができる。

【0450】

[0223]いくつかの例では、プロセス1900は、前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することを含む。そのような例では、プロセス1900は、現在の時間ステップの動作中に、現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することによって現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することを含む。そのような例では、プロセス1900は、現在の時間ステップの動作中にデコーダ部分によって、現在の時間ステップの動作のための第1の再構成された残差データを決定することを含むことができる。そのような例では、プロセス1900は、現在の時間ステップのために決定された再構成された残差データにワーピングした再構成されたビデオフレームを加算することによって現在の時間ステップの動作のための再構成されたビデオフレームを生成することを含むことができる。プロセス1900は、現在の時間ステップの動作中にデコーダ部分によって、現在の時間ステップの動作中にエンコーダ部分によって処理された前の入力ビデオフレームに少なくとも部分的に基づいて前の時間ステップの動作からの第2の再構成された残差データを決定することを含むことができる。そのような例では、プロセス1900は、現在の時間ステップの動作中に、第2の再構成された残差データを使用して前に再構成されたビデオフレームを修正

することと、現在の時間ステップの動作のための最終の出力ビデオフレームとして修正された前に再構成されたビデオフレームを出力することを含むことができる。たとえば、図 13 に関して上記で説明されたように、修正された前に再構成されたビデオフレーム

【0451】

【数130】

$$\hat{x}_t$$

【0452】

は、時間ステップ $t+1$ において再構成されたビデオフレーム

【0453】

【数131】

$$\hat{x}_{t+1}$$

【0454】

を出力する代わりに、時間ステップ $t+1$ において（たとえば、表示、記憶、送信、および／または他の動作のために）出力され得る。そのような例では、再構成されたビデオフレーム

【0455】

【数132】

$$\hat{x}_{t+1}$$

【0456】

は、次の時間ステップ $t+2$ における修正のために出力され得る。

【0457】

[0224]いくつかの例では、本明細書で説明されるプロセス（たとえば、プロセス1800、プロセス1900、および／または本明細書で説明される他のプロセス）は、図20に示されているコンピューティングデバイスアーキテクチャ2000を有するコンピューティングデバイスなど、コンピューティングデバイスまたは装置によって実施され得る。いくつかの例は、コンピューティングデバイスは、モバイルデバイス、ウェアラブルデバイス、エクステンデッドリアリティデバイス（たとえば、仮想現実（VR）デバイス、拡張現実（AR）デバイス、もしくは複合現実（MR）デバイス）、パーソナルコンピュータ、ラップトップコンピュータ、ビデオサーバ、テレビジョン、または他のデバイスを含むことができる。一例では、プロセス1800および／またはプロセス1900は、（図5A～図15に示されているアーキテクチャのいずれかまたはそれらの任意の組合せを使用して）システム400またはシステム500を実装するコンピューティングデバイスアーキテクチャ2000をもつコンピューティングデバイスによって実施され得る。コンピューティングデバイスは、自律車両、ロボットデバイス、モバイルデバイス（たとえば、スマートフォン）、デスクトップコンピューティングデバイス、タブレットコンピューティングデバイス、ウェアラブルデバイス、および／またはプロセス1800を含む本明細書で説明されるプロセスを実施するリソース能力をもつ任意の他のコンピューティングデバイスなどの任意の好適なデバイスを含むことができる。場合によっては、コンピューティングデバイスまたは装置は、1つもしくは複数の入力デバイス、1つもしくは複数の出力デバイス、1つもしくは複数のプロセッサ、1つもしくは複数のマイクロプロセッサ、1つもしくは複数のマイクロコンピュータ、1つもしくは複数の送信機、受信機もしくは（たとえば、トランシーバと呼ばれる）組み合わされた送受信装置、1つもしくは複数のカメラ、1つもしくは複数のセンサー、および／または本明細書で説明されるプロセスのステップを行うように構成された他の構成要素などの様々な構成要素を含み得る。いくつかの例では、コンピューティングデバイスは、ディスプレイ、データを通信および／もしくは受信するように構成されたネットワークインターフェース、それらの任意の組合せ、ならびに／または他の構成要素を含み得る。ネットワークインターフェースは、インター

10

20

30

40

50

ネットプロトコル（ＩＰ）ベースのデータまたは他のタイプのデータを通信および／または受信するように構成され得る。

【０４５８】

【０２２５】コンピューティングデバイスの構成要素は、回路で実装され得る。たとえば、構成要素は、本明細書で説明される様々な動作を実施するために、１つまたは複数のプログラマブル電子回路（たとえば、マイクロプロセッサ、グラフィックス処理ユニット（GPU）、デジタル信号プロセッサ（DSP）、中央処理ユニット（CPU）、ニューラルプロセッシングユニット（NPU）、および／または他の好適な電子回路）を含むことができる、電子回路もしくは他の電子ハードウェアを含むことができ、および／またはそれらを使用して実装され得、ならびに／あるいは、コンピュータソフトウェア、ファームウェア、もしくはそれらの任意の組合せを含むことができ、および／またはそれらを使用して実装され得る。

10

【０４５９】

【０２２６】プロセス１８００およびプロセス１９００は、論理フロー図として示され、その動作は、ハードウェア、コンピュータ命令、またはそれらの組合せにおいて実施され得る動作のシーケンスを表す。コンピュータ命令の文脈では、動作は、１つまたは複数のプロセッサによって実行されたときに、記載された動作を実行する、１つまたは複数のコンピュータ可読記憶媒体に記憶されたコンピュータ実行可能命令を表す。概して、コンピュータ実行可能命令は、特定の機能を実施するか、または特定のデータタイプを実装する、ルーチン、プログラム、オブジェクト、構成要素、データ構造などを含む。動作が説明される順序は、限定として解釈されるものではなく、任意の数の説明される動作は、プロセスを実装するために任意の順序でおよび／または並行して組み合わせられ得る。

20

【０４６０】

【０２２７】さらに、（プロセス１８００、プロセス１９００、および／または本明細書で説明される他のプロセスを含む）本明細書で説明されるプロセスは、実行可能命令で構成された１つまたは複数のコンピュータシステムの制御下で実施され得、まとめて１つまたは複数のプロセッサ上で、ハードウェアによって、またはそれらの組合せで実行するコード（たとえば、実行可能命令、１つまたは複数のコンピュータプログラム、または１つまたは複数のアプリケーション）として実装され得る。上述のように、コードは、たとえば、１つまたは複数のプロセッサによって実行可能な複数の命令を備えるコンピュータプログラムの形態で、コンピュータ可読または機械可読記憶媒体に記憶され得る。コンピュータ可読または機械可読記憶媒体は非一時的であり得る。

30

【０４６１】

【０２２８】図２０は、本明細書で説明される様々な技法を実装することができる例示的なコンピューティングデバイスの例示的なコンピューティングデバイスアーキテクチャ２０００を示す。いくつかの例は、コンピューティングデバイスは、モバイルデバイス、ウェアラブルデバイス、エクステンデッドリアリティデバイス（たとえば、仮想現実（VR）デバイス、拡張現実（AR）デバイス、もしくは複合現実（MR）デバイス）、パーソナルコンピュータ、ラップトップコンピュータ、ビデオサーバ、または他のデバイスを含むことができる。たとえば、コンピューティングデバイスアーキテクチャ２０００は、（図５Ａ～図１５に示されているアーキテクチャのうちの任意の１つまたはそれらの任意の組合せを含む）システム４００および／またはシステム５００を実装することができる。コンピューティングデバイスアーキテクチャ２０００の構成要素は、バスなどの接続部２００５を使用して、互いに電気通信するように図示されている。例示的なコンピューティングデバイスアーキテクチャ２０００は、処理ユニット（CPUまたはプロセッサ）２０１０と、読み出し専用メモリ（ROM）２０２０およびランダムアクセスメモリ（RAM）２０２５などのコンピューティングデバイスメモリ２０１５を含む様々なコンピューティングデバイス構成要素をプロセッサ２０１０に結合するコンピューティングデバイス接続部２００５とを含む。

40

【０４６２】

50

【0229】コンピューティングデバイスアーキテクチャ2000は、プロセッサ2010と直接接続された、プロセッサ2010に極めて近接した、またはプロセッサ2010の一部として統合された、高速メモリのキャッシュを含むことができる。コンピューティングデバイスアーキテクチャ2000は、プロセッサ2010による迅速なアクセスのために、メモリ2015および／または記憶デバイス2030からキャッシュ2012にデータをコピーすることができる。このようにして、キャッシュは、データを待機する間のプロセッサ2010遅延を回避する性能の向上を提供することができる。これらのモジュールおよび他のモジュールは、様々なアクションを実施するために、プロセッサ2010を制御することができ、またはプロセッサ2010を制御するように構成され得る。他のコンピューティングデバイスメモリ2015も、使用のために利用可能であり得る。メモリ2015は、異なる性能特性をもつ、複数の異なるタイプのメモリを含むことができる。プロセッサ2010は、任意の汎用プロセッサ、ならびに、プロセッサ2010を制御するように構成された、記憶デバイス2030に記憶されたサービス1 2032、サービス2 2034、およびサービス3 2036などの、ハードウェアまたはソフトウェアサービス、ならびにソフトウェア命令がプロセッサ設計に組み込まれる専用プロセッサを含むことができる。プロセッサ2010は、複数のコアまたはプロセッサと、バスと、メモリコントローラと、キャッシュなどを含んでいる、自己完結型システムであり得る。マルチコアプロセッサは、対称的または非対称的であり得る。

【0463】

【0230】コンピューティングデバイスアーキテクチャ2000とのユーザ対話を可能にするために、入力デバイス2045は、音声のためのマイクロフォン、ジェスチャーまたはグラフィカル入力のためのタッチセンシティブスクリーン、キーボード、マウス、動き入力、音声など、任意の数の入力機構を表現することができる。出力デバイス2035も、ディスプレイ、プロジェクタ、テレビジョン、スピーカーデバイスなど、当業者に知られているいくつかの出力機構のうちの1つまたは複数であり得る。いくつかの事例では、マルチモーダルコンピューティングデバイスは、ユーザが、コンピューティングデバイスアーキテクチャ2000と通信するために複数のタイプの入力を提供することを可能にすることができる。通信インターフェース2040は、概して、ユーザ入力およびコンピューティングデバイス出力を統制および管理することができる。任意の特定のハードウェア構成上で動作することに対する制限はなく、したがって、ここでの基本的特徴は、改善されたハードウェア構成またはファームウェア構成が開発されるにつれて、それらで容易に代用され得る。

【0464】

【0231】記憶デバイス2030は、不揮発性メモリであり、磁気カセット、フラッシュメモリカード、固体メモリデバイス、デジタル多用途ディスク、カートリッジ、ランダムアクセスメモリ（RAM）2025、読取り専用メモリ（ROM）2020、およびそれらのハイブリッドなど、コンピュータによってアクセス可能であるデータを記憶することができるハードディスクまたは他のタイプのコンピュータ可読媒体であり得る。記憶デバイス2030は、プロセッサ2010を制御するためのサービス2032、2034、2036を含むことができる。他のハードウェアモジュールまたはソフトウェアモジュールが企図される。記憶デバイス2030は、コンピューティングデバイス接続2005に接続され得る。一態様では、特定の機能を実施するハードウェアモジュールは、その機能を行うために、プロセッサ2010、接続2005、出力デバイス2035など、必要なハードウェア構成要素に関して、コンピュータ可読媒体に記憶されたソフトウェア構成要素を含むことができる。

【0465】

【0232】本開示の態様は、1つまたは複数のアクティブ深度感知システムを含むか、またはそれに結合された（セキュリティシステム、スマートフォン、タブレット、ラップトップコンピュータ、車両、ドローン、または他のデバイスなどの）任意の好適な電子デバイスに適用可能である。1つの光プロジェクタを有するか、またはそれに結合されたデバイ

スに関して以下で説明されるが、本開示の態様は、任意の数の光プロジェクタを有するデバイスに適用可能であり、したがって、特定のデバイスに限定されない。

【0466】

【0233】「デバイス」という用語は、（1つのスマートフォン、1つのコントローラ、1つの処理システムなど）1つまたは特定の数の物理オブジェクトに限定されない。本明細書で使用するデバイスは、1つまたは複数の部分をもつ任意の電子デバイスであり得、本開示の少なくともいくつかの部分を実装し得る。以下の説明および例が本開示の様々な態様について説明するために「デバイス」という用語を使用するが、「デバイス」という用語は特定の構成、タイプ、または数のオブジェクトに限定されない。さらに、「システム」という用語は、複数の構成要素または特定の実施形態に限定されない。たとえば、システムは、1つまたは複数のプリント回路板または他の基板上に実装され得、可動のまたは静的な構成要素を有し得る。以下の説明および例が本開示の様々な態様について説明するために「システム」という用語を使用するが、「システム」という用語は特定の構成、タイプ、または数のオブジェクトに限定されない。

