

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7520118号
(P7520118)

(45)発行日 令和6年7月22日(2024.7.22)

(24)登録日 令和6年7月11日(2024.7.11)

(51)国際特許分類

F I

H O 4 N 19/593 (2014.01)

H O 4 N 19/593

請求項の数 8 (全43頁)

(21)出願番号	特願2022-533334(P2022-533334)	(73)特許権者	520353802
(86)(22)出願日	令和3年8月27日(2021.8.27)		テンセント・アメリカ・エルエルシー
(65)公表番号	特表2023-504817(P2023-504817 A)		アメリカ合衆国 9 4 3 0 6 カリフォルニア州 パロアルト パーク・ブルヴァード 2 7 4 7
(43)公表日	令和5年2月7日(2023.2.7)	(74)代理人	100107766
(86)国際出願番号	PCT/US2021/047996		弁理士 伊東 忠重
(87)国際公開番号	WO2022/072093	(74)代理人	100070150
(87)国際公開日	令和4年4月7日(2022.4.7)		弁理士 伊東 忠彦
審査請求日	令和4年6月3日(2022.6.3)	(74)代理人	100135079
(31)優先権主張番号	63/085,900		弁理士 宮崎 修
(32)優先日	令和2年9月30日(2020.9.30)	(72)発明者	ディン, ディン
(33)優先権主張国・地域又は機関	米国(US)		アメリカ合衆国 9 4 3 0 6 カリフォルニア州 パロアルト パーク・ブルバード 2 7 4 7 テンセント アメリカ エル
(31)優先権主張番号	63/085,908		
(32)優先日	令和2年9月30日(2020.9.30)		
	最終頁に続く		最終頁に続く

(54)【発明の名称】 ポスト・フィルタリングを用いるブロック・ワイズ・ニューラル画像圧縮のための方法及び装置

(57)【特許請求の範囲】

【請求項1】

ブロック・ワイズ・ニューラル画像の処理方法であって：

第1ニューラル・ネットワークを使用して、入力画像のブロックを符号化するステップであって、符号化されたブロックは、再構成されたブロックを生成するために、第2ニューラル・ネットワークを使用してデコーダによって復号化される、ステップ；

前記入力画像のブロックに関してより少ない損失をもたらすものを、前記再構成されたブロック及び生成され復元された予測ブロックの中から選択するステップ；

予測ブロックを生成するために、第3ニューラル・ネットワークを使用して、前記選択するステップにおいて選択されたブロックに関してイントラ予測を実行するステップ；

予測残差を生成するために、前記入力画像のブロックと生成された予測ブロックとの間の差分を決定するステップ；

第4ニューラル・ネットワークを使用して、前記差分を決定するステップにおいて生成された予測残差を符号化するステップであって、符号化された予測残差は、第5ニューラル・ネットワークを使用して前記デコーダにより復号化される、ステップ；及び

復元された予測ブロックを生成するために、復号化された予測残差を、前記生成された予測ブロックに加えるステップ；

を含み、前記方法は、更に：

複数の再構成されたブロックの各々から、アーチファクトを含む境界エリアを取り除くステップ；

1つ以上のニューラル・ネットワークを使用して、取り除かれた境界エリアに関してデブロッキングを実行して、前記アーチファクトを伴わないデブロック化されたエリアを生成するステップ；及び

前記複数の再構成されたブロックの各々を、生成されたデブロック化されたエリアでアップデートするステップ；

を含む、方法。

【請求項2】

請求項1に記載の方法において、更に：

前記再構成されたブロック及び復元された予測ブロックの中から選択されたものを示す選択信号を、前記再構成されたブロック及び復元された予測ブロックの中から選択されたものに対応する、前記符号化されたブロック及び符号化された予測残差のうちのもの、とともに前記デコーダへ送信するステップ；

を含み、前記符号化されたブロック及び符号化された予測残差のうちの送信されたものは、送信された選択信号に基づいて前記デコーダにより復号化される、方法。

【請求項3】

請求項1に記載の方法において、前記第1ニューラル・ネットワークによるブロックの符号化及び前記第2ニューラル・ネットワークによるブロックの復号化は、変分オートエンコーダ（VAE）構造に基づいて行われ、前記再構成されたブロックに関してイントラ予測を実行するステップは：

複数の再構成されたブロックを1つ以上のグループに、前記複数の再構成されたブロックの位置に基づいてグループ化するステップ；

前記VAE構造における1つ以上の潜在空間を生成するために、前記1つ以上のグループの各々に関して個々の変換を実行するステップ；

前記個々の変換を実行するステップにおいて生成された1つ以上の潜在空間を統合するステップ；及び

前記予測ブロックを生成するために、前記統合するステップにおいて統合された1つ以上の潜在空間に関して前記イントラ予測を実行するステップ；

を含む、方法。

【請求項4】

請求項1に記載の方法において、更に：

前記複数の再構成されたブロックの各々から、取り除かれた境界エリアを伴わない残りのブロック・エリアを取り除くステップ；

エンハンスド・ブロック・エリアを生成するために、少なくとも1つのニューラル・ネットワークを使用して、取り除かれた残りのブロック・エリアに関してポスト・エンハンスメントを実行するステップ；及び

前記複数の再構成されたブロックの各々を、生成されたエンハンスド・ブロック・エリアでアップデートするステップ；

を含む、方法。

【請求項5】

請求項1に記載の方法において、前記第1ニューラル・ネットワークによるブロックの符号化及び前記第2ニューラル・ネットワークによるブロックの復号化は、変分オートエンコーダ（VAE）構造に基づいて行われ、前記再構成されたブロックに関してイントラ予測を実行するステップは：

前記再構成されたブロックの上にある再構成されたピクセルの第1コンテキスト・エリアと、前記再構成されたブロックの左にある再構成されたピクセルの第2コンテキスト・エリアとを取得するステップ；

前記VAE構造における潜在表現を生成するために、取得された前記第1コンテキスト・エリアと取得された前記第2コンテキスト・エリアの各々に関して個々の変換を実行するステップ；

前記個々の変換を実行するステップにおいて生成された潜在表現を統合するステップ；

10

20

30

40

50

及び

前記予測ブロックを生成するために、前記統合するステップにおいて統合された潜在表現に関して前記イントラ予測を実行するステップ；

を含む、方法。

【請求項6】

請求項1に記載の方法において、前記第1ニューラル・ネットワークによるブロックの符号化及び前記第2ニューラル・ネットワークによるブロックの復号化は、変分オートエンコーダ（VAE）構造に基づいて行われ、前記再構成されたブロックに関してイントラ予測を実行するステップは：

前記再構成されたブロックの上にある再構成されたピクセルの第1コンテキスト・エリアと、前記再構成されたブロックの左にある再構成されたピクセルの第2コンテキスト・エリアと、前記第1コンテキスト・エリアの下及び前記第2コンテキスト・エリアの右にあるパディング・エリアとを取得するステップ；

矩形エリアを取得するために、取得された前記第1コンテキスト・エリアと、取得された前記第2コンテキスト・エリアと、取得された前記パディング・エリアを統合するステップ；

前記VAE構造における潜在表現を生成するために、取得された前記矩形エリアに関して変換を実行するステップ；及び

前記予測ブロックを生成するために、前記変換を実行するステップにおいて生成された潜在表現に関して前記イントラ予測を実行するステップ；

を含む、方法。

【請求項7】

ブロック・サイズ・ニューラル画像処理のための装置であって、前記装置はエンコーダとして実装され、前記装置は：

プログラム・コードを記憶するように構成された少なくとも1つのメモリ；及び

前記プログラム・コードを読み込み、前記プログラム・コードにより指示されるとおりに動作するように構成された少なくとも1つのプロセッサ；

を含み、前記プログラム・コードは、請求項1-6のうちの何れか一項に記載の方法を前記少なくとも1つのプロセッサに実行させる、装置。

【請求項8】

請求項1-6のうちの何れか一項に記載の方法を少なくとも1つのプロセッサに実行させるコンピュータ・プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願の相互参照

本件出願は、米国特許商標庁に2020年9月30日付で出願された米国仮特許出願第63/085,900号、2020年9月30日付で出願された米国仮特許出願第63/085,908号、2020年10月7日付で出願された米国仮特許出願第63/088,675号、及び2021年6月29日付で出願された米国特許出願第17/362,003号を基礎とし且つ優先権を主張しており、それらの開示全体が参照により本件に援用される。

【背景技術】

【0002】

従来、オリジナル画像全体を直接的に符号化する代わりに、ブロック・ベースのイントラ予測&残差コーディング機構は、高効率ビデオ・コーディング(High Efficiency Video Coding, HEVC)や汎用ビデオ・コーディング(Versatile Video Coding, VVC)規格のような現代のビデオ・コーディング・システムにおいて、画像フレームを圧縮することに非常に効果的であることが実証されている。画像全体は先ずコーディング・ツリー・ユニット(CTU)にパーティション化される。各CTUに関し、それは、種々のサイズのブロックにパーティション化される可能性があり、予測ブロックは、種々の角度方向に沿って

、以前の圧縮されたブロックの境界ピクセルをコピーすることによって生成される。次いで、オリジナル・ブロックと予測ブロックの間の残差が圧縮される。残差は、オリジナル・ピクセルよりもはるかに効率的に符号化することが可能であり、より良いコーディング・パフォーマンスを達成することができる。