10

【0467】

【0234】本明細書で提供される実施形態および例の完全な理解を与えるために、具体的な詳細が上記の説明で与えられる。ただし、実施形態はこれらの具体的な詳細なしに実施され得ることを当業者は理解されよう。説明の明快のために、いくつかの事例では、本技術は、デバイス、デバイス構成要素、ソフトウェアで具現される方法におけるステップまたはルーチン、あるいはハードウェアとソフトウェアとの組合せを備える機能ブロックを含む個々の機能ブロックを含むものとして提示され得る。図に示されているおよび／または本明細書で説明される構成要素以外の追加の構成要素が使用され得る。たとえば、回路、システム、ネットワーク、プロセス、および他の構成要素は、実施形態を不要な詳細で不明瞭にしないためにブロック図の形態で構成要素として示され得る。他の事例では、実施形態を不明瞭にすることを回避するために、よく知られている回路、プロセス、アルゴリズム、構造、および技法が不要な詳細なしに示され得る。

20

【0468】

【0235】個々の実施形態は、フローチャート、フロー図、データフロー図、構造図、またはブロック図として描かれるプロセスまたは方法として上述され得る。フローチャートは、動作を逐次的なプロセスとして説明し得るが、動作の多くは、並行してまたは同時に実施され得る。さらに、動作の順序は並べ替えられ得る。その動作が完了されるとき、プロセスは終了されるが、図中に含まれない追加のステップを有し得る。プロセスは、方法、関数、プロシージャ、サブルーチン、サブプログラムなどに対応し得る。プロセスが機能に対応する場合、その終了は、呼出関数またはメイン関数への関数の戻り値に対応し得る。

30

【0469】

【0236】上述した例によるプロセスおよび方法は、記憶されている、またはさもなければコンピュータ可読媒体から利用可能である、コンピュータ実行可能な命令を使用して実装され得る。そのような命令は、たとえば、汎用コンピュータ、専用コンピュータ、もしくは処理デバイスに、一定の機能もしくは機能のグループを実行させる、またはさもなければ、汎用コンピュータ、専用コンピュータ、もしくは処理デバイスを一定の機能もしくは機能のグループを実施するように構成する、命令およびデータを含むことができる。使用されるコンピュータリソースの部分は、ネットワーク上でアクセス可能とすることができる。コンピュータ実行可能命令は、たとえば、バイナリ、アセンブリ言語などの中間フォーマット命令、ファームウェア、ソースコードなどであり得る。

40

【0470】

【0237】「コンピュータ可読媒体」という用語は、限定はしないが、ポータブルまたは非ポータブル記憶デバイス、光記憶デバイス、ならびに（1つまたは複数の）命令および／またはデータを記憶、含有、または搬送することが可能な様々な他の媒体を含む。コンピュータ可読媒体は、データがそこに記憶され得る非一時的媒体を含み得、それは、ワイヤ

50

レスにまたはワイヤード接続を介して伝搬する搬送波および／または一時的電子信号を含まない。非一時的媒体の例は、限定はしないが、特に、磁気ディスクまたはテープ、フラッシュメモリなどの光記憶媒体、メモリまたはメモリデバイス、磁気または光学ディスク、フラッシュメモリ、不揮発性メモリを備えたUSBデバイス、ネットワーク接続された記憶デバイス、コンパクトディスク（CD）またはデジタル多用途ディスク（DVD）、それらの任意の好適な組合せを含み得る。コンピュータ可読媒体は、プロシージャ、関数、サブプログラム、プログラム、ルーチン、サブルーチン、モジュール、ソフトウェアパッケージ、クラス、または命令、データ構造、もしくはプログラムステートメントの任意の組合せを表し得るコードおよび／または機械実行可能命令をその上に記憶していることがある。コードセグメントは、情報、データ、引数、パラメータ、またはメモリコンテンツをパスおよび／または受信することによって、別のコードセグメントまたはハードウェア回路に結合され得る。情報、引数、パラメータ、データなどは、メモリ共有、メッセージパッシング、トークンパッシング、ネットワーク送信などを含む、任意の好適な手段を介してパス、フォワーディング、または送信され得る。

【0471】

【0238】いくつかの実施形態では、コンピュータ可読記憶デバイス、媒体、およびメモリは、ビットストリームなどを含んでいるケーブル信号またはワイヤレス信号を含むことができる。しかしながら、述べられるとき、非一時的コンピュータ可読記憶媒体は、エネルギー、キャリア信号、電磁波、および信号自体などの媒体を明確に除外する。

【0472】

【0239】これらの開示によるプロセスおよび方法を実装するデバイスは、ハードウェア、ソフトウェア、ファームウェア、ミドルウェア、マイクロコード、ハードウェア記述言語、またはそれらの任意の組合せを含むことができ、様々なフォームファクタのいずれかをとることができる。ソフトウェア、ファームウェア、ミドルウェア、またはマイクロコードにおいて実装される場合、必要なタスクを実施するためのプログラムコードまたはコードセグメント（たとえば、コンピュータプログラム製品）は、コンピュータ可読媒体または機械可読媒体に記憶され得る。プロセッサは、必要なタスクを実施し得る。フォームファクタの典型的な例は、ラップトップコンピュータ、スマートフォン、携帯電話、タブレットデバイスまたは他の小型フォームファクタパーソナルコンピュータ、携帯情報端末、ラックマウントデバイス、スタンドアロンデバイス等を含む。本明細書において説明される機能性は、周辺機器またはアドインカードにおいても具体化され得る。そのような機能はまた、さらなる例として、単一のデバイスにおいて実行する異なるチップまたは異なるプロセスの間で回路板上に実装され得る。

【0473】

【0240】命令、そのような命令を伝達するための媒体、それらを実行するためのコンピューティングリソース、およびそのようなコンピューティングリソースをサポートするための他の構造は、本開示において説明される機能を提供するための例示的な手段である。

【0474】

【0241】上記の説明では、本出願の態様がその特定の実施形態を参照しながら説明されたが、本出願はそれに限定されないことを、当業者は認識されよう。したがって、本出願の例示的な実施形態が本明細書で詳細に説明されているが、従来技術によって限定される場合を除いて、発明的概念が、場合によっては様々な具体化および採用され得、添付の特許請求の範囲が、そのような変形形態を含むように解釈されるものであることを理解されたい。上記で説明された適用例の様々な特徴および態様は、個々にまたは一緒に使用され得る。さらに、実施形態は、本明細書のより広い趣旨および範囲から逸脱せずに、本明細書において説明される環境およびアプリケーションを超える、任意の数の環境およびアプリケーションにおいて利用され得る。したがって、本明細書および図面は、制限的なものというよりもむしろ例示的なものとして考慮されるべきである。例示の目的のために、方法は、特定の順序で説明された。代替実施形態では、方法は、説明された順序とは異なる順序で実施され得ることを諒解されたい。

【0475】

[0242]本明細書で使用される、よりも小さい（「<」）、および、よりも大きい（「>」）のシンボルまたは専門用語は、本明細書の範囲から逸脱することなく、それぞれ、よりも小さいかまたはそれに等しい（「≤」）、および、よりも大きいまたはそれに等しい（「≥」）のシンボルと置き換えられ得ることを、当業者は諒解されよう。

【0476】

[0243]構成要素が、ある動作を実施する「ように構成されて」いるものとして説明される場合、そのような構成は、たとえば、動作を実施するように電子回路または他のハードウェアを設計することによって、動作を実施するようにプログラマブル電子回路（たとえば、マイクロプロセッサ、または他の好適な電子回路）をプログラムすることによって、またはそれらの任意の組合せによって、達成され得る。

10

【0477】

[0244]「に結合される」という句は、直接的にもしくは間接的に別の構成要素に物理的に接続される任意の構成要素、および／または、直接的もしくは間接的に別の構成要素と通信する（たとえば、有線接続もしくはワイヤレス接続、および／または他の適切な通信インターフェース上で、他の構成要素に接続された）任意の構成要素を指す。

【0478】

[0245]セット「のうちの少なくとも1つ」、および／またはセットのうちの「1つもしくは複数」を具陳する特許請求の範囲の文言または他の文言は、セットのうちの1つのメンバーまたは（任意の組合せにおける）セットのうちの複数のメンバーが特許請求の範囲を満たすことを示す。たとえば、「AおよびBのうちの少なくとも1つ」あるいは「AまたはBのうちの少なくとも1つ」を具陳する特許請求の範囲の文言は、A、B、またはAおよびBを意味する。別の例では、「A、B、およびCのうちの少なくとも1つ」あるいは「A、B、またはCのうちの少なくとも1つ」を具陳する特許請求の範囲の文言は、A、B、C、もしくはAおよびB、もしくはAおよびC、もしくはBおよびC、もしくはAおよびBおよびCを意味する。文言セットの「うちの少なくとも1つ」および／またはセットのうちの「1つまたは複数」は、セットにリストされている項目にセットを制限しない。たとえば、「AおよびBのうちの少なくとも1つ」または「AもしくはBのうちの少なくとも1つ」を具陳するクレームの文言は、A、B、またはAおよびBを意味することができ、さらに、AおよびBのセットにリストされていない項目を含むことができる。

20

30

【0479】

[0246]本明細書において開示される実施形態に関連して説明される、様々な例示的な論理ブロック、モジュール、回路、およびアルゴリズムステップは、電子ハードウェア、コンピュータソフトウェア、ファームウェア、または、これらの組合せとして実装され得る。ハードウェアとソフトウェアとのこの互換性を明確に例示するために、様々な例示的な構成要素、ブロック、モジュール、回路、およびステップは、一般に、それらの機能性の観点から上述されている。そのような機能性がハードウェアとして実装されるか、またはソフトウェアとして実装されるかは、特定の適用例およびシステム全体に課される設計制約条件に依存する。熟練した技術者は、説明された機能性を特定の適用例ごとに様々な手法で実装し得るが、そのような実装上の決定は、本出願の範囲からの逸脱を引き起こすものとして解釈されるべきではない。

40

【0480】

[0247]本明細書において説明される技法は、電子ハードウェア、コンピュータソフトウェア、ファームウェア、または、これらの任意の組合せにおいても実装され得る。そのような技法は、汎用コンピュータ、ワイヤレス通信デバイスハンドセット、またはワイヤレス通信デバイスハンドセットおよび他のデバイスにおける適用を含む複数の用途を有する集積回路デバイスなど、様々なデバイスのいずれかにおいて実装され得る。モジュールまたは構成要素として説明された特徴は、集積論理デバイスと一緒に、または個別であるが相互運用可能な論理デバイスとして別個に実装され得る。ソフトウェアで実装された場合、本技法は、実行されたとき、上記で説明された方法のうちの1つまたは複数を実施する

50

命令を含むプログラムコードを備えるコンピュータ可読データ記憶媒体によって、少なくとも部分的に実現され得る。コンピュータ可読データ記憶媒体は、パッケージング材料を含み得るコンピュータプログラム製品の一部を形成し得る。コンピュータ可読媒体は、同期型ダイナミックランダムアクセスメモリ（SDRAM）などのランダムアクセスメモリ（RAM）、読み取り専用メモリ（ROM）、不揮発性ランダムアクセスメモリ（NVRAM）、電氣的消去可能プログラマブル読み取り専用メモリ（EEPROM（登録商標））、FLASH（登録商標）メモリ、磁気もしくは光学データ記憶媒体など、メモリまたはデータ記憶媒体を備え得る。本技法は、追加または代替として、伝搬信号または電波など、命令またはデータ構造の形式でプログラムコードを搬送または通信し、コンピュータによってアクセスされ、読み取られ、および／または実行され得るコンピュータ可読通信媒体によって少なくとも部分的に実現され得る。

【0481】

【0248】プログラムコードは、1つまたは複数のデジタル信号プロセッサ（DSP）、汎用マイクロプロセッサ、特定用途向け集積回路（ASIC）、フィールドプログラマブル論理アレイ（FPGA）、または他の等価な集積回路またはディスクリート論理回路など、1つまたは複数のプロセッサを含み得るプロセッサによって実行され得る。そのようなプロセッサは、本開示において説明される技法のうちのいずれかを実施するように構成され得る。汎用プロセッサはマイクロプロセッサであり得るが、代替として、プロセッサは、任意の従来のプロセッサ、コントローラ、マイクロコントローラ、または状態機械であり得る。プロセッサは、コンピューティングデバイスの組合せ、たとえば、DSPとマイクロプロセッサとの組合せ、複数のマイクロプロセッサ、DSPコアと連携する1つまたは複数のマイクロプロセッサ、または、任意の他のそのような構成としても実装され得る。したがって、本明細書で使用される「プロセッサ」という用語は、上記の構造、上記の構造の任意の組合せ、または、本明細書で説明される技法の実装に適した任意の他の構造もしくは装置のいずれかを指し得る。

【0482】

【0249】本開示の例示的な例は、以下を含む。

【0483】

【0250】態様1：ビデオデータを処理する方法であって、ニューラルネットワークシステムのエンコーダ部分によって、ニューラルネットワークシステムの現在の時間ステップの動作のための入力ビデオフレームと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された動き推定データと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からのニューラルネットワークシステムのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データと、を取得することと；ニューラルネットワークシステムのエンコーダ部分によって、ニューラルネットワークシステムの現在の時間ステップの動作のための出力データを生成することと、出力データは、入力ビデオフレームと、前の時間ステップの動作からの再構成された動き推定データと、前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて生成される；ニューラルネットワークシステムのデコーダ部分によって、現在の時間ステップの動作のための出力データと前の時間ステップの動作からのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと；出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、再構成されたビデオフレームは、入力ビデオフレームを表す、を備える方法。

【0484】

【0251】態様2：ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて現在の時間ステップの動作のための動き推定データを決定することと；前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと；現在の時間ステップ

のために決定された動き推定データを使用して前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することに少なくとも部分的によって現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと、をさらに備える、態様1に記載の方法。

【0485】

[0252]態様3：ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成された残差データを決定することと；現在の時間ステップのために決定された再構成された残差データにワーピングした再構成されたビデオフレームを加算することに少なくとも部分的によって、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、をさらに備える、態様2に記載の方法。

10

【0486】

[0253]態様4：ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成された残差データを決定することと、デコーダ部分のポストワーピングネットワークによって、ワーピングした再構成されたビデオフレームと現在の時間ステップのために決定された再構成された残差データとを処理することと、ここにおいて、再構成されたビデオフレームは、ポストワーピングネットワークによるワーピングした再構成されたビデオフレームと再構成された残差データとの処理に基づいて現在の時間ステップの動作のために生成される、をさらに備える、態様2および3のいずれかに記載の方法。

20

【0487】

[0254]態様5：ニューラルネットワークシステムの動き推定ネットワークによって、入力ビデオフレームと前の時間ステップの動作からの前に再構成されたビデオフレームまたは前の時間ステップの動作からの前の入力ビデオフレームとのうちの少なくとも1つとに少なくとも部分的に基づいて、現在の時間ステップの動作のための動き推定データを決定することと；現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームまたは前の入力ビデオフレームとのうちの少なくとも1つの1つまたは複数のピクセルを修正することに少なくとも部分的によって、現在の時間ステップのためのワーピングした再構成されたビデオフレームを生成することと；ニューラルネットワークシステムのエンコーダ部分にワーピングした再構成されたビデオフレームと現在の時間ステップの動作のために決定された動き推定データとを送ることと、ここにおいて、現在の時間ステップの動作のために生成された出力データは、ワーピングした再構成されたビデオフレームと現在の時間ステップの動作のために決定された動き推定データとに少なくとも部分的に基づいてニューラルネットワークシステムのエンコーダ部分によって生成される、をさらに備える、態様1から4のいずれかに記載の方法。

30

【0488】

[0255]態様6：1つまたは複数のトレーニング反復について、第1の損失関数を使用してニューラルネットワークシステムをトレーニングすることと、第1の損失関数は、1つまたは複数の入力ビデオフレームと1つまたは複数のワーピングした再構成されたビデオフレームとの間の損失を決定する；第1の損失関数を非活性化することと；1つまたは複数のトレーニング反復の後に実施される1つまたは複数の後続のトレーニング反復について、第2の損失関数を使用してニューラルネットワークシステムをトレーニングすることと、第2の損失関数は、少なくとも1つの入力ビデオフレームと少なくとも1つの再構成されたビデオフレームとの間の損失を決定する、をさらに備える、態様5に記載の方法。

40

【0489】

[0256]態様7：再帰状態データは、少なくとも1つの再帰レイヤのノードの1つまたは複数の値を含む、態様1から6のいずれかに記載の方法。

【0490】

[0257]態様8：ニューラルネットワークシステムのデコーダ部分は、複数の再帰レイヤ

50

を含む、複数の再帰レイヤの各再帰レイヤは、それぞれの再帰状態データを与える、態様 1 から 7 のいずれかに記載の方法。

【0491】

[0258]態様 9：現在の時間ステップの動作または前の時間ステップの動作のために決定された動き推定データを使用して前の時間ステップの動作からの再帰状態データを修正することに少なくとも部分的によって、現在の時間ステップの動作のためのワーピングした再帰状態データを生成すること、ここにおいて、現在の時間ステップの動作のために生成された再構成されたビデオフレームは、ワーピングした再帰状態データに少なくとも部分的に基づいて生成される、をさらに備える、態様 1 から 8 のいずれかに記載の方法。

【0492】

[0259]態様 10：デコーダ部分によって、再構成されたマスクを生成することと、再構成されたマスクは、前の時間ステップの動作中に生成された前に再構成されたビデオフレーム中のデータの存在を示す 1 つまたは複数の値を含む；再構成されたマスクと、現在の時間ステップの動作のために生成されるワーピングした再構成されたビデオフレームと、現在の時間ステップの動作のために決定された再構成された残差データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、再構成されたマスクは、ワーピングした再構成されたビデオフレームと再構成された残差データとの寄与を重み付けする、をさらに備える、態様 1 から 9 のいずれかに記載の方法。

【0493】

[0260]態様 11：前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと；現在の時間ステップの動作中に、現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームの 1 つまたは複数のピクセルを修正することに少なくとも部分的によって、現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと；現在の時間ステップの動作中にデコーダ部分によって、現在の時間ステップの動作のための第 1 の再構成された残差データを決定することと；現在の時間ステップのために決定された再構成された残差データにワーピングした再構成されたビデオフレームを加算することに少なくとも部分的によって、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと；現在の時間ステップの動作中にデコーダ部分によって、現在の時間ステップの動作中にエンコーダ部分によって処理された前の入力ビデオフレームに少なくとも部分的に基づいて、前の時間ステップの動作からの第 2 の再構成された残差データを決定することと；現在の時間ステップの動作中に、第 2 の再構成された残差データを使用して前に再構成されたビデオフレームを修正することと；現在の時間ステップの動作のための最終の出力ビデオフレームとして修正された前に再構成されたビデオフレームを出力することと、をさらに備える、態様 1 から 10 のいずれかに記載の方法。

【0494】

[0261]態様 12：ニューラルネットワークシステムのエンコーダ部分によって、1 つまたは複数のひずみマップを取得することと、ニューラルネットワークシステムのエンコーダ部分によって、1 つまたは複数のひずみマップに少なくとも部分的に基づいて出力データを生成することとをさらに備える、態様 1 から 11 のいずれかに記載の方法。

【0495】

[0262]態様 13：1 つまたは複数のひずみマップは、現在の時間ステップの動作のための入力ビデオフレームと前の時間ステップの動作からの前に再構成されたビデオフレームとの間のひずみを示す第 1 のひずみマップを含む、態様 12 に記載の方法。

【0496】

[0263]態様 14：1 つまたは複数のひずみマップは、現在の時間ステップの動作のための入力ビデオフレームとワーピングした再構成されたビデオフレームとの間のひずみを示す第 2 のひずみマップを含み、ワーピングした再構成されたビデオフレームは、前の時間ステップの動作からの前に再構成されたビデオフレームの 1 つまたは複数のピクセルを修

正することに少なくとも部分的によって生成される、態様 1 2 および 1 3 のいずれかに記載の方法。

【0497】

[0264]態様 1 5：ニューラルネットワークシステムのエンコーダ部分の第 1 のレイヤによって、入力データを処理することと；エンコーダ部分の第 1 のレイヤによって、エンコーダ部分の第 1 のレイヤのための複数の重み値を決定することと；複数の重み値にニューラルネットワークシステムのエンコーダ部分の第 2 のレイヤの重みを設定することと；エンコーダ部分の第 2 のレイヤによって、複数の重み値に設定された重みを使用して入力データを処理することと、をさらに備える、態様 1 から 1 4 のいずれかに記載の方法。

【0498】

[0265]態様 1 6：記憶媒体中に出力データを記憶することをさらに備える、態様 1 から 1 5 のいずれかに記載の方法。

【0499】

[0266]態様 1 7：記憶媒体中に再構成されたビデオフレームを記憶することをさらに備える、態様 1 から 1 6 のいずれかに記載の方法。

【0500】

[0267]態様 1 8：少なくとも 1 つのデバイスに送信媒体を介して再構成されたビデオフレームを送ることをさらに備える、態様 1 から 1 7 のいずれかに記載の方法。

【0501】

[0268]態様 1 9：ビデオデータを処理するための装置。装置は、（たとえば、回路中に実装された）メモリとメモリに結合されたプロセッサ（または複数のプロセッサ）とを含む。プロセッサは、ニューラルネットワークシステムのエンコーダ部分によって、ニューラルネットワークシステムの現在の時間ステップの動作のための入力ビデオフレームと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された動き推定データと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からのニューラルネットワークシステムのデコーダ部分の少なくとも 1 つの再帰レイヤからの再帰状態データと、を取得することと；ニューラルネットワークシステムのエンコーダ部分によって、ニューラルネットワークシステムの現在の時間ステップの動作のための出力データを生成することと、出力データは、入力ビデオフレームと、前の時間ステップの動作からの再構成された動き推定データと、前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて生成される；ニューラルネットワークシステムのデコーダ部分によって、現在の時間ステップの動作のための出力データと前の時間ステップの動作からのデコーダ部分の少なくとも 1 つの再帰レイヤからの再帰状態データとを取得することと；出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、再構成されたビデオフレームは、入力ビデオフレームを表す、を行うように構成される。

【0502】

[0269]態様 2 0：プロセッサは、ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて現在の時間ステップの動作のための動き推定データを決定することと；前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと；現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームの 1 つまたは複数のピクセルを修正することに少なくとも部分的によって現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと、を行うように構成された、態様 1 9 に記載の装置。

【0503】

[0270]態様 2 1：プロセッサは、ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的

10

20

30

40

50

に基づいて現在の時間ステップの動作のための再構成された残差データを決定することと；現在の時間ステップのために決定された再構成された残差データにワーピングした再構成されたビデオフレームを加算することに少なくとも部分的によって、現在の時間ステップの動作のための再構成されたビデオフレームを生成することを行うように構成された、態様 20 に記載の装置。

【0504】

[0271]態様 22：プロセッサは、ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成された残差データを決定することと；デコーダ部分のポストワーピングネットワークによって、ワーピングした再構成されたビデオフレームと現在の時間ステップのために決定された再構成された残差データとを処理することと、再構成されたビデオフレームは、ポストワーピングネットワークによるワーピングした再構成されたビデオフレームと再構成された残差データとの処理に基づいて現在の時間ステップの動作のために生成される、を行うように構成された、態様 20 および 21 のいずれかに記載の装置。

【0505】

[0272]態様 23：プロセッサは、ニューラルネットワークシステムの動き推定ネットワークによって、入力ビデオフレームと前の時間ステップの動作からの前に再構成されたビデオフレームまたは前の時間ステップの動作からの前の入力ビデオフレームとのうちの少なくとも 1 つとに少なくとも部分的に基づいて、現在の時間ステップの動作のための動き推定データを決定することと；現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームまたは前の入力ビデオフレームとのうちの少なくとも 1 つの 1 つまたは複数のピクセルを修正することに少なくとも部分的によって現在の時間ステップのためのワーピングした再構成されたビデオフレームを生成することと；ニューラルネットワークシステムのエンコーダ部分にワーピングした再構成されたビデオフレームと現在の時間ステップの動作のために決定された動き推定データとを送ることと、現在の時間ステップの動作のために生成された出力データは、ワーピングした再構成されたビデオフレームと現在の時間ステップの動作のために決定された動き推定データとに少なくとも部分的に基づいてニューラルネットワークシステムのエンコーダ部分によって生成される、を行うように構成された、態様 19 から 22 のいずれかに記載の装置。

【0506】

[0273]態様 24：プロセッサは、1 つまたは複数のトレーニング反復について、第 1 の損失関数を使用してニューラルネットワークシステムをトレーニングすることと、第 1 の損失関数は、1 つまたは複数の入力ビデオフレームと 1 つまたは複数のワーピングした再構成されたビデオフレームとの間の損失を決定する；1 つまたは複数のトレーニング反復の後に実施される 1 つまたは複数の後続のトレーニング反復について、第 2 の損失関数を使用してニューラルネットワークシステムをトレーニングすることと、第 2 の損失関数は、少なくとも 1 つの入力ビデオフレームと少なくとも 1 つの再構成されたビデオフレームとの間の損失を決定する、を行うように構成された、態様 23 に記載の装置。

【0507】

[0274]態様 25：再帰状態データは、少なくとも 1 つの再帰レイヤのノードの 1 つまたは複数の値を含む、態様 19 から 24 のいずれかに記載の装置。

【0508】

[0275]態様 26：ニューラルネットワークシステムのデコーダ部分は、複数の再帰レイヤを含む、複数の再帰レイヤの各再帰レイヤは、それぞれの再帰状態データを与える、態様 19 から 25 のいずれかに記載の装置。

【0509】

[0276]態様 27：プロセッサは、現在の時間ステップの動作または前の時間ステップの動作のために決定された動き推定データを使用して前の時間ステップの動作からの再帰状態データを修正することに少なくとも部分的によって、現在の時間ステップの動作のため

のワーピングした再帰状態データを生成すること、現在の時間ステップの動作のために生成された再構成されたビデオフレームは、ワーピングした再帰状態データに少なくとも部分的に基づいて生成される、を行うように構成された、態様 19 から 26 のいずれかに記載の装置。

【0510】

[0277]態様 28：プロセッサは、デコーダ部分によって、再構成されたマスクを生成することと、再構成されたマスクは、前の時間ステップの動作中に生成された前に再構成されたビデオフレーム中のデータの存在を示す 1 つまたは複数の値を含む；再構成されたマスクと、現在の時間ステップの動作のために生成されるワーピングした再構成されたビデオフレームと、現在の時間ステップの動作のために決定された再構成された残差データとに少なくとも部分的に基づいて現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、再構成されたマスクは、ワーピングした再構成されたビデオフレームと再構成された残差データとの寄与を重み付けする、を行うように構成された、態様 19 から 27 のいずれかに記載の装置。