【0003】

ISO/IEC MPEG (JTC 1/SC 29/WG 11)は、将来のビデオ・コーディング技術の標準化に関する潜在的なニーズを積極的に探求している。ISO/IEC JPEGは、ディープ・ニューラル・ネットワーク(Deep Neural Networks, DNN)を用いるAIベースのエンド・ツー・エンド・ニューラル画像圧縮に焦点を当てたJPEG - AIグループを設立している。また、中国のAVS規格も、ニューラル画像及びビデオ圧縮技術に取り組むためにAVS-AIスペシャル・グループを結成している。最近のアプローチの成功により、高度なニューラル画像及びビデオ圧縮方法論に産業界の関心が益々高まっている。

10

【0004】

入力画像 x が与えられた場合に、ニューラル画像圧縮(Neural Image Compression, NIC)のターゲットは、画像 x をDNNエンコーダへの入力として使用して、記憶及び伝送のためにコンパクトにした圧縮された表現

【数1】

$$\hat{x}$$

20

(以下、混乱が生じない限り、 x ハット又は $x^\wedge$ と言及する)を計算し、次いで、 $x^\wedge$ をDNNデコーダ入力として使用して、画像

【数2】

$$\bar{x}$$

(以下、混乱が生じない限り、 x バー又は x^- と言及する)を再構築することである。従来のNIC法は、変分オートエンコーダ(variational autoencoder, VAE)構造を採用しており、DNNエンコーダは、画像全体 x をその入力として直接的に使用し、その入力は、出力表現 $x^\wedge$ を計算するブラック・ボックスのように機能する一組のネットワーク層を通過する。それに対応して、DNNデコーダは、全体表現 $x^\wedge$ をその入力として採用し、その入力は、再構成された x^- を計算する別のブラック・ボックスのように機能する別の一組のネットワーク層を通過する。

30

【0005】

HEVCでは、デブロッキング・フィルタ・プロセスが、復号化プロセスと同じ順序で各コーディング・ユニット(CU)に対して実行される。まず垂直エッジがフィルタリング(水平フィルタリング)され、次いで水平エッジがフィルタリング(垂直フィルタリング)される。フィルタリングは、ルマ及びクロマ成分の両方に対してフィルタリングされるように決定された 8×8 ブロック境界に適用される。 4×4 ブロック境界は、複雑さを減らすために処理されない。

40

【0006】

境界強度(boundary strength, Bs)は、どの程度強いフィルタリング・プロセスが境界に必要とされるのかを反映することができる。Bsに関して2という値は強いフィルタリングを示し、1は弱いフィルタリングを示し、0はデブロッキング・フィルタリングを行わないことを示す。

【0007】

PとQを、フィルタリングに関与するブロックとして定義し、ここで、Pは境界の左側(垂直エッジの場合)又は上方(水平エッジの場合)に位置するブロックを表し、Qは境界の右側(垂直エッジの場合)又は下方(水平エッジの場合)に位置するブロックを表す。図1は、イントラ・コーディング・モード、非ゼロ変換係数の存在、参照ピクチャ、動きベクトル数、及

50

び動きベクトル差分に基づいて、Bs値がどのように計算されるかを示す。例えば、MV差分が閾値Tを下回る場合、Bsは0に等しい。閾値Tは1ピクセルに設定される。HEVCでは、MV精度は1/4ピクセルであり、MV差分閾値は4に設定されている。VTMでは、MV精度は1/16であり、MV差分は16に設定されている。

【0008】

Bsは4×4ブロック・ベースで計算されるが、8×8グリッドに再マッピングされる。4×4グリッド内のラインで構成される8ピクセルに対応するBsの2つの値のうちの最大値が、8×8グリッドの境界に対するBsとして選択される。

【0009】

VTM5では、デブロッキング・フィルタリング・プロセスはHEVCにおけるものとほとんど同じである。ただし、以下の変更が加えられる。

10

1) 再構成されたサンプルの平均ルマ・レベルに依存するデブロッキング・フィルタのフィルタ強度

2) デブロッキングtCテーブルの拡張

3) ルマ用のより強力なデブロッキング・フィルタ

4) クロマ用のより強力なデブロッキング・フィルタ

5) 4×4サンプル・グリッドに関するルマ・デブロッキング、及び8×8サンプル・グリッドに関するクロマ・デブロッキング

【0010】

HEVCでは、デブロッキング・フィルタのフィルタ強度は、平均化された量子化パラメータqPLから導出される変数βとtCによって制御される。VTM5では、デブロッキング・フィルタは、再構成されたサンプルのルマ・レベルに従って、qPLにオフセットを加えることによって、デブロッキング・フィルタの強度を制御する。再構成されたルマ・レベルLLは、次のように導出される：

20

【0011】

$LL = ((p_{0,0} + p_{0,3} + q_{0,0} + q_{0,3}) \gg 2) / (1 \ll \text{bitDepth})$ (Eq. 1),

【0012】

ここで、サンプル値 $p_{i,k}$ 及び $q_{i,k}$ は、($i = 0..3, k = 0$ 及び 3)は、図2に示されているように導出される。

【0013】

30

変数qPLは、次のように導出される：

【0014】

$qPL = ((QpQ + QpP + 1) \gg 1) + qpOffset$ (Eq. 2),

【0015】

ここで、QpQ及びQpPはそれぞれサンプル $q_{0,0}$ 及び $p_{0,0}$ を含むコーディング・ユニットの量子化パラメータを示す。伝達関数に依存するオフセットqpOffset、その値はSPSでシグナリングされる。

【0016】

VTM5では、Maximum QPは、51から63へ変更されており、ブロックQPに基づいてデブロッキング・パラメータtCの値を導出するデブロッキング・テーブルに対する相応の変化を反映することが望ましい。以下は、QP範囲の拡張に対処するためにアップデートされたtCテーブルである。

40

【0017】

$tC = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 5, 5, 6, 6, 7, 8, 9, 10, 11, 13, 14, 16, 18, 20, 22, 25, 28, 31, 35, 39, 44, 50, 56, 63, 70, 79, 88, 99]$ (Eq. 3).

【0018】

バイリニア・フィルタ(より強力なデブロッキング・フィルタ)は、境界の片側のサンプルがラージ・ブロックに属する場合に使用される。ラージ・ブロックに属するサンプルは、

50

垂直エッジについて幅が32以上であり、水平エッジについて高さが32以上である場合として定義される。ブロック境界サンプル p_i ($i=0$ ないし S_p-1)及び q_i ($j=0$ ないし S_q-1)は、次のように線形補間で置換される：

【0 0 1 9】

-

【0 0 2 0】

【数 3】

$p_i' = (f_i * Middle_{s,t} + (64 - f_i) * P_s + 32) \gg 6$, $p_i \pm tcPD_i$ にクリップされる (Eq. 4); 及び
 $q_j' = (g_j * Middle_{s,t} + (64 - g_j) * Q_s + 32) \gg 6$, $q_j \pm tcPD_j$ にクリップされる (Eq. 5),

10

【0 0 2 1】

ここで、 $tcPD_i$ 及び $tcPD_j$ の項は位置依存クリッピングであり、 g_j , f_i , $Middle_{s,t}$, P_s 及び Q_s は以下の通りである：

テーブル1 - ルマに対する強デブロッキング・パラメータの導出

【数 4】

20

30

40

50

Sp, Sq 7, 7 (p側: 7, q側: 7)	$f_i = 59 - i * 9$, 次のようにも記述できる: $f = \{59, 50, 41, 32, 23, 14, 5\}$ $g_j = 59 - j * 9$, 次のようにも記述できる: $g = \{59, 50, 41, 32, 23, 14, 5\}$ $Middle_{7,7} = (2 * (p_0 + q_0) + p_1 + q_1 + p_2 + q_2 + p_3 + q_3 + p_4 + q_4 + p_5 + q_5 + p_6 + q_6 + 8) \gg 4$ $P_7 = (p_6 + p_7 + 1) \gg 1, \quad Q_7 = (q_6 + q_7 + 1) \gg 1$
7, 3 (p側: 7 q側: 3)	$f_i = 59 - i * 9$, 次のようにも記述できる: $f = \{59, 50, 41, 32, 23, 14, 5\}$ $g_j = 53 - j * 21$, 次のようにも記述できる: $g = \{53, 32, 11\}$ $Middle_{7,3} = (2 * (p_0 + q_0) + q_0 + 2 * (q_1 + q_2) + p_1 + q_1 + p_2 + p_3 + p_4 + p_5 + p_6 + 8) \gg 4$ $P_7 = (p_6 + p_7 + 1) \gg 1, \quad Q_3 = (q_2 + q_3 + 1) \gg 1$
3, 7 (p側: 3 q側: 7)	$g_j = 59 - j * 9$, 次のようにも記述できる: $g = \{59, 50, 41, 32, 23, 14, 5\}$ $f_i = 53 - i * 21$, 次のようにも記述できる: $f = \{53, 32, 11\}$ $Middle_{3,7} = (2 * (q_0 + p_0) + p_0 + 2 * (p_1 + p_2) + q_1 + p_1 + q_2 + q_3 + q_4 + q_5 + q_6 + 8) \gg 4$ $Q_7 = (q_6 + q_7 + 1) \gg 1, \quad P_3 = (p_2 + p_3 + 1) \gg 1$
7, 5 (p側: 7 q側: 5)	$g_j = 58 - j * 13$, 次のようにも記述できる: $g = \{58, 45, 32, 19, 6\}$ $f_i = 59 - i * 9$, 次のようにも記述できる: $f = \{59, 50, 41, 32, 23, 14, 5\}$ $Middle_{7,5} = (2 * (p_0 + q_0 + p_1 + q_1) + q_2 + p_2 + q_3 + p_3 + q_4 + p_4 + q_5 + p_5 + 8) \gg 4$ $Q_5 = (q_4 + q_5 + 1) \gg 1, \quad P_7 = (p_6 + p_7 + 1) \gg 1$
5, 7 (p側: 5 q側: 7)	$g_j = 59 - j * 9$, 次のようにも記述できる: $g = \{59, 50, 41, 32, 23, 14, 5\}$ $f_i = 58 - i * 13$, 次のようにも記述できる: $f = \{58, 45, 32, 19, 6\}$ $Middle_{5,7} = (2 * (q_0 + p_0 + p_1 + q_1) + q_2 + p_2 + q_3 + p_3 + q_4 + p_4 + q_5 + p_5 + 8) \gg 4$ $Q_7 = (q_6 + q_7 + 1) \gg 1, \quad P_5 = (p_4 + p_5 + 1) \gg 1$
5, 5 (p側: 5 q側: 5)	$g_j = 58 - j * 13$, 次のようにも記述できる: $g = \{58, 45, 32, 19, 6\}$ $f_i = 58 - i * 13$, 次のようにも記述できる: $f = \{58, 45, 32, 19, 6\}$ $Middle_{5,5} = (2 * (q_0 + p_0 + p_1 + q_1 + q_2 + p_2) + q_3 + p_3 + q_4 + p_4 + 8) \gg 4$ $Q_5 = (q_4 + q_5 + 1) \gg 1, \quad P_5 = (p_4 + p_5 + 1) \gg 1$
5, 3 (p側: 5 q側: 3)	$g_j = 53 - j * 21$, 次のようにも記述できる: $g = \{53, 32, 11\}$ $f_i = 58 - i * 13$, 次のようにも記述できる: $f = \{58, 45, 32, 19, 6\}$ $Middle_{5,3} = (q_0 + p_0 + p_1 + q_1 + q_2 + p_2 + q_3 + p_3 + 4) \gg 3$ $Q_3 = (q_2 + q_3 + 1) \gg 1, \quad P_5 = (p_4 + p_5 + 1) \gg 1$
3, 5 (p側: 3 q側: 5)	$g_j = 58 - j * 13$, 次のようにも記述できる: $g = \{58, 45, 32, 19, 6\}$ $f_i = 53 - i * 21$, 次のようにも記述できる: $f = \{53, 32, 11\}$ $Middle_{3,5} = (q_0 + p_0 + p_1 + q_1 + q_2 + p_2 + q_3 + p_3 + 4) \gg 3$ $Q_5 = (q_4 + q_5 + 1) \gg 1, \quad P_3 = (p_2 + p_3 + 1) \gg 1$

10

20

30

【0022】

上記のより強いルマ・フィルタは、Condition1(条件1)、Condition2(条件2)、Condition 3(条件3)の全てがTRUEである場合にのみ使用される。条件1は“ラージ・ブロック条件”である。この条件は、P側とQ側のサンプルがラージ・ブロックに属しているかどうかを検出する。条件2と条件3は以下により決定される：

40

【0023】

Condition2 = (d < β) ? TRUE: FALSE (Eq. 6); 及び

【0024】

Condition3 = StrongFilterCondition = (dpq is less than (β > 2), sp3 + sq3 が(3 * β > 5)より小さい, 且つAbs(p0 - q0)が(5 * tC + 1) > 1)より小さい ? TRUE : FALSE (Eq. 7).

【0025】

クロマに対する強デブロック・フィルタは以下のように定義される：

【0026】

50

$p2' = (3 \cdot p3 + 2 \cdot p2 + p1 + p0 + q0 + 4) \gg 3$ (Eq. 8);

【0027】

$p1' = (2 \cdot p3 + p2 + 2 \cdot p1 + p0 + q0 + q1 + 4) \gg 3$ (Eq. 9); 及び

【0028】

$p0' = (p3 + p2 + p1 + 2 \cdot p0 + q0 + q1 + q2 + 4) \gg 3$ (Eq. 10).

【0029】

上記のクロマ・フィルタは、 8×8 クロマ・サンプル・グリッドでデブロッキングを実行する。クロマ強フィルタは、ブロック境界の両側で使用される。ここで、クロマ・フィルタは、クロマ・エッジの両側が（クロマ・サンプル単位で）8以上であり、続く3条件判定が満たされる場合に選択される。第1のものはラージ・ブロックに加えて境界強度の判定に関するものである。第2及び第3のものは、基本的にはHEVCルマ判定と同じであり、これらはそれぞれon/off判定及び強フィルタ判定である。第1判定では、bSが、テーブル2に示すようにクロマ・フィルタリングに関して修正される。テーブル1の条件が順次チェックされる。条件が満たされると、より低い優先度の残りの条件はスキップされる。

テーブル2 — 修正される境界強度

【数5】

優先度	条件	Y	U	V
5	隣接するブロックの少なくとも1つはイントラである。	2	2	2
4	隣接するブロックの少なくとも1つは非ゼロ変換係数を有する。	1	1	1
3	隣接するブロックに属する動きベクトル間の絶対差分は整数ルマ・サンプル1つ以上である。	1	N/A	N/A
2	ベクトルを参照する隣接するブロックの動き予測は異なる。	1	N/A	N/A
1	それ以外	0	0	0

【0030】

クロマ・デブロッキングは、bSが2に等しい場合、又はラージ・ブロック境界が検出された場合にはbSが1に等しい場合に、実行される。第2及び第3条件は、基本的にはHEVCルマ強フィルタ決定と同じである。

【0031】

現在のVVCワーキング・ドラフトでは、デブロッキング・フィルタは、ルマの場合は 4×4 グリッド、クロマの場合は 8×8 グリッドでイネーブルにされる。

デブロッキング・フィルタリング・プロセスは、サブブロック境界に加えてCU境界にも適用される。サブブロック境界は、STMVP及びアフィン・モードによって導入される予測ユニット境界や、SBT及びISPモードによって導入される変換ユニット境界を含む。

【0032】

SBT及びISPサブブロックに対しては、HEVCデブロッキング・フィルタでのTUにおける同じロジックが適用される。エッジを横切るどちらのサブブロックにも非ゼロ係数が存在する場合に、デブロッキング・フィルタがTU境界に適用される。

【0033】

4×4 グリッド上のSbTMVP及びアフィン・サブブロックについては、HEVCデブロッキング・フィルタでのPUにおける同じロジックが適用される。PU境界については、動きベクトルと隣接サブブロックの参照ピクチャとの間の差を考慮して、デブロッキング・フィルタが適用される。

【0034】

既存のネットワーク・アーキテクチャを利用し、より高いコーディング効率を達成できることが可能な、より効率的なNIC技術を開発する必要があるかもしれない。

【0035】

イントラ予測は、HEVC及びVVCのような画像及びビデオ圧縮アルゴリズムにおいて広く

使用されている重要な機構である。予測のゴールは、以前に再構成された周囲のピクセルから、ピクセルのブロックを推論することである。オリジナル・ピクセルと予測されたピクセルとの間で計算された残差を符号化/復号することにより、圧縮方法はより効率的になり得る。

【0036】

従来のビデオ・コーデック(HEVC, VVC)におけるイントラ予測は、予測されるブロックとその近傍との間の依存性を当てにしている。例えば、VVCでは、レート歪・基準に従って、予測方法は、35の固定されたシンプルな予測関数の中から1つのモードに従って選択し、指定された方向に沿ってピクセル値を伝搬させる。周囲のピクセルから因果関係によって単純にピクセルを伝搬させる代わりに、以前の提案では、予測されるブロックに対して、画像内で最良に一致するものを発見するための探索方法が導入された。更なる研究では、近傍のk個の最も近いパッチの線形結合として定義される、画像内の自己類似性(self-similarities)を活用している。

10

【0037】

近年、リカレント・ニューラル・ネットワークは、画像ピクセルの予測について信頼性の高いモデルを学習する能力を実証している。例えば、リカレント・ニューラル・ネットワークは、画像の既知の領域内のピクセル間の依存性のそれらの内部表現を逐次的に更新し、次いで、画像の未知の領域内の次のピクセルを生成する。

【0038】

また、イントラ予測のための全結合(fully connected, FC)層ベースの方法も提案されている。FCニューラル・ネットワークは、隣接する再構成ピクセルから現在のブロックへのエンド・ツー・エンド・マッピングを学習することができる。これは、コンテキスト・エリア(モデルへの入力)として、ピクセルの複数の参照ラインを使用する。しかしながら、この方法は非常に小数のピクセルしか予測することができない。幾つかの以前の研究では、イントラ予測のための畳み込み層ベースの方法が提案された。これは隣接する再構成ピクセルを取り込み、それを予測のために畳み込みニューラル・ネットワーク(convolutional neural network, CNN)ネットワークへ送る。

20

【0039】

現在の予測方法は、ビデオ・イントラ予測モジュール用に設計されており、ニューラル・ベースの画像圧縮に直接的に使用することはできない。更に、現在のニューラル・ベースの予測方法は、再構成されたピクセル(コンテキスト・エリア)を2つの分離されたエリアにパーティション化するが、以下の欠点を招く可能性がある：1) それらは2つの分離されたコンテキスト・エリア間の境界エリアを学習することができない；2) 2つのコンテキスト・エリア(又はこれら2つのコンテキスト・エリアによって変換された潜在表現)を統合するために、予測方法は、演算時間に加えて、膨大な量の演算リソースを必要とする可能性がある。