【0511】

[0278]態様 29：プロセッサは、前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと；現在の時間ステップの動作中に、現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームの 1 つまたは複数のピクセルを修正することに少なくとも部分的によって、現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと；現在の時間ステップの動作中にデコーダ部分によって、現在の時間ステップの動作のための第 1 の再構成された残差データを決定することと；現在の時間ステップのために決定された再構成された残差データにワーピングした再構成されたビデオフレームを加算することに少なくとも部分的によって、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと；現在の時間ステップの動作中にデコーダ部分によって、現在の時間ステップの動作中にエンコーダ部分によって処理された前の入力ビデオフレームに少なくとも部分的に基づいて、前の時間ステップの動作からの第 2 の再構成された残差データを決定することと；現在の時間ステップの動作中に、第 2 の再構成された残差データを使用して前に再構成されたビデオフレームを修正することと；現在の時間ステップの動作のための最終の出力ビデオフレームとして修正された前に再構成されたビデオフレームを出力することと、を行うように構成された、態様 19 から 28 のいずれかに記載の装置。

【0512】

[0279]態様 30：プロセッサは、ニューラルネットワークシステムのエンコーダ部分によって、1 つまたは複数のひずみマップを取得することと；ニューラルネットワークシステムのエンコーダ部分によって、1 つまたは複数のひずみマップに少なくとも部分的に基づいて出力データを生成することと、を行うように構成された、態様 19 から 29 のいずれかに記載の装置。

【0513】

[0280]態様 31：1 つまたは複数のひずみマップは、現在の時間ステップの動作のための入力ビデオフレームと前の時間ステップの動作からの前に再構成されたビデオフレームとの間のひずみを示す第 1 のひずみマップを含む、態様 30 に記載の装置。

【0514】

[0281]態様 32：1 つまたは複数のひずみマップは、現在の時間ステップの動作のための入力ビデオフレームとワーピングした再構成されたビデオフレームとの間のひずみを示す第 2 のひずみマップを含み、ワーピングした再構成されたビデオフレームは、前の時間ステップの動作からの前に再構成されたビデオフレームの 1 つまたは複数のピクセルを修正することに少なくとも部分的によって生成される、態様 30 および 31 のいずれかに記載の装置。

【0515】

[0282]態様 33：プロセッサは、ニューラルネットワークシステムのエンコーダ部分の

第1のレイヤによって、入力データを処理することと；エンコーダ部分の第1のレイヤによって、エンコーダ部分の第1のレイヤのための複数の重み値を決定することと；複数の重み値にニューラルネットワークシステムのエンコーダ部分の第2のレイヤの重みを設定することと；エンコーダ部分の第2のレイヤによって、複数の重み値に設定された重みを使用して入力データを処理することと、を行うように構成された、態様19から32のいずれかに記載の装置。

【0516】

[0283]態様34：プロセッサは、出力データをメモリ中に記憶させることを行うように構成された、態様19から33のいずれかに記載の装置。

【0517】

[0284]態様35：プロセッサは、再構成されたビデオフレームをメモリ中に記憶させることを行うように構成された、態様19から34のいずれかに記載の装置。

【0518】

[0285]態様36：少なくとも1つのデバイスに送信媒体を介して再構成されたビデオフレームを送信するように構成された送信機をさらに備える、態様19から35のいずれかに記載の装置。

【0519】

[0286]態様37：プロセッサがニューラルプロセッシングユニット（NPU）を含む、態様19から36のいずれかに記載の装置。

【0520】

[0287]態様38：装置がモバイルデバイスを備える、態様19から37のいずれかに記載の装置。

【0521】

[0288]態様39：装置がエクステンデッドリアリティデバイスを備える、態様19から37に記載の装置。

【0522】

[0289]態様40：装置がテレビジョンを備える、態様19から37に記載の装置。

【0523】

[0290]態様41：ディスプレイをさらに備える、態様19から39のいずれかに記載の装置。

【0524】

[0291]態様42：装置は、1つまたは複数のビデオフレームをキャプチャするように構成されたカメラを備える、態様19から41のいずれかに記載の装置。

【0525】

[0292]態様43：ビデオデータを処理する方法であって、ニューラルネットワークシステムのデコーダ部分によって、ニューラルネットワークシステムの現在の時間ステップの動作のための出力データとニューラルネットワークシステムの前の時間ステップの動作からのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと；ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための動き推定データを決定することと；前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと；現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することに少なくとも部分的によって、現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと；ワーピングした再構成されたビデオフレームに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、を備える方法。

【0526】

[0293]態様44：ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて現在

10

20

30

40

50

の時間ステップの動作のための再構成された残差データを決定することと；現在の時間ステップのために決定された再構成された残差データにワーピングした再構成されたビデオフレームを加算することに少なくとも部分的によって現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、をさらに備える、態様 4 3 に記載の方法。
【0527】

【0294】態様 4 5：現在の時間ステップの動作中にデコーダ部分によって、現在の時間ステップの動作中にニューラルネットワークシステムのエンコーダ部分によって処理された前の入力ビデオフレームに少なくとも部分的に基づいて、前の時間ステップの動作からの再構成された残差データを決定することと；現在の時間ステップの動作中に、前の時間ステップの動作からの再構成された残差データを使用して前に再構成されたビデオフレームを修正することと；現在の時間ステップの動作のための最終の出力ビデオフレームとして修正された前に再構成されたビデオフレームを出力することと、をさらに備える、態様 4 3 および 4 4 のいずれかに記載の方法。

10

【0528】

【0295】態様 4 6：ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成された残差データを決定することと；デコーダ部分のポストワーピングネットワークによって、ワーピングした再構成されたビデオフレームと現在の時間ステップのために決定された再構成された残差データとを処理することと；ポストワーピングネットワークによるワーピングした再構成されたビデオフレームと再構成された残差データとの処理に少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、をさらに備える、態様 4 3 から 4 5 のいずれかに記載の方法。

20

【0529】

【0296】態様 4 7：現在の時間ステップの動作のために決定された動き推定データまたは前の時間ステップの動作のために決定された動き推定を使用して前の時間ステップの動作からの再帰状態データを修正することに少なくとも部分的によって、現在の時間ステップの動作のためのワーピングした再帰状態データを生成することと；ワーピングした再帰状態データに少なくとも部分的に基づいて現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、をさらに備える、態様 4 3 から 4 6 のいずれかに記載の方法。

30

【0530】

【0297】態様 4 8：デコーダ部分によって、再構成されたマスクを生成することと、再構成されたマスクは、前の時間ステップの動作中に生成された前に再構成されたビデオフレーム中のデータの存在を示す 1 つまたは複数の値を含み；再構成されたマスクと、現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームと、現在の時間ステップの動作のために決定された再構成された残差データとに少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、再構成されたマスクは、ワーピングした再構成されたビデオフレームと再構成された残差データとの寄与を重み付けする、をさらに備える、態様 4 3 から 4 7 のいずれかに記載の方法。

40

【0531】

【0298】態様 4 9：ニューラルネットワークシステムのエンコーダ部分によって、現在の時間ステップの動作のための入力ビデオフレームと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された動き推定データと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からのニューラルネットワークシステムのデコーダ部分の少なくとも 1 つの再帰レイヤからの再帰状態データとを取得することと；ニューラルネットワークシステムのエンコーダ部分によって、現在の時間ステップの動作のための出力データを生成することと、出力データは、入力ビデオフレームと、前の時間ステップの動作からの再構成された動き

50

推定データと、前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて生成される、をさらに備える、態様 43 から 48 のいずれかに記載の方法。

【0532】

【0299】態様 50：ビデオデータを処理するための装置。装置は、（たとえば、回路中に実装された）メモリとメモリに結合されたプロセッサ（または複数のプロセッサ）とを含む。プロセッサは、ニューラルネットワークシステムのデコーダ部分によって、ニューラルネットワークシステムの現在の時間ステップの動作のための出力データとニューラルネットワークシステムの前の時間ステップの動作からのデコーダ部分の少なくとも 1 つの再帰レイヤからの再帰状態データとを取得することと；ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて現在の時間ステップの動作のための動き推定データを決定することと；前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと；現在の時間ステップのために決定された動き推定データを使用して前に再構成されたビデオフレームの 1 つまたは複数のピクセルを修正することに少なくとも部分的によって現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと；ワーピングした再構成されたビデオフレームに少なくとも部分的に基づいて現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、を行うように構成される。

【0533】

【0300】態様 51：プロセッサは、ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて現在の時間ステップの動作のための再構成された残差データを決定することと；現在の時間ステップのために決定された再構成された残差データにワーピングした再構成されたビデオフレームを加算することに少なくとも部分的によって現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、を行うように構成された、態様 50 に記載の装置。

【0534】

【0301】態様 52：プロセッサは、現在の時間ステップの動作中にデコーダ部分によって、現在の時間ステップの動作中にニューラルネットワークシステムのエンコーダ部分によって処理された前の入力ビデオフレームに少なくとも部分的に基づいて、前の時間ステップの動作からの再構成された残差データを決定することと；現在の時間ステップの動作中に、前の時間ステップの動作からの再構成された残差データを使用して前に再構成されたビデオフレームを修正することと；現在の時間ステップの動作のための最終の出力ビデオフレームとして修正された前に再構成されたビデオフレームを出力することと、を行うように構成された、態様 50 および 51 のいずれかに記載の装置。

【0535】

【0302】態様 53：プロセッサは、ニューラルネットワークシステムのデコーダ部分によって、出力データと前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて現在の時間ステップの動作のための再構成された残差データを決定することと；デコーダ部分のポストワーピングネットワークによって、ワーピングした再構成されたビデオフレームと現在の時間ステップのために決定された再構成された残差データとを処理することと；ポストワーピングネットワークによるワーピングした再構成されたビデオフレームと再構成された残差データとの処理に少なくとも部分的に基づいて、現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、を行うように構成された、態様 50 から 52 のいずれかに記載の装置。

【0536】

【0303】態様 54：プロセッサは、現在の時間ステップの動作のために決定された動き推定データまたは前の時間ステップの動作のために決定された動き推定を使用して前の時間ステップの動作からの再帰状態データを修正することに少なくとも部分的によって、現在

の時間ステップの動作のためのワーピングした再帰状態データを生成することと；ワーピングした再帰状態データに少なくとも部分的に基づいて現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、を行うように構成された、態様 50 から 53 のいずれかに記載の装置。

【0537】

[0304]態様 55：プロセッサは、デコーダ部分によって、再構成されたマスクを生成することと、再構成されたマスクは、前の時間ステップの動作中に生成された前に再構成されたビデオフレーム中のデータの存在を示す 1 つまたは複数の値を含み；再構成されたマスクと、現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームと、現在の時間ステップの動作のために決定された再構成された残差データとに少なくとも部分的に基づいて現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、再構成されたマスクは、ワーピングした再構成されたビデオフレームと再構成された残差データと寄与を重み付けする、を行うように構成された、態様 50 から 54 のいずれかに記載の装置。

【0538】

[0305]態様 56：プロセッサは、ニューラルネットワークシステムのエンコーダ部分によって、現在の時間ステップの動作のための入力ビデオフレームと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された動き推定データと、ニューラルネットワークシステムの前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からのニューラルネットワークシステムのデコーダ部分の少なくとも 1 つの再帰レイヤからの再帰状態データとを取得することと；ニューラルネットワークシステムのエンコーダ部分によって、現在の時間ステップの動作のための出力データを生成することと、出力データは、入力ビデオフレームと、前の時間ステップの動作からの再構成された動き推定データと、前の時間ステップの動作からの再構成された残差データと、前の時間ステップの動作からの再帰状態データとに少なくとも部分的に基づいて生成される、を行うように構成された、態様 50 から 55 のいずれかに記載の装置。

【0539】

[0306]態様 57：プロセッサは、再構成されたビデオフレームをメモリ中に記憶させることを行うように構成された、態様 50 から 56 のいずれかに記載の装置。

【0540】

[0307]態様 58：プロセッサは、再構成されたビデオフレームを表示させることを行うように構成された、態様 50 から 57 のいずれかに記載の装置。

【0541】

[0308]態様 59：再構成されたビデオフレームを表示するように構成されたディスプレイをさらに備える、態様 50 から 58 のいずれかに記載の装置。

【0542】

[0309]態様 60：プロセッサがニューラルプロセッシングユニット（NPU）を含む、態様 50 から 59 のいずれかに記載の装置。

【0543】

[0310]態様 61：装置がモバイルデバイスを備える、態様 50 から 60 のいずれかに記載の装置。

【0544】

[0311]態様 62：装置がエクステンデッドリアリティデバイスを備える、態様 50 から 60 に記載の装置。

【0545】

[0312]態様 63：装置がテレビジョンを備える、態様 50 から 60 に記載の装置。

【0546】

[0313]態様 64：装置は、1 つまたは複数のビデオフレームをキャプチャするように構成されたカメラを備える、態様 50 から 63 のいずれかに記載の装置。

【0547】

10

20

30

40

50

[0314]態様65：実行されたとき、請求項1から18の動作のいずれかを実施することを1つまたは複数のプロセッサに行わせる命令を記憶したコンピュータ可読記憶媒体。

【0548】

[0315]態様66：請求項1から18の動作のいずれかを実施するための手段を備える装置。

【0549】

[0316]態様67：実行されたとき、請求項43から49の動作のいずれかを実施することを1つまたは複数のプロセッサに行わせる命令を記憶したコンピュータ可読記憶媒体。

【0550】

[0317]態様68：請求項43から49の動作のいずれかを実施するための手段を備える装置。

【0551】

[0318]態様69：実行されたとき、請求項1から18の動作のいずれかと請求項43から49の動作のいずれかとを実施することを1つまたは複数のプロセッサに行わせる命令を記憶したコンピュータ可読記憶媒体。

【0552】

[0319]態様70：請求項1から18の動作のいずれかと請求項8から49の動作のいずれかとを実施するための手段を備える装置。

以下に本願の出願当初の特許請求の範囲に記載された発明を付記する。

〔C1〕 ビデオデータを処理する方法であって、

ニューラルネットワークシステムのエンコーダ部分によって、前記ニューラルネットワークシステムの現在の時間ステップの動作のための入力ビデオフレームと、前記ニューラルネットワークシステムの前の時間ステップの動作からの再構成された動き推定データと、前記ニューラルネットワークシステムの前記前の時間ステップの動作からの再構成された残差データと、前記前の時間ステップの動作からの前記ニューラルネットワークシステムのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと、

前記ニューラルネットワークシステムの前記エンコーダ部分によって、前記ニューラルネットワークシステムの前記現在の時間ステップの動作のための出力データを生成することと、前記出力データは、前記入力ビデオフレームと、前記前の時間ステップの動作からの前記再構成された動き推定データと、前記前の時間ステップの動作からの前記再構成された残差データと、前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて生成され、

前記ニューラルネットワークシステムの前記デコーダ部分によって、前記現在の時間ステップの動作のための前記出力データと、前記前の時間ステップの動作からの前記デコーダ部分の前記少なくとも1つの再帰レイヤからの前記再帰状態データとを取得することと、

前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、前記再構成されたビデオフレームは、前記入力ビデオフレームを表し、

を備える方法。

〔C2〕 前記ニューラルネットワークシステムの前記デコーダ部分によって、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための動き推定データを決定することと、

前記前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと、

前記現在の時間ステップのために決定された前記動き推定データを使用して、前記前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することに少なくとも部分的によって、前記現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと、

10

20

30

40

50

をさらに備える、C 1に記載の方法。

〔C 3〕 前記ニューラルネットワークシステムの前記デコーダ部分によって、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成された残差データを決定することと、前記現在の時間ステップのために決定された前記再構成された残差データに前記ワーピングした再構成されたビデオフレームを加算することに少なくとも部分的によって、前記現在の時間ステップの動作のための前記再構成されたビデオフレームを生成することと、をさらに備える、C 2に記載の方法。

〔C 4〕 前記ニューラルネットワークシステムの前記デコーダ部分によって、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成された残差データを決定することと、前記デコーダ部分のポストワーピングネットワークによって、前記ワーピングした再構成されたビデオフレームと前記現在の時間ステップのために決定された前記再構成された残差データとを処理することと、

をさらに備え、

前記再構成されたビデオフレームは、前記ポストワーピングネットワークによる前記ワーピングした再構成されたビデオフレームと前記再構成された残差データとの処理に基づいて、前記現在の時間ステップの動作のために生成される、

C 2に記載の方法。

〔C 5〕 前記ニューラルネットワークシステムの動き推定ネットワークによって、前記入力ビデオフレームと、前記前の時間ステップの動作からの前に再構成されたビデオフレームまたは前記前の時間ステップの動作からの前の入力ビデオフレームのうちの少なくとも1つと、に少なくとも部分的に基づいて、前記現在の時間ステップの動作のための動き推定データを決定することと、

前記現在の時間ステップのために決定された前記動き推定データを使用して、前記前に再構成されたビデオフレームまたは前記前の入力ビデオフレームのうちの少なくとも1つの1つまたは複数のピクセルを修正することに少なくとも部分的によって、前記現在の時間ステップのためのワーピングした再構成されたビデオフレームを生成することと、

前記ニューラルネットワークシステムの前記エンコーダ部分に、前記ワーピングした再構成されたビデオフレームと前記現在の時間ステップの動作のために決定された前記動き推定データとを送ることと、

をさらに備え、

前記現在の時間ステップの動作のために生成された前記出力データは、前記ワーピングした再構成されたビデオフレームと前記現在の時間ステップの動作のために決定された前記動き推定データとに少なくとも部分的に基づいて、前記ニューラルネットワークシステムの前記エンコーダ部分によって生成される、

C 1に記載の方法。

〔C 6〕 1つまたは複数のトレーニング反復について、第1の損失関数を使用して前記ニューラルネットワークシステムをトレーニングすることと、前記第1の損失関数は、1つまたは複数の入力ビデオフレームと1つまたは複数のワーピングした再構成されたビデオフレームとの間の損失を決定し、

前記第1の損失関数を非活性化することと、

前記1つまたは複数のトレーニング反復の後に実施される1つまたは複数の後続のトレーニング反復について、第2の損失関数を使用して前記ニューラルネットワークシステムをトレーニングすることと、前記第2の損失関数は、少なくとも1つの入力ビデオフレームと少なくとも1つの再構成されたビデオフレームとの間の損失を決定し、

をさらに備える、C 5に記載の方法。

〔C 7〕 前記再帰状態データは、前記少なくとも1つの再帰レイヤのノードの1つまたは複数の値を含む、C 1に記載の方法。

〔C 8〕 前記ニューラルネットワークシステムの前記デコーダ部分は、複数の再帰レイ

10

20

30

40

50

ヤを含み、前記複数の再帰レイヤの各再帰レイヤは、それぞれの再帰状態データを与える、C 1に記載の方法。

〔C 9〕 前記現在の時間ステップの動作または前記前の時間ステップの動作のために決定された動き推定データを使用して、前記前の時間ステップの動作からの前記再帰状態データを修正することに少なくとも部分的によって、前記現在の時間ステップの動作のためのワーピングした再帰状態データを生成すること、

をさらに備え、

前記現在の時間ステップの動作のために生成された前記再構成されたビデオフレームは、前記ワーピングした再帰状態データに少なくとも部分的に基づいて生成される、

C 1に記載の方法。

10

〔C 10〕 前記デコーダ部分によって、再構成されたマスクを生成することと、前記再構成されたマスクは、前記前の時間ステップの動作中に生成された前に再構成されたビデオフレーム中のデータの存在を示す1つまたは複数の値を含み、

前記再構成されたマスクと、前記現在の時間ステップの動作のために生成されるワーピングした再構成されたビデオフレームと、前記現在の時間ステップの動作のために決定された再構成された残差データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための前記再構成されたビデオフレームを生成することと、前記再構成されたマスクは、前記ワーピングした再構成されたビデオフレームと前記再構成された残差データとの寄与を重み付けし、

をさらに備える、C 1に記載の方法。

20

〔C 11〕 前記前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと、

前記現在の時間ステップの動作中に、前記現在の時間ステップのために決定された動き推定データを使用して、前記前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することに少なくとも部分的によって、前記現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと、

前記現在の時間ステップの動作中に前記デコーダ部分によって、前記現在の時間ステップの動作のための第1の再構成された残差データを決定することと、

前記現在の時間ステップのために決定された前記再構成された残差データに前記ワーピングした再構成されたビデオフレームを加算することに少なくとも部分的によって、前記現在の時間ステップの動作のための前記再構成されたビデオフレームを生成することと、前記現在の時間ステップの動作中に前記デコーダ部分によって、前記現在の時間ステップの動作中に前記エンコーダ部分によって処理された前の入力ビデオフレームに少なくとも部分的に基づいて、前記前の時間ステップの動作からの第2の再構成された残差データを決定することと、

30

前記現在の時間ステップの動作中に、前記第2の再構成された残差データを使用して、前記前に再構成されたビデオフレームを修正することと、

前記現在の時間ステップの動作のための最終の出力ビデオフレームとして、前記修正された前に再構成されたビデオフレームを出力することと、

をさらに備える、C 1に記載の方法。

40

〔C 12〕 前記ニューラルネットワークシステムの前記エンコーダ部分によって、1つまたは複数のひずみマップを取得することと、

前記ニューラルネットワークシステムの前記エンコーダ部分によって、前記1つまたは複数のひずみマップに少なくとも部分的に基づいて前記出力データを生成することと、をさらに備える、C 1に記載の方法。

〔C 13〕 前記1つまたは複数のひずみマップは、前記現在の時間ステップの動作のための前記入力ビデオフレームと前記前の時間ステップの動作からの前に再構成されたビデオフレームとの間のひずみを示す第1のひずみマップを含む、C 12に記載の方法。

〔C 14〕 前記1つまたは複数のひずみマップは、前記現在の時間ステップの動作のための前記入力ビデオフレームとワーピングした再構成されたビデオフレームとの間のひず

50

みを示す第2のひずみマップを含み、

前記ワーピングした再構成されたビデオフレームは、前記前の時間ステップの動作からの前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することに少なくとも部分的によって生成される、C12に記載の方法。

〔C15〕 前記ニューラルネットワークシステムの前記エンコーダ部分の第1のレイヤによって、入力データを処理することと、

前記エンコーダ部分の前記第1のレイヤによって、前記エンコーダ部分の前記第1のレイヤのための複数の重み値を決定することと、

前記複数の重み値に前記ニューラルネットワークシステムの前記エンコーダ部分の第2のレイヤの重みを設定することと、

前記エンコーダ部分の前記第2のレイヤによって、前記複数の重み値に設定された前記重みを使用して前記入力データを処理することと、

をさらに備える、C1に記載の方法。

〔C16〕 記憶媒体中に前記出力データを記憶すること、

をさらに備える、C1に記載の方法。

〔C17〕 記憶媒体中に前記再構成されたビデオフレームを記憶すること、

をさらに備える、C1に記載の方法。

〔C18〕 少なくとも1つのデバイスに、送信媒体を介して前記再構成されたビデオフレームを送ること、

をさらに備える、C1に記載の方法。

〔C19〕 ビデオデータを処理するための装置であって、

メモリと、

前記メモリに結合されたプロセッサと、

を備え、前記プロセッサは、

ニューラルネットワークシステムのエンコーダ部分を使用して、前記ニューラルネットワークシステムの現在の時間ステップの動作のための入力ビデオフレームと、前記ニューラルネットワークシステムの前の時間ステップの動作からの再構成された動き推定データと、前記ニューラルネットワークシステムの前記前の時間ステップの動作からの再構成された残差データと、前記前の時間ステップの動作からの前記ニューラルネットワークシステムのデコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと、

前記エンコーダ部分を使用して、前記ニューラルネットワークシステムの前記現在の時間ステップの動作のための出力データを生成することと、前記出力データは、前記入力ビデオフレームと、前記前の時間ステップの動作からの前記再構成された動き推定データと、前記前の時間ステップの動作からの前記再構成された残差データと、前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて生成され、

ニューラルネットワークシステムの前記デコーダ部分を使用して、前記現在の時間ステップの動作のための前記出力データと、前記前の時間ステップの動作からの前記デコーダ部分の前記少なくとも1つの再帰レイヤからの前記再帰状態データとを取得することと、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、前記再構成されたビデオフレームは、前記入力ビデオフレームを表し、

を行うように構成された、装置。

〔C20〕 前記プロセッサは、

前記ニューラルネットワークシステムの前記デコーダ部分を使用して、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための動き推定データを決定することと、

前記前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと、

前記現在の時間ステップのために決定された前記動き推定データを使用して、前記前に

10

20

30

40

50

再構成されたビデオフレームの1つまたは複数のピクセルを修正することによって、前記現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと、

を行うように構成された、C 1 9 に記載の装置。

〔C 2 1〕 前記プロセッサは、

前記ニューラルネットワークシステムの前記デコーダ部分を使用して、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成された残差データを決定することと、前記現在の時間ステップのために決定された前記再構成された残差データに前記ワーピングした再構成されたビデオフレームを加算することによって、前記現在の時間ステップの動作のための前記再構成されたビデオフレームを生成することと、

を行うように構成された、C 2 0 に記載の装置。

〔C 2 2〕 前記プロセッサは、

前記ニューラルネットワークシステムの前記デコーダ部分を使用して、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成された残差データを決定することと、前記デコーダ部分のポストワーピングネットワークを使用して、前記ワーピングした再構成されたビデオフレームと前記現在の時間ステップのために決定された前記再構成された残差データとを処理することと、

を行うように構成され、

前記再構成されたビデオフレームは、前記ポストワーピングネットワークによる前記ワーピングした再構成されたビデオフレームと前記再構成された残差データとの処理に基づいて、前記現在の時間ステップの動作のために生成される、

C 2 1 に記載の装置。

〔C 2 3〕 前記プロセッサは、

前記ニューラルネットワークシステムの動き推定ネットワークを使用して、前記入力ビデオフレームと、前記前の時間ステップの動作からの前に再構成されたビデオフレームまたは前記前の時間ステップの動作からの前の入力ビデオフレームのうちの少なくとも1つと、に少なくとも部分的に基づいて、前記現在の時間ステップの動作のための動き推定データを決定することと、

前記現在の時間ステップのために決定された前記動き推定データを使用して、前記前に再構成されたビデオフレームまたは前記前の入力ビデオフレームのうちの少なくとも1つの1つまたは複数のピクセルを修正することによって、前記現在の時間ステップのためのワーピングした再構成されたビデオフレームを生成することと、