30

【発明の概要】

【0040】

実施形態によれば、ポスト・フィルタリングを用いるブロック・ワイズ・ニューラル画像圧縮の方法は、エンコーダの少なくとも1つのプロセッサにより実行されるものであり、第1ニューラル・ネットワークを使用して、入力画像のブロックを符号化するステップであって、符号化されたブロックは、再構成されたブロックを生成するために、第2ニューラル・ネットワークを使用してデコーダによって復号化される、ステップ；及び予測ブロックを生成するために、第3ニューラル・ネットワークを使用して、前記再構成されたブロックに関してイントラ予測を実行するステップを含む。方法は、更に、予測残差を生成するために、前記入力画像のブロックと生成された予測ブロックとの間の差分を決定するステップ；第4ニューラル・ネットワークを使用して、生成された予測残差を符号化するステップであって、符号化された予測残差は、第5ニューラル・ネットワークを使用して前記デコーダにより復号化される、ステップ；及び復元された予測ブロックを生成するために、復号化された予測残差を、前記生成された予測ブロックに加えるステップを含む。

40

50

【0041】

別の実施形態によれば、ポスト・フィルタリングを用いるブロック・ワイズ・ニューラル画像圧縮のための装置は、エンコーダとして実装されるものであり、プログラム・コードを記憶するように構成された少なくとも1つのメモリ；及び前記プログラム・コードを読み込み、前記プログラム・コードにより指示されるとおりに動作するように構成された少なくとも1つのプロセッサを含む。前記プログラム・コードは：第1ニューラル・ネットワークを使用して、入力画像のブロックを符号化することを、前記少なくとも1つのプロセッサに行わせるように構成された第1符号化コードであって、符号化されたブロックは、再構成されたブロックを生成するために、第2ニューラル・ネットワークを使用してデコーダによって復号化される、第1符号化コード；及び予測ブロックを生成するために、第3ニューラル・ネットワークを使用して、前記再構成されたブロックに関してイントラ予測を実行することを、前記少なくとも1つのプロセッサに行わせるように構成された第1実行コードを含む。プログラム・コードは、更に、予測残差を生成するために、前記入力画像のブロックと生成された予測ブロックとの間の差分を決定することを、前記少なくとも1つのプロセッサに行わせるように構成された決定コード；第4ニューラル・ネットワークを使用して、生成された予測残差を符号化することを、前記少なくとも1つのプロセッサに行わせるように構成された第2符号化コードであって、符号化された予測残差は、第5ニューラル・ネットワークを使用して前記デコーダにより復号化される、第2符号化コード；及び復元された予測ブロックを生成するために、復号化された予測残差を、前記生成された予測ブロックに加えることを、前記少なくとも1つのプロセッサに行わせるように構成された加算コードを含む。

10

20

【0042】

実施形態によれば、命令を記憶する非一時的なコンピュータ読み取り可能な媒体において、前記命令は、エンコーダのポスト・フィルタリングを用いるブロック・ワイズ・ニューラル画像圧縮のために、少なくとも1つのプロセッサにより実行されると、前記少なくとも1つのプロセッサに：第1ニューラル・ネットワークを使用して、入力画像のブロックを符号化するステップであって、符号化されたブロックは、再構成されたブロックを生成するために、第2ニューラル・ネットワークを使用してデコーダによって復号化される、ステップ；予測ブロックを生成するために、第3ニューラル・ネットワークを使用して、前記再構成されたブロックに関してイントラ予測を実行するステップを行わせる。命令は、少なくとも1つのプロセッサにより実行されると、少なくとも1つのプロセッサに、更に：予測残差を生成するために、前記入力画像のブロックと生成された予測ブロックとの間の差分を決定するステップ；第4ニューラル・ネットワークを使用して、生成された予測残差を符号化するステップであって、符号化された予測残差は、第5ニューラル・ネットワークを使用して前記デコーダにより復号化される、ステップ；及び復元された予測ブロックを生成するために、復号化された予測残差を、前記生成された予測ブロックに加えるステップを行わせる。

30

【発明を実施するための形態】

【0043】

図1は、境界強度(Bs)を計算するフロー図である。

40

【0044】

図2は、サンプル位置の図である。

【0045】

図3は、本件で説明される方法、装置及びシステムが実施形態に従って実施されることが可能な環境の図である。

【0046】

図4は、図3の1つ以上のデバイスの例示的なコンポーネントのブロック図である。

【0047】

図5Aは、実施形態によるブロック・ワイズ・ニューラル画像圧縮のためのエンコーダのブロック図である。

50

【0048】

図5Bは、実施形態による図5Aのエンコーダのイントラ予測コンポーネントのブロック図である。

【0049】

図6は、実施形態によるブロック・ワイズ・ニューラル画像圧縮のためのデコーダのブロック図である。

【0050】

図7は、実施形態による後処理コンポーネントのブロック図である。

【0051】

図8は、図7の後処理コンポーネントのデブロッキング・コンポーネントの図である。

10

【0052】

図9は、実施形態による境界エリアを取り出す方法を示す図である。

【0053】

図10は、実施形態によるマルチ・モデルのデブロッキング・コンポーネントの図である。

【0054】

図11は、図7の後処理コンポーネントのポスト・エンハンスメント・コンポーネントの図である。

【0055】

図12は、実施形態によるマルチ・モデルのポスト・エンハンスメント・コンポーネントの図である。

20

【0056】

図13は、実施形態によるポスト・エンハンスメント・コンポーネントの図である。

【0057】

図14は、実施形態によるポスト・エンハンスメント&デブロッキング・コンポーネントを有する重複エリアの図である。

【0058】

図15は、実施形態によるDNN符号化及びDNN予測残差符号化のための装置のブロック図である。

【0059】

図16は、実施形態によるDNN復号化及びDNN予測残差復号化のための装置のブロック図である。

30

【0060】

図17Aは、予測ブロック及び再構成されたピクセルの図である。

【0061】

図17Bは、実施形態による2つのコンテキスト・エリア予測方法を示す図である。

【0062】

図17Cは、実施形態によるパディング予測方法を示す図である。

【0063】

図17Dは、実施形態による変換予測方法を示す図である。

【0064】

図18は、実施形態によるポスト・フィルタリングを用いたブロック・ワイズ・ニューラル画像圧縮方法のフローチャートである。

40

【0065】

図19は、実施形態によるポスト・フィルタリングを用いたブロック・ワイズ・ニューラル画像圧縮装置のブロック図である。

【0066】

本開示は、ブロック・ベースのイントラ予測からの残差を用いてDNNによって入力画像を圧縮するためのニューラル画像圧縮(NIC)フレームワークを記述している。

【0067】

図3は、本件で説明される方法、装置、及びシステムが実装され得る実施形態による環境3

50

00を示す図である。

【0068】

図3に示すように、環境300は、ユーザー装置310、プラットフォーム320、及びネットワーク330を含んでもよい。環境300のうちの装置は、有線接続、無線接続、又は有線及び無線接続の組み合わせを介して相互接続されてもよい。

【0069】

ユーザー装置310は、プラットフォーム320に関連する情報を、受信、生成、記憶、処理、及び/又は提供することが可能な1つ以上の装置を含む。例えば、ユーザー装置310は、演算装置(例えば、デスクトップ・コンピュータ、ラップトップ・コンピュータ、タブレット・コンピュータ、ハンドヘルド・コンピュータ、スマート・スピーカ、サーバー等)、携帯電話(例えば、スマートフォン、無線電話など)、ウェアラブル装置(例えば、一対のスマート・グラス又はスマート・ウォッチ)、又は類似の装置を含んでもよい。幾つかの実装では、ユーザー装置310は、プラットフォーム320から情報を受信し及び/又はプラットフォーム320へ情報を送信することができる。

10

【0070】

プラットフォーム320は、本件の他の箇所で説明されるように、1つ以上の装置を含む。幾つかの実装では、プラットフォーム320は、クラウド・サーバー又はクラウド・サーバーのグループを含んでもよい。幾つかの実装では、プラットフォーム320は、ソフトウェア・コンポーネントが交換できるようにモジュールとして設計されてもよい。このように、プラットフォーム320は、異なる用途のために容易に及び/又は迅速に再構成されてもよい。

20

【0071】

幾つかの実装では、図示のように、プラットフォーム320は、クラウド演算環境322でホストされてもよい。特に、本件で説明される実装は、プラットフォーム320を、クラウド演算環境322でホストされるものとして説明しているが、一部の実装では、プラットフォーム320は、クラウド・ベースでなくともよく(即ち、クラウド・演算環境の外部で実装されてもよく)、又は部分的にクラウド・ベースであってもよい。

【0072】

クラウド演算環境322は、プラットフォーム320をホストする環境を含む。クラウド・演算環境322は、プラットフォーム320をホストするシステム及び/又は装置の物理的な位置及び構成に関する知識をエンド・ユーザー(例えば、ユーザー装置310)に要求しない演算、ソフトウェア、データ・アクセス、ストレージ等のサービスを提供することができる。図示のように、クラウド演算環境322は、演算リソース324のグループ(全体的には“演算リソース(複数形)324”と称し、個々には“演算リソース(単数形)324”と称する)を含んでもよい。

30

【0073】

演算リソース324は、1つ以上のパーソナル・コンピュータ、ワークステーション・コンピュータ、サーバー装置、又は他のタイプの演算及び/又は通信装置を含む。幾つかの実装では、演算リソース324は、プラットフォーム320をホストしてもよい。クラウド・リソースは、演算リソース324において実行する演算インスタンス、演算リソース324において提供される記憶装置、演算リソース324によって提供されるデータ転送装置などを含んでもよい。幾つかの実装では、演算リソース324は、有線接続、無線接続、又は有線及び無線接続の組み合わせを介して、他の演算リソース324と通信することができる。