前記ニューラルネットワークシステムの前記エンコーダ部分に、前記ワーピングした再構成されたビデオフレームと前記現在の時間ステップの動作のために決定された前記動き推定データとを送ることと、

を行うように構成され、

前記現在の時間ステップの動作のために生成された前記出力データは、前記ワーピングした再構成されたビデオフレームと前記現在の時間ステップの動作のために決定された前記動き推定データとに少なくとも部分的に基づいて、前記ニューラルネットワークシステムのの前記エンコーダ部分によって生成される、

C 1 9 に記載の装置。

〔C 2 4〕 前記プロセッサは、

1つまたは複数のトレーニング反復について、第1の損失関数を使用して前記ニューラルネットワークシステムをトレーニングすることと、前記第1の損失関数は、1つまたは複数の入力ビデオフレームと1つまたは複数のワーピングした再構成されたビデオフレームとの間の損失を決定し、

前記第1の損失関数を非活性化することと、

前記1つまたは複数のトレーニング反復の後に実施される1つまたは複数の後続のトレ

10

20

30

40

50

ーニング反復について、第2の損失関数を使用して前記ニューラルネットワークシステムをトレーニングすることと、前記第2の損失関数は、少なくとも1つの入力ビデオフレームと少なくとも1つの再構成されたビデオフレームとの間の損失を決定し、
を行うように構成された、C 2 3に記載の装置。

〔C 2 5〕 前記再帰状態データは、前記少なくとも1つの再帰レイヤのノードの1つまたは複数の値を含む、C 1 9に記載の装置。

〔C 2 6〕 前記ニューラルネットワークシステムの前記デコーダ部分は、複数の再帰レイヤを含み、前記複数の再帰レイヤの各再帰レイヤは、それぞれの再帰状態データを与える、C 1 9に記載の装置。

〔C 2 7〕 前記プロセッサは、
前記現在の時間ステップの動作または前記前の時間ステップの動作のために決定された動き推定データを使用して、前記前の時間ステップの動作からの前記再帰状態データを修正することによって、前記現在の時間ステップの動作のためのワーピングした再帰状態データを生成すること、

を行うように構成され、
前記現在の時間ステップの動作のために生成された前記再構成されたビデオフレームは、前記ワーピングした再帰状態データに少なくとも部分的に基づいて生成される、
C 1 9に記載の装置。

〔C 2 8〕 前記プロセッサは、
前記デコーダ部分を使用して、再構成されたマスクを生成することと、前記再構成されたマスクは、前記前の時間ステップの動作中に生成された前に再構成されたビデオフレーム中のデータの存在を示す1つまたは複数の値を含み、

前記デコーダ部分を使用して、前記再構成されたマスクと、前記現在の時間ステップの動作のために生成されるワーピングした再構成されたビデオフレームと、前記現在の時間ステップの動作のために決定された再構成された残差データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための前記再構成されたビデオフレームを生成することと、前記再構成されたマスクは、前記ワーピングした再構成されたビデオフレームと前記再構成された残差データとの寄与を重み付けし、
を行うように構成された、C 1 9に記載の装置。

〔C 2 9〕 前記プロセッサは、
前記前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと、

前記デコーダ部分を使用して、前記現在の時間ステップの動作中に、前記現在の時間ステップのために決定された動き推定データを使用して、前記前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することによって、前記現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと、

前記現在の時間ステップの動作中に前記デコーダ部分を使用して、前記現在の時間ステップの動作のための第1の再構成された残差データを決定することと、

前記現在の時間ステップのために決定された前記再構成された残差データに前記ワーピングした再構成されたビデオフレームを加算することによって、前記現在の時間ステップの動作のための前記再構成されたビデオフレームを生成することと、

前記現在の時間ステップの動作中に前記デコーダ部分を使用して、前記現在の時間ステップの動作中に前記エンコーダ部分によって処理された前の入力ビデオフレームに少なくとも部分的に基づいて、前記前の時間ステップの動作からの第2の再構成された残差データを決定することと、

前記現在の時間ステップの動作中に前記デコーダ部分を使用して、前記第2の再構成された残差データを使用して、前記前に再構成されたビデオフレームを修正することと、
前記現在の時間ステップの動作のための最終の出力ビデオフレームとして、前記修正された前に再構成されたビデオフレームを出力することと、
を行うように構成された、C 1 9に記載の装置。

10

20

30

40

50

〔C 3 0〕 前記プロセッサは、

前記ニューラルネットワークシステムの前記エンコーダ部分を使用して、1つまたは複数のひずみマップを取得することと、

前記ニューラルネットワークシステムの前記エンコーダ部分を使用して、前記1つまたは複数のひずみマップに少なくとも部分的に基づいて前記出力データを生成することと、を行うように構成された、C 1 9に記載の装置。

〔C 3 1〕 前記1つまたは複数のひずみマップは、前記現在の時間ステップの動作のための前記入力ビデオフレームと前記前の時間ステップの動作からの前に再構成されたビデオフレームとの間のひずみを示す第1のひずみマップを含む、C 3 0に記載の装置。

〔C 3 2〕 前記1つまたは複数のひずみマップは、前記現在の時間ステップの動作のための前記入力ビデオフレームとワーピングした再構成されたビデオフレームとの間のひずみを示す第2のひずみマップを含み、

前記ワーピングした再構成されたビデオフレームは、前記前の時間ステップの動作からの前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することによって生成される、C 3 0に記載の装置。

〔C 3 3〕 前記プロセッサは、

前記ニューラルネットワークシステムの前記エンコーダ部分の第1のレイヤを使用して、入力データを処理することと、

前記エンコーダ部分の前記第1のレイヤを使用して、前記エンコーダ部分の前記第1のレイヤのための複数の重み値を決定することと、

前記複数の重み値に前記ニューラルネットワークシステムの前記エンコーダ部分の第2のレイヤの重みを設定することと、

前記エンコーダ部分の前記第2のレイヤを使用して、前記複数の重み値に設定された前記重みを使用して前記入力データを処理することと、

を行うように構成された、C 1 9に記載の装置。

〔C 3 4〕 前記プロセッサは、

前記出力データを前記メモリ中に記憶させることと、

を行うように構成された、C 1 9に記載の装置。

〔C 3 5〕 前記プロセッサは、

前記再構成されたビデオフレームを前記メモリ中に記憶させることと、

を行うように構成された、C 1 9に記載の装置。

〔C 3 6〕 少なくとも1つのデバイスに、送信媒体を介して前記再構成されたビデオフレームを送信するように構成された送信機、

をさらに備える、C 1 9に記載の装置。

〔C 3 7〕 前記プロセッサは、ニューラルプロセッシングユニット（NPU）を含む、C 1 9に記載の装置。

〔C 3 8〕 前記装置はモバイルデバイスを備える、C 1 9に記載の装置。

〔C 3 9〕 前記装置がエクステンデッドリアリティデバイスを備える、C 1 9に記載の装置。

〔C 4 0〕 ディスプレイをさらに備える、C 1 9に記載の装置。

〔C 4 1〕 前記装置はテレビジョンを備える、C 1 9に記載の装置。

〔C 4 2〕 前記装置は、1つまたは複数のビデオフレームをキャプチャするように構成されたカメラを備える、C 1 9に記載の装置。

〔C 4 3〕 ビデオデータを処理する方法であって、

ニューラルネットワークシステムのデコーダ部分によって、前記ニューラルネットワークシステムの現在の時間ステップの動作のための出力データと、前記ニューラルネットワークシステムの前の時間ステップの動作からの前記デコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと、

前記ニューラルネットワークシステムの前記デコーダ部分によって、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて

10

20

30

40

50

、前記現在の時間ステップの動作のための動き推定データを決定することと、
前記前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと、
前記現在の時間ステップのために決定された前記動き推定データを使用して、前記前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することに少なくとも部分的によって、前記現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと、
前記ワーピングした再構成されたビデオフレームに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、
を備える方法。

〔C 4 4〕 前記ニューラルネットワークシステムの前記デコーダ部分によって、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成された残差データを決定することと、前記現在の時間ステップのために決定された前記再構成された残差データに前記ワーピングした再構成されたビデオフレームを加算することに少なくとも部分的によって、前記現在の時間ステップの動作のための前記再構成されたビデオフレームを生成することと、をさらに備える、C 4 3に記載の方法。

〔C 4 5〕 前記現在の時間ステップの動作中に前記デコーダ部分によって、前記現在の時間ステップの動作中に前記ニューラルネットワークシステムのエンコーダ部分によって処理された前の入力ビデオフレームに少なくとも部分的に基づいて、前記前の時間ステップの動作からの再構成された残差データを決定することと、
前記現在の時間ステップの動作中に、前記前の時間ステップの動作からの前記再構成された残差データを使用して、前記前に再構成されたビデオフレームを修正することと、
前記現在の時間ステップの動作のための最終の出力ビデオフレームとして、前記修正された前に再構成されたビデオフレームを出力することと、
をさらに備える、C 4 4に記載の方法。

〔C 4 6〕 前記ニューラルネットワークシステムの前記デコーダ部分によって、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成された残差データを決定することと、前記デコーダ部分のポストワーピングネットワークによって、前記現在の時間ステップのために決定された前記再構成された残差データと前記ワーピングした再構成されたビデオフレームとを処理することと、
前記ポストワーピングネットワークによる前記ワーピングした再構成されたビデオフレームと前記再構成された残差データとの処理に少なくとも部分的に基づいて、前記現在の時間ステップの動作のための前記再構成されたビデオフレームを生成することと、
をさらに備える、C 4 3に記載の方法。

〔C 4 7〕 前記現在の時間ステップの動作のために決定された前記動き推定データまたは前記前の時間ステップの動作のために決定された動き推定データを使用して、前記前の時間ステップの動作からの前記再帰状態データを修正することに少なくとも部分的によって、前記現在の時間ステップの動作のためのワーピングした再帰状態データを生成することと、
前記ワーピングした再帰状態データに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための前記再構成されたビデオフレームを生成することと、
をさらに備える、C 4 3に記載の方法。

〔C 4 8〕 前記デコーダ部分によって、再構成されたマスクを生成することと、前記再構成されたマスクは、前記前の時間ステップの動作中に生成された前記前に再構成されたビデオフレーム中のデータの存在を示す1つまたは複数の値を含み、
前記再構成されたマスクと、前記現在の時間ステップの動作のための前記ワーピングした再構成されたビデオフレームと、前記現在の時間ステップの動作のために決定された再構成された残差データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための前記再構成されたビデオフレームを生成することと、前記再構成されたマスクは

10

20

30

40

50

、前記ワーピングした再構成されたビデオフレームと前記再構成された残差データとの寄与を重み付けし、

をさらに備える、C 4 3に記載の方法。

〔C 4 9〕ニューラルネットワークシステムのエンコーダ部分によって、前記現在の時間ステップの動作のための入力ビデオフレームと、前記ニューラルネットワークシステムの前の時間ステップの動作からの再構成された動き推定データと、前記ニューラルネットワークシステムの前記前の時間ステップの動作からの再構成された残差データと、前記前の時間ステップの動作からの前記ニューラルネットワークシステムの前記デコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと、

前記ニューラルネットワークシステムの前記エンコーダ部分によって、前記現在の時間ステップの動作のための前記出力データを生成することと、前記出力データは、前記入力ビデオフレームと、前記前の時間ステップの動作からの前記再構成された動き推定データと、前記前の時間ステップの動作からの前記再構成された残差データと、前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて生成され、

をさらに備える、C 4 3に記載の方法。

〔C 5 0〕ビデオデータを処理するための装置であって、

メモリと、

前記メモリに結合されたプロセッサと、

を備え、前記プロセッサは、

ニューラルネットワークシステムのデコーダ部分を使用して、前記ニューラルネットワークシステムの現在の時間ステップの動作のための出力データと、前記ニューラルネットワークシステムの前の時間ステップの動作からの前記デコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと、

前記ニューラルネットワークシステムの前記デコーダ部分を使用して、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための動き推定データを決定することと、

前記前の時間ステップの動作中に生成された前に再構成されたビデオフレームを取得することと、

前記現在の時間ステップのために決定された前記動き推定データを使用して、前記前に再構成されたビデオフレームの1つまたは複数のピクセルを修正することに少なくとも部分的によって、前記現在の時間ステップの動作のためのワーピングした再構成されたビデオフレームを生成することと、

前記ワーピングした再構成されたビデオフレームに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成されたビデオフレームを生成することと、

を行うように構成された、装置。

〔C 5 1〕前記プロセッサは、

前記ニューラルネットワークシステムの前記デコーダ部分を使用して、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成された残差データを決定することと、前記現在の時間ステップのために決定された前記再構成された残差データに前記ワーピングした再構成されたビデオフレームを加算することに少なくとも部分的によって、前記現在の時間ステップの動作のための前記再構成されたビデオフレームを生成することと、

を行うように構成された、C 5 0に記載の装置。

〔C 5 2〕前記プロセッサは、

前記現在の時間ステップの動作中に前記デコーダ部分を使用して、前記現在の時間ステップの動作中に前記ニューラルネットワークシステムのエンコーダ部分によって処理された前の入力ビデオフレームに少なくとも部分的に基づいて、前記前の時間ステップの動作からの再構成された残差データを決定することと、