40

【0074】

図3に更に示すように、演算リソース324は、1つ以上のアプリケーション(“APP”)324-1、1つ以上の仮想マシン(“VM”)324-2、仮想化されたストレージ(“VS”)324-3、1つ以上のハイパーバイザ(“HYP”)324-4のようなクラウド・リソースのグループを含む。

【0075】

アプリケーション324-1は、ユーザー装置310及び/又はプラットフォーム320によって提供又はアクセスされることが可能な1つ以上のソフトウェア・アプリケーションを含む

50

。アプリケーション324-1は、ユーザー装置310上にソフトウェア・アプリケーションをインストールして実行する必要性を排除することができる。例えば、アプリケーション324-1は、プラットフォーム320に関連するソフトウェア、及び/又はクラウド演算環境322を介して提供されることが可能な他の任意のソフトウェアを含んでもよい。幾つかの実装では、1つのアプリケーション324-1は、仮想マシン324-2を介して、1つ以上の他のアプリケーション324-1へ/からの情報を送信/受信することができる。

【0076】

仮想マシン324-2は、物理マシンのようなプログラムを実行するマシン(例えば、コンピュータ)のソフトウェア実装を含む。仮想マシン324-2は、仮想マシン324-2による現実の何らかのマシンに対する用途及び対応の程度に応じて、システム仮想マシン又はプロセス仮想マシンの何れかであってもよい。システム仮想マシンは、完全なオペレーティング・システム(“OS”)の実行をサポートする完全なシステム・プラットフォームを提供することができる。プロセス仮想マシンは、単一のプログラムを実行してもよく、単一のプロセスをサポートしてもよい。幾つかの実装では、仮想マシン324-2は、ユーザー(例えば、ユーザー装置310)に代わって実行することができ、データ管理、同期化、又は長時間データ転送のようなクラウド演算環境322のインフラストラクチャを管理することができる。

【0077】

仮想化された記憶装置324-3は、演算リソース324の記憶システム又は装置内で仮想化技術を使用する1つ以上の記憶システム及び/又は1つ以上の装置を含む。幾つかの実装では、記憶システムのコンテキスト内で、仮想化のタイプは、ブロック仮想化及びファイル仮想化を含んでもよい。ブロック仮想化は、物理ストレージ又はヘテロジニアス構造に関係なく、ストレージ・システムがアクセスを受けることができるように、物理ストレージからの論理ストレージの抽象化(又はセパレーション)を参照することができる。このセパレーションは、エンド・ユーザーのストレージをアドミニストレータがどのように管理するかについての柔軟性を、ストレージ・システムのアドミニストレータに許可することができる。ファイル仮想化は、ファイル・レベルでアクセスされるデータと、ファイルが物理的に格納される場所との間の依存関係を排除することができる。これは、ストレージの使用、サーバーの統合、及び/又は非破壊的なファイル移動のパフォーマンスを最適化できるようにする可能性がある。

【0078】

ハイパーバイザ324-4は、複数のオペレーティング・システム(例えば、“ゲスト・オペレーティング・システム”)が、演算リソース324のようなホスト・コンピュータ上で同時に実行することを可能にするハードウェア仮想化技術を提供することができる。ハイパーバイザ324-4は、ゲスト・オペレーティング・システムに対して仮想オペレーティング・プラットフォームを提示し、ゲスト・オペレーティング・システムの実行を管理することができる。様々なオペレーティング・システムの複数のインスタンスが、仮想化されたハードウェア・リソースを共有する可能性がある。

【0079】

ネットワーク330は、1つ以上の有線及び/又は無線ネットワークを含む。例えば、ネットワーク330は、セルラー・ネットワーク(例えば、第5世代(5G)ネットワーク、ロング・ターム・エボリューション(LTE)ネットワーク、第3世代(3G)ネットワーク、符号分割多元接続(CDMA)ネットワーク等)、公衆陸上移動網(PLMN)、ローカル・エリア・ネットワーク(LAN)、ワイド・エリア・ネットワーク(WAN)、メトロポリタン・エリア・ネットワーク(MAN)、電話網(例えば、公衆交換電話網(PSTN))、プライベート・ネットワーク、アドホック・ネットワーク、イントラネット、インターネット、光ファイバ・ベースのネットワーク等、及び/又はこれら又は他のタイプのネットワークの組み合わせを含んでもよい。

【0080】

図3に示すデバイス及びネットワークの数及び配置は、一例として提供されている。実際には、図3に示すものに対して、追加の装置及び/又はネットワーク、より少ない装置及び/又はネットワーク、異なる装置及び/又はネットワーク、又は別様に配置された装置及び/

10

20

30

40

50

又はネットワークが存在してもよい。更に、図3に示す2つ以上のデバイスは、単一のデバイス内に実装されてもよいし、又は図3に示す単一のデバイスは、複数の分散されたデバイスとして実装されてもよい。追加的又は代替的に、環境300のデバイス・セット(例えば、1つ以上のデバイス)は、環境300の別のデバイス・セットによって実行されるものとして説明される1つ以上の機能を実行してもよい。

【0081】

図4は、図3の1つ以上のデバイスの例示的なコンポーネントのブロック図である。

【0082】

装置400は、ユーザー装置310及び/又はプラットフォーム340に対応してもよい。図4に示すように、装置400は、バス410、プロセッサ420、メモリ430、記憶コンポーネント440、入力コンポーネント450、出力コンポーネント460、及び通信インターフェース470を含んでもよい。

【0083】

バス410は、装置400のコンポーネント間の通信を可能にするコンポーネントを含む。プロセッサ420は、ハードウェア、ファームウェア、又はハードウェアとソフトウェアの組み合わせで実装される。プロセッサ420は、中央処理ユニット(CPU)、グラフィックス処理ユニット(GPU)、加速処理ユニット(APU)、マイクロプロセッサ、マイクロコントローラ、デジタル信号プロセッサ(DSP)、フィールド・プログラマブル・ゲート・アレイ(FPGA)、特定用途向け集積回路(ASIC)、又は別のタイプの処理コンポーネントである。幾つかの実装では、プロセッサ420は、機能を実行するようにプログラムされることが可能な1つ以上のプロセッサを含む。メモリ430は、ランダム・アクセス・メモリ(RAM)、リード・オンリー・メモリ(ROM)、及び/又は、別のタイプのダイナミック又はスタティック記憶装置(例えば、フラッシュ・メモリ、磁気メモリ、及び/又は、光メモリ)であって、プロセッサ420による使用のための情報及び/又は命令を記憶するものを含む。

【0084】

記憶コンポーネント440は、装置400の動作及び使用に関連する情報及び/又はソフトウェアを記憶する。例えば、記憶コンポーネント440は、ハード・ディスク(例えば、磁気ディスク、光ディスク、光磁気ディスク、及び/又はソリッド・ステート・ディスク)、コンパクト・ディスク(CD)、デジタル多用途ディスク(DVD)、フロッピー・ディスク、カートリッジ、磁気テープ、及び/又は他のタイプの非一時的なコンピュータ読み取り可能な媒体を、対応するドライブと共に含んでもよい。

【0085】

入力コンポーネント450は、ユーザー入力(例えば、タッチ・スクリーン・ディスプレイ、キーボード、キーパッド、マウス、ボタン、スイッチ、及び/又はマイクロホン)等を介して情報を受信することを、装置400が実行できるようにするコンポーネントを含む。追加的又は代替的に、入力コンポーネント450は、情報をセンシングするためのセンサー(例えば、グローバル・ポジショニング・システム(GPS)コンポーネント、加速度計、ジャイロスコープ、及び/又はアクチュエータ)を含んでもよい。出力コンポーネント460は、デバイス400(例えば、ディスプレイ、スピーカ、及び/又は1つ以上の発光ダイオード(LED))からの出力情報を提供するコンポーネントを含む。

【0086】

通信インターフェース470は、装置400が有線接続、無線接続、又は有線及び無線接続の組み合わせ等を介して他の装置と通信することを可能にするトランシーバのようなコンポーネント(例えば、トランシーバ及び/又は別個の受信機と送信機)を含む。通信インターフェース470は、デバイス400が、他のデバイスから情報を受信し、及び/又は他のデバイスへ情報を提供することを可能にする。例えば、通信インターフェース470は、イーサネット・インターフェース、光インターフェース、同軸インターフェース、赤外線インターフェース、無線周波数(RF)インターフェース、ユニバーサル・シリアル・バス(USB)インターフェース、Wi-Fiインターフェース、セルラー・ネットワーク・インターフェース等を含んでもよい。

10

20

30

40

50

【0087】

装置400は、本件で説明される1つ以上のプロセスを実行することができる。装置400は、メモリ430及び/又は記憶コンポーネント440のような非一時的なコンピュータ読み取り可能な媒体によって記憶されたソフトウェア命令を実行するプロセッサ420に応答して、これらのプロセスを実行してもよい。コンピュータ読み取り可能な媒体は、本件では非一時的なメモリ・デバイスとして定義される。メモリ・デバイスは、単一の物理ストレージ・デバイス内のメモリ空間、又は複数の物理ストレージ・デバイスに分散されたメモリ空間を含む。

【0088】

ソフトウェア命令は、メモリ430及び/又はストレージ・コンポーネント440内に、別のコンピュータ読み取り可能な媒体から、又は通信インターフェース470を介して別の装置から読み込まれてもよい。実行されると、メモリ430及び/又はストレージ・コンポーネント440に記憶されたソフトウェア命令は、プロセッサ420に、本件で説明される1つ以上のプロセスを実行させることができる。追加的又は代替的に、ハードワイヤード回路が、ソフトウェア命令の代わりに又はそれと組み合わせて使用され、本件で説明される1つ以上のプロセスを実行することができる。従って、本件で説明される実装は、ハードウェア回路とソフトウェアの何らかの特定の組み合わせに限定されない。

【0089】

図4に示すコンポーネントの数及び配置は、一例として提供されている。実際には、装置400は、図4に示されるものに対して、追加のコンポーネント、より少ないコンポーネント、異なるコンポーネント、又は別様に配置されたコンポーネントを含んでもよい。追加的又は代替的に、デバイス400のコンポーネントのセット(例えば、1つ以上のコンポーネント)は、デバイス400のコンポーネントの別のセットによって実行されるものとして説明される1つ以上の機能を実行してもよい。

【0090】

ポスト・フィルタリングを用いたブロック・ワイズ・ニューラル画像圧縮のための方法及び装置をこれから詳細に説明する。

【0091】

本開示はNICのためのブロック・ベースのイントラ予測及び残差コーディング・フレームワークを記述している。従来の予測メカニズムはNICに修正され、NICコーディング効率は大幅に改善される。オリジナル・ピクセルを符号化する代わりに、予測ブロックとオリジナル・ブロックの間の残差が符号化される。

【0092】

図5A及び図6はそれぞれNICのためのエンコーダ500及びデコーダ600の実施形態によるワークフローを示す。エンコーダ側では、入力画像 x が与えられると、それは先ず n 個のブロック $\{b_0, b_1, \dots, b_n\}$ にパーティション化される。ブロック b_i に関し、それは2つの圧縮方法(即ち、予測ベースの圧縮方法又はレギュラーDNNベースの圧縮方法)によって圧縮され、より少ない損失の再構成されたブロック b_{i-} がデコーダ側に選択される。レギュラーDNNベースの圧縮方法では、 b_i は、 $b_i^{\wedge}$ を生成するためにDNNエンコーダ510によって直接的に符号化され、 $b_i^{\wedge}$ は、再構成されたブロック b_{i-} を生成するために、DNNデコーダ520によって将来復号化されることが可能である。予測ベースの圧縮方法に関し、 $\{b_{0-}, b_{1-}, \dots, b_{i-1-}\}$ を、以前に再構成されたブロックとし、これらは、予測ブロック b_i' を計算するために、イントラ予測コンポーネント530のDNNを通る。予測残差 $r-b_i$ は、 b_i と b_i' の間の差に基づいて、加算器540によって計算することができる。この残差 $r-b_i$ は、DNN予測残差エンコーダ550によって符号化され、別の符号化されたコンパクト残差 $r-b_i^{\wedge}$ を生成する。次いで、符号化された残差 $r-b_i^{\wedge}$ は、復号化された予測残差 $r-b_{i-}$ を計算するために、DNN予測残差デコーダ560を通り、加算器570によって b_i' に加算されて、新たに復元された予測ブロック b_{i-} を計算する。ブロック選択コンポーネント580は、(b_i と比較して)より少ない損失を有する再構成されたブロック(上記の2つの方法からのもの)を選択する。選択信号 s_i は、デコーダ600にシグナリングされる。ブロック b_i に関し、 $b_i^{\wedge}$

10

20

30

40

50

又は $r\text{-}bl^{\wedge}$ (再構成された bi に対する有用な情報を含む)のうちの一方のみが、ビットストリームを介してデコーダへ送られることに留意されたい。表記を簡易にするために、ブロック bi に関し、 ui は、ビットストリームにおける記憶及び伝送のためにコンパクトである有用な情報として示される。

【0093】

予測コンポーネント530のDNNは、少数の畳み込み層及びデコンボリューション層を含んでもよい。異なる数の再構成されたブロックを、コンテキスト・エリアとしてサポートするために、完全に接続された層が、イントラ予測コンポーネント530のDNNで使用されることも可能である。例えば、画像超解像度で使用されるDNNを、(出力サイズを入力サイズと同じであるように変更すること等によって)本件で使用する事が可能である。

10

【0094】

図6に記載されているように、デコーダ600は、ブロック選択信号、符号化されたブロック残差 $r\text{-}bl^{\wedge}$ 、又は符号化されたブロック $bl^{\wedge}$ を含むビットストリームにおけるブロック・レベル情報に従って画像を再構成する。ブロック選択信号は、現在のブロックを復号化する方法を示す。ブロックがレギュラーDNNベースの圧縮方法を使用する場合、それは、 $bl^{\wedge}$ をDNNデコーダ520へ送信して、再構成された対応するブロック bl を生成し、後にこれは一組の以前の再構成ブロックに追加される。ブロックが予測ベースの圧縮方法を使用する場合、一組の以前の再構成ブロック $\{b0-, b1-, \dots, bi-1-\}$ が与えられると、(エンコーダ側と同じ)イントラ予測コンポーネント530のDNNは、予測された現在のブロック bi' を計算するために使用され、これは、復号化された残差ブロック $r\text{-}bl$ に加算器570によって加算されて、再構成ブロック bl を生成する。次いで、この再構成されたブロック bl は、一組の以前の再構成されたブロック $\{b0-, b1-, \dots, bi-1-\}$ に追加される。実施形態では、デコーダ側のイントラ予測コンポーネント530、DNNデコーダ520、DNN予測残差デコーダ560は、エンコーダ側と同じブロック・ワイズ対応メカニズムを使用する。デコーダ600はまた、デブロッキングDNN及びポスト・エンハンスメント型DNNを有する。再構成されたブロックは、ブロック間のアーチファクトを含む可能性がある $x-$ を形成する。従って $x-$ はそのようなアーチファクトを最小化するために使用されるデブロッキングDNNへ送られる。将来、損失を低減するために、画像は、最終的に再構成された $x-$ を生成するために、ポスト・エンハンスメントDNNへ送られる。デブロッキングDNN及びポスト・エンハンスメントDNNは、少数の畳み込み層を含んでもよい。例えば、画像超解像度で使用されるDNNを、(出力サイズを入力サイズと同じであるように変更すること等によって)本件で使用する事が可能である。

20

30

【0095】

実施形態では、DNNエンコーダ510は、DNNエンコーダを使用してブロック bi を符号化する。DNN予測残差エンコーダ550は、予測残差符号化DNNを使用して、残差 r_bi を符号化する。DNNデコーダ520は、DNNデコーダを使用して、符号化された残差 $bl^{\wedge}$ を復号化する。DNN予測残差デコーダ560は、予測残差復号化DNNを使用して、符号化された残差 $r_bl^{\wedge}$ を復号化する。DNNエンコーダとDNNデコーダはVAE構造をとり、予測残差符号化DNNと予測残差復号化DNNもVAE構造をとる。DNNエンコーダ/デコーダと予測残差符号化/復号化DNNの構造は、異なってもよいし或いは同じであってもよい。

40

【0096】

本開示は、イントラ予測コンポーネント530のDNN、DNNエンコーダ/デコーダ、及び予測残差符号化/復号化DNN、デブロッキングDNN、及びポスト・エンハンスメントDNNのための特定のネットワーク構造に関して如何なる制約も課していない。本開示は、ブロック・サイズ、画像をブロックに分割する方法、及びブロックの走査順序に関して如何なる制限も課していない。

【0097】

以下は、上述の予測ベースのDNNエンコーダ/デコーダ・アーキテクチャを実装するための幾つかの実施形態である。

【0098】

50

実施形態では、HEVC及びVVCにおけるもののような従来のビデオ・コーディング・システムにおけるブロッキング・パーティショニングを採用することが可能である。サンプルの正方形配列であるコーディング・ツリー・ユニット(coding tree unit, CTU)は、基本的な処理単位と考えられる。ピクチャは一連のCTUに分割されることが可能である。潜在的な更なるパーティション化が可能であり、CTUの下で、より小さなブロックを使用することが可能である。CTUサイズは、CTU内部の 32×32 、 64×64 、 128×128 等であることが可能である。上述の予測ベースのDNNエンコーダ/デコーダを実施するために使用されるブロックは、CTU内のコーディング・ブロック又は予測ブロックであってもよい。

【0099】

別の実施形態では、従来のビデオ・コーディング・システムにおけるものを採用する以外に、簡略化されたブロック・パーティショニング方法が使用されてもよい。例えば、ピクチャは一連の $M \times N$ ブロックに分割されることが可能である。各ブロックに関し、それは、四分木パーティションを使用して、4つのより小さなブロックに更に分割されてもよい。ブロックを分割するかどうかは、ビットストリームでシグナリングすることが可能であり、或いは学習ベースのアプローチで推論することが可能である。

【0100】

実施形態では、上述の予測ベースの方法が、画像全体に関して使用される(そして画像全体はブロックにパーティション化される)。別の実施形態では、最初に画像が幾つかのパッチに分割される可能性があり、各パッチについて、更にブロックにパーティション化され、画像圧縮がパッチ・レベルで並列に実行される。

【0101】

実施形態では、スキャン開始ブロックは、左上コーナー、右上コーナー、左下コーナー、及び右下コーナーにあってもよい。そして、スキャン順序は、幅を最初にするだけでなく、高さを最初にすることも可能である。

【0102】

実施形態では、予測DNNは、ブロックを予測するために、全ての再構成されたブロックを入力として取り込む。別の実施形態では、予測DNNは、ブロックを予測するために、部分的に再構成されたブロック(例えば、予測されるべきものに隣接するブロック)を取り込む。