前記現在の時間ステップの動作中に、前記前の時間ステップの動作からの前記再構成された残差データを使用して、前記前に再構成されたビデオフレームを修正することと、

10

20

30

40

50

前記現在の時間ステップの動作のための最終の出力ビデオフレームとして、前記修正された前に再構成されたビデオフレームを出力することと、

を行うように構成された、C 5 1 に記載の装置。

〔C 5 3〕 前記プロセッサは、

前記ニューラルネットワークシステムの前記デコーダ部分を使用して、前記出力データと前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための再構成された残差データを決定することと、前記デコーダ部分のポストワーピングネットワークを使用して、前記現在の時間ステップのために決定された前記再構成された残差データと前記ワーピングした再構成されたビデオフレームとを処理することと、

10

前記ポストワーピングネットワークによる前記ワーピングした再構成されたビデオフレームと前記再構成された残差データとの処理に少なくとも部分的に基づいて、前記現在の時間ステップの動作のための前記再構成されたビデオフレームを生成することと、

を行うように構成された、C 5 0 に記載の装置。

〔C 5 4〕 前記プロセッサは、

前記現在の時間ステップの動作のために決定された前記動き推定データまたは前記前の時間ステップの動作のために決定された動き推定を使用して、前記前の時間ステップの動作からの前記再帰状態データを修正することによって、前記現在の時間ステップの動作のためのワーピングした再帰状態データを生成することと、

前記ワーピングした再帰状態データに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための前記再構成されたビデオフレームを生成することと、

20

を行うように構成された、C 5 0 に記載の装置。

〔C 5 5〕 前記プロセッサは、

前記デコーダ部分を使用して、再構成されたマスクを生成することと、前記再構成されたマスクは、前記前の時間ステップの動作中に生成された前記前に再構成されたビデオフレーム中のデータの存在を示す1つまたは複数の値を含み、

前記再構成されたマスクと、前記現在の時間ステップの動作のための前記ワーピングした再構成されたビデオフレームと、前記現在の時間ステップの動作のために決定された再構成された残差データとに少なくとも部分的に基づいて、前記現在の時間ステップの動作のための前記再構成されたビデオフレームを生成することと、前記再構成されたマスクは、前記ワーピングした再構成されたビデオフレームと前記再構成された残差データとの寄与を重み付けし、

30

を行うように構成された、C 5 0 に記載の装置。

〔C 5 6〕 前記プロセッサは、

ニューラルネットワークシステムのエンコーダ部分を使用して、前記現在の時間ステップの動作のための入力ビデオフレームと、前記ニューラルネットワークシステムの前の時間ステップの動作からの再構成された動き推定データと、前記ニューラルネットワークシステムの前記前の時間ステップの動作からの再構成された残差データと、前記前の時間ステップの動作からの前記ニューラルネットワークシステムの前記デコーダ部分の少なくとも1つの再帰レイヤからの再帰状態データとを取得することと、

40

前記ニューラルネットワークシステムの前記エンコーダ部分を使用して、前記現在の時間ステップの動作のための前記出力データを生成することと、前記出力データは、前記入力ビデオフレームと、前記前の時間ステップの動作からの前記再構成された動き推定データと、前記前の時間ステップの動作からの前記再構成された残差データと、前記前の時間ステップの動作からの前記再帰状態データとに少なくとも部分的に基づいて生成され、