【0103】

実施形態では、予測DNNは：1)再構成されたブロックを潜在空間に変換するための複数の畳み込み層のセット；2)全結合層、又は連結演算、又は空間ピラミッド・プーリング層であって、異なる再構成されたブロックからの潜在空間を一緒に統合するもの；及び3)統合された潜在空間からターゲット・ブロックの予測を生成するために使用されるデコンボリューション・ネットワーク、という3つの主要なコンポーネントから構成される。

【0104】

実施形態では、イントラ予測コンポーネント530は、図5Bに示すように、3つのコンポーネントから構成される。再構成されたブロックは、まず、それら位置に応じて1つ以上のグループにグループ化される。図5Bは一例として3つのグループを示していることに留意されたい。実施形態では、グループの数は、1(即ち、1つのグループ内に全ての再構成されたブロックがある)からブロックの数(即ち、各々の再構成されたブロックはグループ内にある)までの範囲内にある。変換コンポーネント532において、各グループについて、変換532A、532B、又は532Nは、ブロックを潜在空間に変換する。変換コンポーネント532は、少数の畳み込み層から構成されてもよい。それはまた、正しい予測を行うために、再構成されたブロックの中で見るべき最も重要な場所を強調することが可能なアテンション・メカニズムから構成されることも可能である。統合コンポーネント534は、潜在空間を一緒に統合する。潜在空間を結合する能力を有する任意のオペレーション又はDNNレイヤが本件で使用可能である。例えば、それは全結合層、連結演算、又は空間ピラミッド・プーリング層であることが可能である。予測コンポーネント536において、予測モジュールは、統合コンポーネント534によって生成された統合された潜在空間に従って、

10

20

30

40

50

ブロックbl-を予測する。通常、それはデコンボリューション層のセットを含む。それはまた、正しい予測を行うために、再構成されたブロックの中で見るべき最も重要な場所を強調することが可能なアテンション・メカニズムから構成されることも可能である。変換コンポーネント532、統合コンポーネント534、及び予測コンポーネント536は、訓練プロセス中に学習が行われるDNNである。

【0105】

上記においてDNNの使用を通じてピクチャを符号化及び復号化する方法に関する幾つかの方法を説明する。サンプルの再構成後に、再構成されたブロックは、ブロックの境界の間にアーチファクトを含む可能性がある $x-'$ を形成する。次いで、 $x-'$ は、このようなアーチファクトを最小化して $x-''$ を生成するために使用されるデブロッキングDNNに送られる。復号化された画像品質を改善するために、画像は、最終的に再構成された $x-$ を生成するために、ポスト・エンハンスメントDNNへ送られる。図7は、後処理コンポーネント700の実施形態のワークフローを示す。デブロッキング・コンポーネント710及びポスト・エンハンスメントコンポーネント720の実施形態の詳細はそれぞれ以下のセクションで行われる。

【0106】

デブロッキング・コンポーネント710及びポスト・エンハンスメント・コンポーネント720に関する処理の順序は、交換されてもよい。即ち、ポスト・エンハンスメント・コンポーネント720が最初に使用され、デブロッキング・コンポーネント710に続いてよい。

【0107】

ブロック間のアーチファクトを更に低減するために、以下のデブロッキング方法のうちの1つ(又はその組み合わせ)を使用することができる。図8は、デブロッキング・コンポーネント710の実施形態のワークフローを示す。描写を容易にするために、図8には4つの等しいサイズのブロックのみが存在している。一般的なケースでは、画像は多数のブロックにパーティション化されることが可能であり、各ブロックのサイズは異なってもよいことに留意されたい。

【0108】

ブロック間の(ブロック境界エリアにおける)アーチファクトを低減するために、DNNベースのデブロッキング・モデルを使用することができる。まず、ブロック810から、隣接するブロックによって誘発されるアーチファクトを含むエリア(このような領域は、以下では境界エリア820として示される)を取り出すことができる。図8に示されるように、境界エリアは、隣接する2つのブロックからの2つのサブ・エリアから構成され、境界エリアのサイズは、 $2m \times n$ 又は $n \times 2m$ であり、 n はブロックの長さであり、 m はブロックから取り出されるエリアの高さである。 m の値は整数として変化する可能性がある。例えば、それは4ピクセルであるとするのが可能である。境界エリアは、アーチファクトを低減して、デブロッキング済みエリア830を生成するためにデブロッキング・コンポーネント710へ送られる。次いで、図8に示すように、デブロッキング済みエリアは、画像内で840にアップデートされる。

【0109】

図8のケースでは、境界エリアは、2つの隣接するブロックから特定される。しかしながら、図9に示されるように一例として、境界エリア910は、複数のブロックから構成されてもよい。

【0110】

図8は、境界エリアを改善するために、単一のデブロッキング・モデルを使用する方法を示す。図10は、マルチ・モデル・デブロッキング・コンポーネント1040A、1040B、及び1040Nの実施形態のワークフローを示す。キーとなるアイデアは、異なるタイプの境界エリア1010に対して、対応するデブロッキング・モデルが、アーチファクトを除去するために使用されることである。図10に示されるように、境界エリア1010が取り出されると、分類コンポーネント1020が(DNNベースのモデル、又はビデオ・コーデックで使用されている従来の方法であるとするのが可能である)、境界エリアを、異なるカテゴリに

10

20

30

40

50

分類するために使用される。次いで、境界エリア1030は、それらのカテゴリに応じて、異なるデブロッキング・コンポーネント1040A、1040B、及び1040Nへ送られる。実際には、境界エリアのタイプを定義するために、以下のメトリックを使用することができる(但し、それらに限定されない)。第1に、境界エリアの内容を使用することが可能である。例えば、高い周波数の中身を有する境界エリアは、低い周波数の中身をデブロッキングするために設計されたモデルとは異なるモデルに送られる。第2に、アーチファクトの強度を使用することが可能である。マルチ・モデル・デブロッキング・コンポーネント1040A、1040B、及び1040Nにおける境界エリアは、図10に描かれているものに限定されないことに留意されたい。図9に描かれているデブロッキング・エリアもまた、マルチ・モデルのデブロッキング方法を使用することが可能である。上記において、エリアの周波数は、エリア内の異なるサンプルの最大差によって測定されてもよい。一例では、境界の一方の側のエッジ近傍のサンプルの差が測定されてもよいし；別の例では、境界の他方の側のエッジ近傍のサンプルの差が測定されてもよい。

【0111】

図8及び10に描かれているDNNデブロッキング・モデルの主なゴールは、ブロック間のアーチファクトを除去することである。境界に近接したピクセルは、境界から離れたピクセルと比較して、より強くデブロッキングされてもよい。デブロッキング・モデルは、少数の畳み込み層を含んでもよい。CNNベースのアテンション・メカニズム(非ローカル・アテンション、SEnetなど)と残差ネット(convnetと活性化関数のセットを含むもの)が使用されてもよい。例えば、画像の超解像度で使用されるDNNが、(出力サイズを入力サイズと同じであるように変更すること等によって)本件で使用されてもよい。

【0112】

上記においてDNN又はその他の学習に基づく方法を利用して境界エリア上でデブロッキングをどのように実行するかについての方法が説明されている。代替的に、エンコーダ及びデコーダは、従来のデブロッキング方法又は本件で説明される方法の中で選択されることが可能である。選択の決定は、スライス/ピクチャ・レベル、ピクチャのグループ、又はシーケンス・レベル、のような様々なレベルで行うことが可能である。選択は、シグナリング・フラグを送信することによって行うことが可能であり、又は内容の分析から推測されることも可能である。

【0113】

上記においてDNN又はその他の学習に基づく方法を利用して境界エリア上でデブロッキングをどのように実行するかについての方法が説明されている。代替的に、エンコーダ及びデコーダは、本件で説明される方法に加えて、種々のレベルの境界強度を適用することが可能であり、但し、ピクセルにおけるDNN導出調整は境界強度のデフォルト・レベルにあると仮定する。境界条件とブロック・コーディング特徴を分析することにより、デフォルト調整を拡大又は縮小するために、異なるレベルのBSが割り当てられてもよい。

【0114】

デブロッキング・モジュールを使用することによりアーチファクトを低減した後、再構成された画像がエンハンスメント・モジュールに送られ、最終的に再構成された画像x-が生成される。図11は、ポスト・エンハンスメント・コンポーネント720の実施形態のワークフローを示す。描写を簡易にするために、図11では、4つの等しいサイズのブロックのみが存在している。一般的なケースでは、画像は多数のブロックにパーティション化される可能性があり、各ブロックのサイズは異なる可能性があることに留意されたい。

【0115】

画像の質を高めるために、DNNベースのポスト・エンハンスメント・モデルを、ポスト・エンハンスメント720コンポーネントとして使用することができる。まず、ブロック1110から、デブロッキング・モデルへは送付されないエリア(このようなエリアは、以下では残りのブロック1120として示される)を取り出すことができる。図11に示すように、残りのブロック1120は、1つのブロックからのエリアで構成され、境界エリアのサイズは、 $(n-m) \times (n-m)$ であり、ここで、 n はブロックの辺の長さであり、 m はデブロッキングのため

のエリアの長さである。mの値は整数である。それは例えば4ピクセルである。残りのブロック1120は、エリアの品質を更に高めて、エンハンスメント・ブロック1140を生成するために、ポスト・エンハンスメント・コンポーネント720へ送られる。次いで、エンハンスメント・ブロック1140は、画像に戻してアップデートされる(1150)。代替的に、残りのブロックは、全体の一部分においてデブロッキング境界エリアとオーバーラップしてもよい。後者の場合、残りのブロックは、コーディング・ブロック全体であってもよい。