を行うように構成された、C 5 0 に記載の装置。

〔C 5 7〕 前記プロセッサは、

前記再構成されたビデオフレームを前記メモリ中に記憶させることと、

を行うように構成された、C 5 0 に記載の装置。

〔C 5 8〕 前記プロセッサは、

50

前記再構成されたビデオフレームを表示させること、
を行うように構成された、C 5 0に記載の装置。

〔C 5 9〕 前記再構成されたビデオフレームを表示するように構成されたディスプレイ、
をさらに備える、C 5 0に記載の装置。

〔C 6 0〕 前記プロセッサはニューラルプロセッシングユニット（NPU）を含む、C
5 0に記載の装置。

〔C 6 1〕 前記装置はモバイルデバイスを備える、C 5 0に記載の装置。

〔C 6 2〕 前記装置はエクステンデッドリアリティデバイスを備える、C 5 0に記載の
装置。

〔C 6 3〕 前記装置はテレビジョンを備える、C 5 0に記載の装置。

〔C 6 4〕 前記装置は、1つまたは複数のビデオフレームをキャプチャするように構成
されたカメラを備える、C 5 0に記載の装置。

【図面】

【図 1】

【図 2 A】

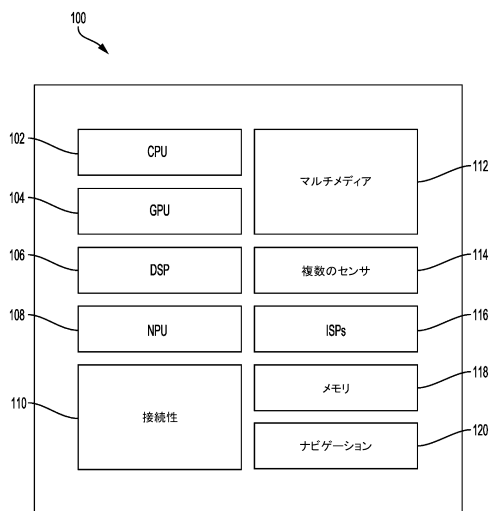


FIG. 1

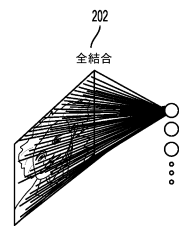


FIG. 2A

【図 2 B】

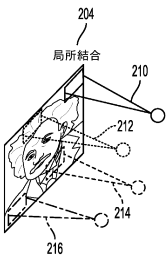


FIG. 2B

【図 2 C】

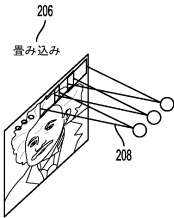


FIG. 2C

10

20

【図 2 D】

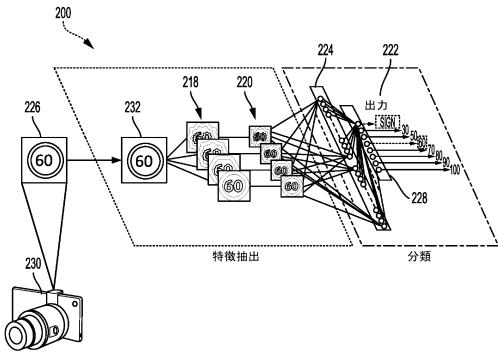


FIG. 2D

【図 3】

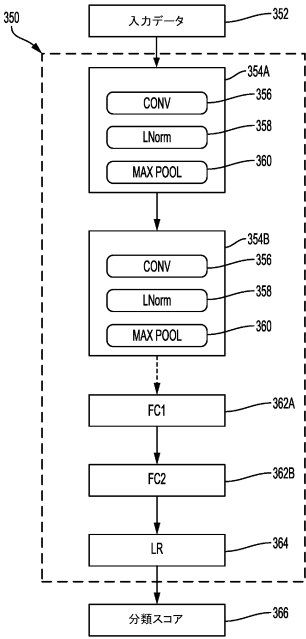


FIG. 3

30

40

50

【図 4】

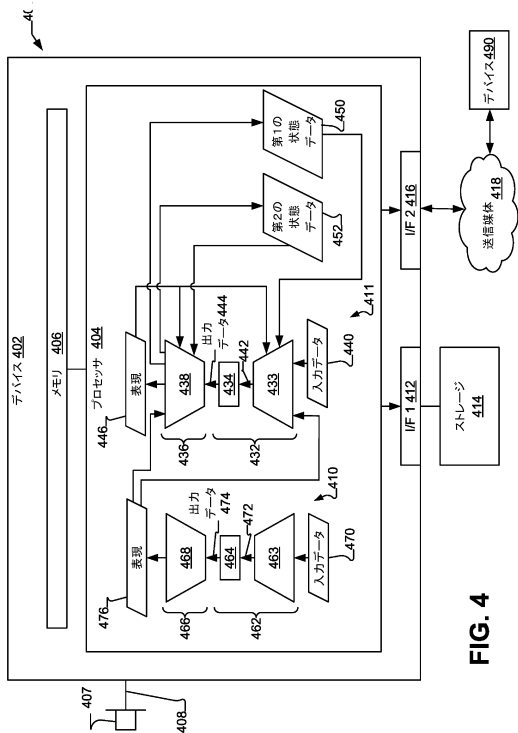


FIG. 4

【図 5 B】

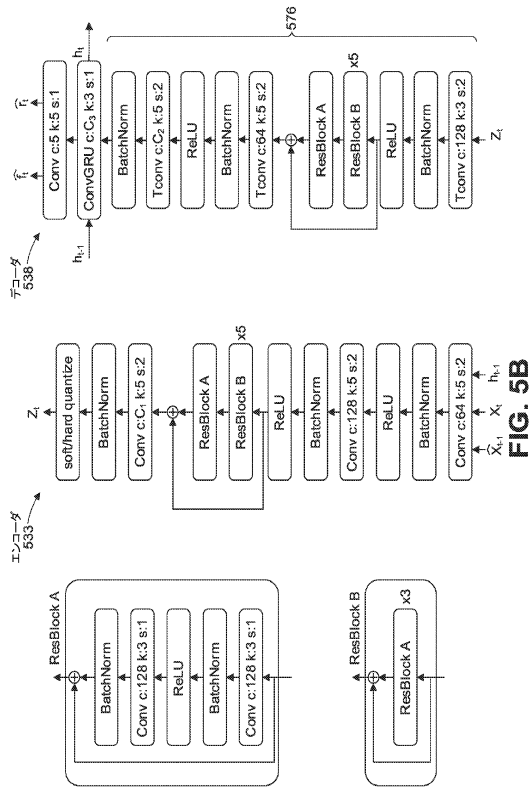


FIG. 5B

【図 5 A】

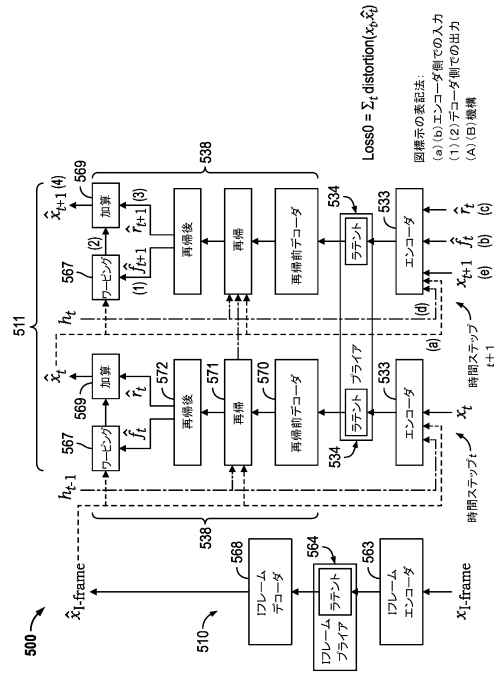


FIG. 5A

【図 6】

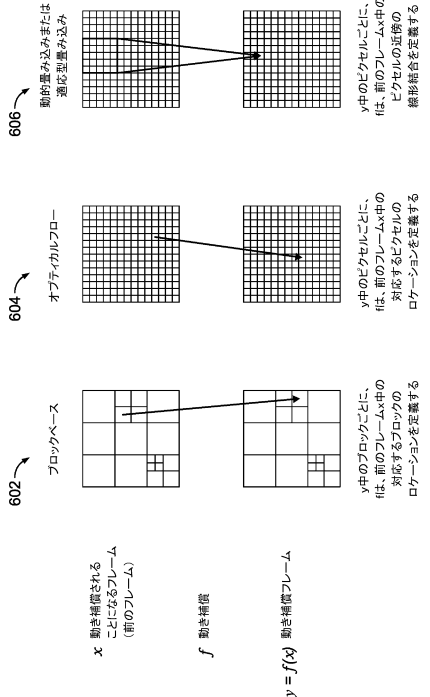


FIG. 6

【図 7】

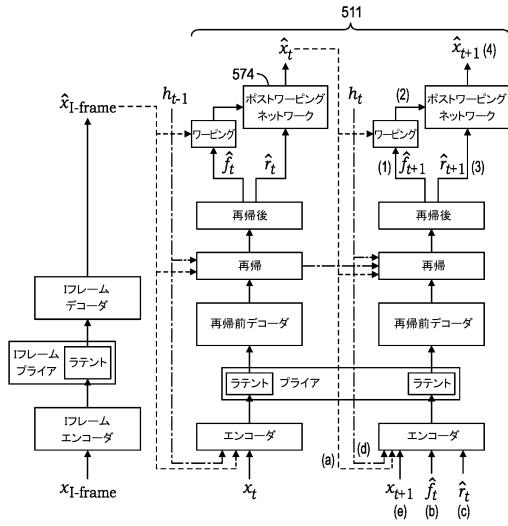


FIG. 7

【図 8】

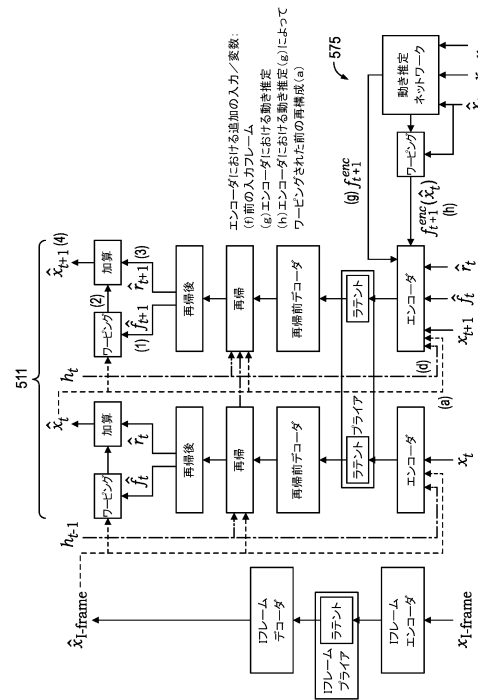


FIG. 8

【図 9】

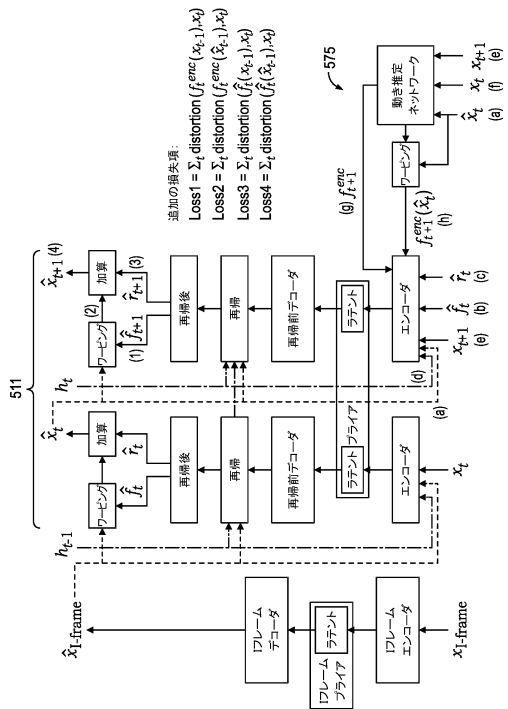


FIG. 9

【図 10】

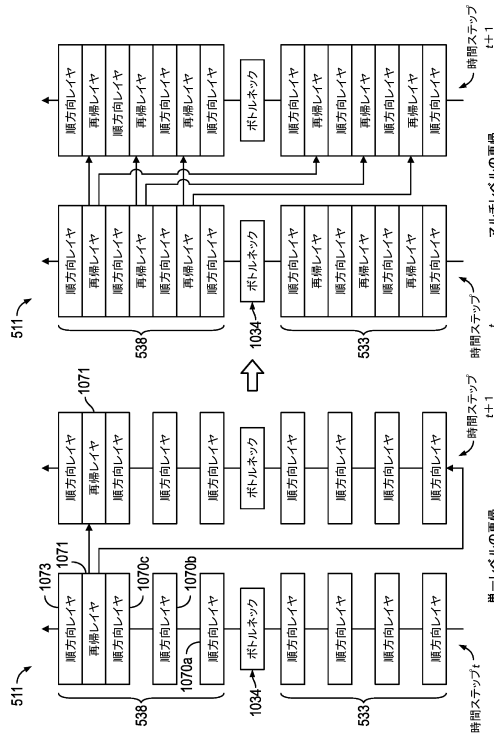


FIG. 10

10

20

30

40

50

【図 1 1】

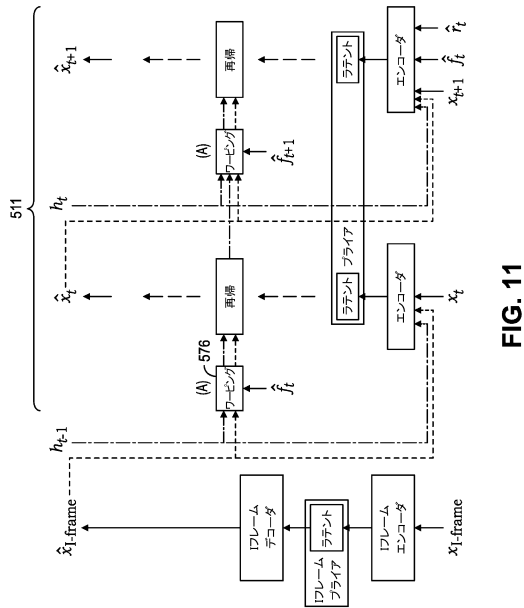


FIG. 11

【図 1 2】

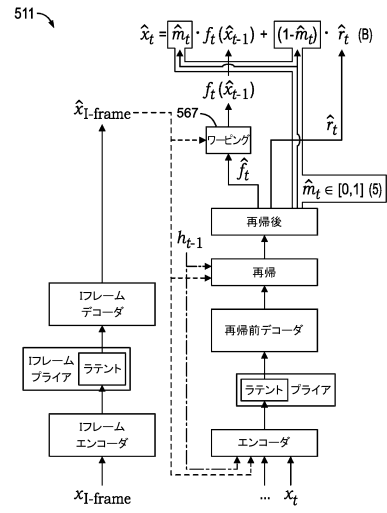


FIG. 12

【図 1 3】

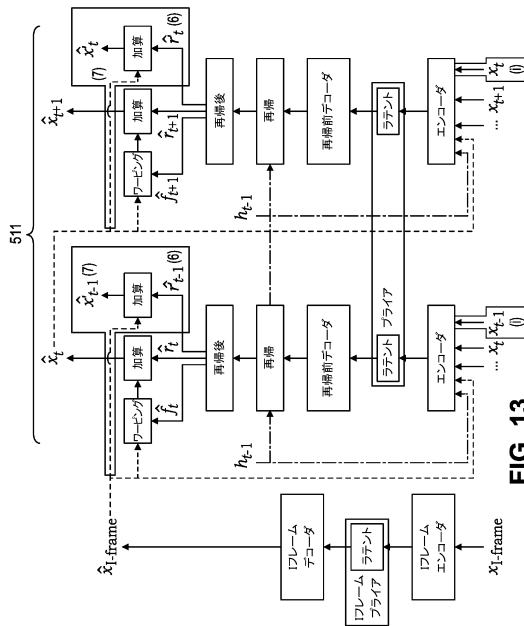


FIG. 13

【図 1 4】

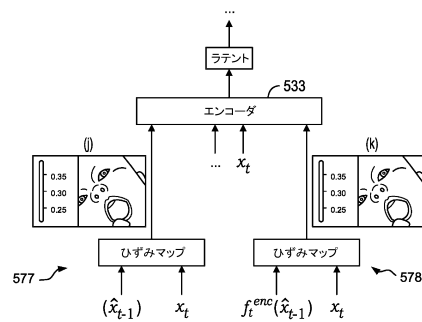


FIG. 14

10

20

30

40

50

【図 15】

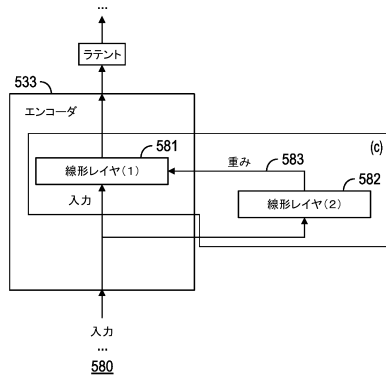


FIG. 15

【図 16】

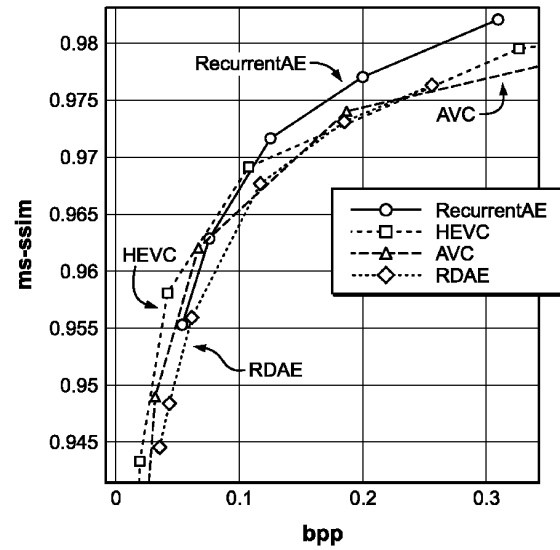


FIG. 16

【図 17】

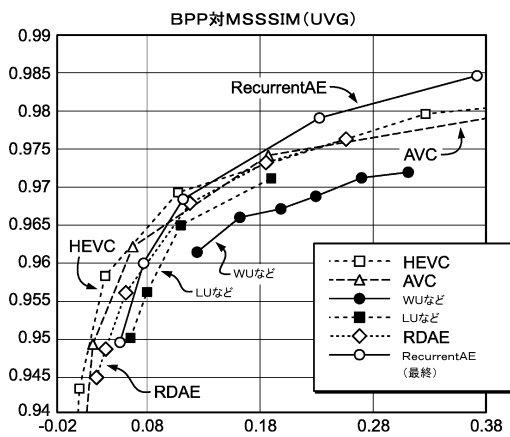


FIG. 17

【図 18】

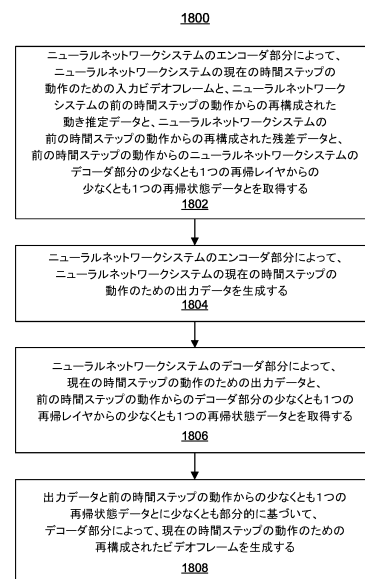


FIG. 18

【図 19】

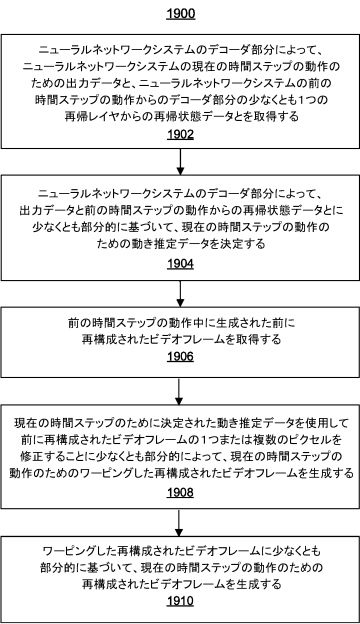


FIG. 19

【図 20】

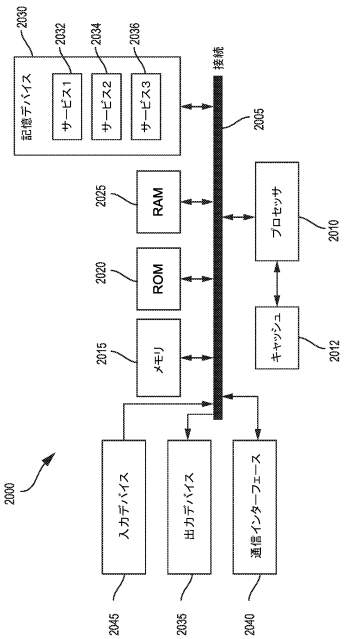


FIG. 20

フロントページの続き

(33)優先権主張国・地域又は機関

米国(US)

(72)発明者 ゴリンスキー、アダム・バルデマール

アメリカ合衆国、カリフォルニア州 9 2 1 2 1-1 7 1 4、サン・ディエゴ、モアハウス・ドライブ 5 7 7 5

(72)発明者 ヤン、ヤン

アメリカ合衆国、カリフォルニア州 9 2 1 2 1-1 7 1 4、サン・ディエゴ、モアハウス・ドライブ 5 7 7 5

(72)発明者 プレザ、レザ

アメリカ合衆国、カリフォルニア州 9 2 1 2 1-1 7 1 4、サン・ディエゴ、モアハウス・ドライブ 5 7 7 5

(72)発明者 ソティエール、ギヨーム・コンラッド

アメリカ合衆国、カリフォルニア州 9 2 1 2 1-1 7 1 4、サン・ディエゴ、モアハウス・ドライブ 5 7 7 5

(72)発明者 ファン・ロゼンダール、ティーズ・ジャン

アメリカ合衆国、カリフォルニア州 9 2 1 2 1-1 7 1 4、サン・ディエゴ、モアハウス・ドライブ 5 7 7 5

(72)発明者 コーエン、タコ・セバ스티アーン

アメリカ合衆国、カリフォルニア州 9 2 1 2 1-1 7 1 4、サン・ディエゴ、モアハウス・ドライブ 5 7 7 5

審査官 田中 崇大

(56)参考文献 米国特許出願公開第 2 0 2 1 / 0 0 4 4 8 0 4 (US, A1)

George Toderici et al., Full Resolution Image Compression with Recurrent Neural Networks, Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017年, pp.5435-5443

Guo Lu et al., DVC: An End-to-end Deep Video Compression Framework, Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019年, pp.10998-11007

(58)調査した分野 (Int.Cl., DB名)

H 0 4 N 1 9 / 0 0 - 1 9 / 9 8

G 0 6 N 3 / 0 4 5 5