【0116】

上記の実施形態は、残りのブロックを改善するために単一のポスト・エンハンスメント・モデルを使用する方法を説明している。図12は、複数モデルのポスト・エンハンスメント・コンポーネント1230A、1230B、及び1230Nの実施形態のワークフローを示す。キーとなるアイデアは、異なるタイプの残りのブロック1120について、対応するポスト・エンハンスメント・モデルが、品質を改善するために使用されることである。図12に示されるように、残りのブロック1120が取り出されると、分類コンポーネント1210が(DNNベースの分類モデル、又はビデオ・コーデックで使用されている従来する方法であるとする)ことが可能である)、残りのブロックを、異なるカテゴリに分類するために使用される。従って、分類された残りのブロック1220は、それらのカテゴリに応じて、異なるポスト・エンハンスメント・コンポーネント1230A、1230B、及び1230Nへ送られる。実際には、残りのブロックのタイプを定義するために、以下のメトリックが使用される(但し、それらに限定されない)。例えば、残りのブロック・エリアの内容が使用してもよい。この例では、高い周波数の中身を有する残りのブロックは、低い周波数の中身を強調するために設計されたモデルとは異なるモデルに送られる。

【0117】

ブロック・レベルで画像を強調する代わりに、画像全体を一度に強調するポスト・エンハンスメント・モデルが考慮されてもよい。図13は、残りのブロック及び境界エリアを含む画像1320全体に対するポスト・エンハンスメント・コンポーネント1310の実施形態のワークフローを示す。描写を簡易にするために、図13では、4つの等しいサイズのブロックのみが存在している。一般的なケースでは、画像は多数のブロックにパーティション化される可能性があり、各ブロックのサイズは異なる可能性があることに留意されたい。このようにして画像を強調することにより、アーチファクト境界は更に小さくなる。

【0118】

図11-13に描かれているDNNエンハンスメント・モデルの主な機能は、画像の品質を高めることである。ポスト・エンハンスメント・コンポーネント720は、少数の畳み込み層を含んでもよい。CNNベースのアテンション・メカニズム(非ローカル・アテンション、SEnetなど)と残差ネット(convnetと活性化関数のセットを含むもの)が、エンハンスメントのゴールを達成するために使用されてもよい。例えば、画像の超解像度で使用されるDNNが、(出力サイズを入力サイズと同じであるように変更すること等によって)本件で使用されてもよい。

【0119】

上述の方法では、2つのモジュールが、重複する領域を有する場合、同じサンプルが、ポスト・エンハンスメント・コンポーネント720及びデブロッキング・コンポーネント710の両方からの修正の影響を受ける。その場合、ポスト・エンハンスメントからの修正されたサンプルは $P_1, P_2, \dots$ と表され;同様に、デブロッキングからの修正されたサンプルは $D_1, D_2, \dots$ と表される。図14に示すように、 P_i 及び D_i サンプル($i=0,1, \dots$)の両方からの加重平均が、 A_i として示される最終的な出力サンプルを生成するために使用される。このプロセスは、境界エリアでのピクセル・ブレンドと言及される。例えば、 $A_i(i=0,1, \dots N-1)$ は、ポスト・エンハンスメント・エリア1410とデブロッキング・エリア1420とのブレンド後の重複エリアにおける2つのサンプルを表す。これらは、 $A_i = S(w_i * P_i + (1 - w_i) * D_i)$ のように導出することが可能であり、ここで、 w_i は、エリア内の位置 i に関して指定される重みパラメータである。異なる位置については、ウェイトが異なる可能性がある。一例では、 w_i は常に0.5に設定されている。別の例では、 w_i 値は2つのモジュールま

での距離に依存する。一般に、一方のモジュールにより近い位置は、同じモジュールにより生成されるサンプルに関して、より大きなウェイトの割り当てを受ける。図14における一例では、 $A1=5/8 \cdot P1+3/8 \cdot D1$; $A2=3/8 \cdot P2+5/8 \cdot D2$ である。

【0120】

訓練プロセスの目標は、予測DNN、符号化DNN、復号化DNN、予測残差符号化DNN、予測残差復号化DNN、デブロッキングDNN、及びポスト・エンハンスメントDNNを学習することである。

【0121】

訓練プロセスにおいて、学習されるべき上述のDNNのウェイト係数は、例えば予め訓練された対応するDNNモデルを使用して、又はそれらを乱数に設定することによって、先ず初期化される。次いで、入力トレーニング画像 x が与えられ、図5Aで説明された符号化プロセスを通して、ビットストリームに符号化された情報を生成し、次いでそれは図6で説明された復号化プロセスを通して、画像 $\hat{x}$ を計算して再構成することになる。NICについては、2つの競合するターゲットが存在する可能性があり：より良い再構成品質_対_より少ないビット消費である。品質損失関数 $D(x, \hat{x})$ は、再構成品質を測定するために使用され、これは、通常、歪損失と呼ばれ、従来のPSNR、MS-SSIM、又は両者の重み付けされた組み合わせのようなものである。レート損失

【数6】

$$R(\sum_1^n s_i, \sum_1^n u_i)$$

が、圧縮された表現のビット消費を測定するために計算される。従って、トレードオフ・ハイパーパラメータ λ を用いて、ジョイント・レート歪(R-D)損失を最適化する：

【0122】

【数7】

$$L(x, \hat{x}, \hat{r}_1, \dots, \hat{r}_N, \hat{y}) = \lambda D(x, \hat{x}) + R(\sum_1^n s_i, \sum_1^n u_i) + \beta E \quad (\text{Eq. 11}),$$

【0123】

ここで、 E は、符号化される前のオリジナル・ブロック残差と比較した場合の、復号化されたブロック残差の歪を測定するものであり、これは、残差符号化/復号化DNNと符号化/復号化DNNに対する正則化損失(regularization loss)として作用する。 β は、正則化損失の重要度のバランスをとるためのハイパーパラメータである。

【0124】

大きな λ を用いる場合の訓練は、より小さな歪を有するが、より多くのビット消費を伴う圧縮モデルをもたらし、逆もまた同様である。DNNのウェイト係数を学習するために、符号化DNNと予測残差符号化DNNは、図15に示すように、2つの部分に分解される。同様に、復号化DNNと予測残差復号化DNNは、図16に示すように、2つの部分に分解される。

【0125】

図15は、実施形態によるDNN符号化及びDNN予測残差符号化のための装置1500のブロック図である。図16は、実施形態によるDNN復号化及びDNN予測残差復号化のための装置1600のブロック図である。

【0126】

図15に示すように、符号化DNNは、ニューラル・エンコーダ1510とそれに続くエンコーダ1520とに分解される。また、対応する処理モジュールは、ニューラル符号化モジュールとそれに続く符号化モジュールとに分解される。同様に、予測残差符号化DNNは、ニューラル残差エンコーダ1530とそれに続く残差エンコーダ1540とに分解される。また、対応する処理モジュールは、ニューラル予測残差符号化モジュールとそれに続く予測残差符

号化モジュールとに分解される。実施形態では、ニューラル・エンコーダ1510又はニューラル残差エンコーダ1530はDNNであり、DNNは訓練プロセス中に学習され、検証段階で展開される。他方、エンコーダ1520又は残差エンコーダ1540は、訓練及び検証で異なる。検証段階では、エンコーダ1520又は残差エンコーダ1540は、実際のエンコーダであり、これは量子化器とそれに続くエントロピー・エンコーダとを含む可能性がある。訓練段階では、エンコーダ1520又は残差エンコーダ1540は、統計データ・サンプラを使用して、実際の量子化器の効果を近似し、エントロピー・エンコーダは訓練中はスキップされる。

【0127】

同様に、図16に示すように、復号化DNNはデコーダ1610とそれに続くニューラル・デコーダ1620とに分解される。また、対応する処理モジュールは、復号化モジュールとそれに続くニューラル復号化モジュールとに分解される。予測残差復号化DNNは、残差デコーダ1630とそれに続くニューラル残差デコーダ1640とに分解される。また、対応する処理モジュールは、予測残差復号化モジュールとそれに続くニューラル予測残差復号化モジュールとに分解される。

【0128】

実施形態では、ニューラル・デコーダ1620又はニューラル残差デコーダ1640はDNNであり、DNNは訓練プロセス中に学習され、検証段階で展開される。他方、デコーダ1610又は残差デコーダ1630は、訓練及び検証で異なる。検証段階では、デコーダ1610又は残差デコーダ1630は、実際のデコーダであり、これはエントロピー・デコーダとそれに続く逆量子化器とを含む可能性がある。訓練段階では、デコーダ1610又は残差デコーダ1540はスキップされる。

【0129】

本開示は、ニューラル・エンコーダ1510、エンコーダ1520、デコーダ1610、ニューラル・デコーダ1620、ニューラル残差エンコーダ1530、残差エンコーダ1540、残差デコーダ1630、又はニューラル残差デコーダ1640で使用される方法に如何なる制限も課していない。換言すれば、本開示のNICフレームワークは、様々なタイプの符号化DNN、復号化DNN、予測残差符号化DNN、及び予測残差復号化DNNを収容することができる。

【0130】

一実施形態では、このフレームワークは、一般に、スケール・ハイパープライア(scale-hyperprior)エンコーダ・デコーダ・フレームワーク(又は混合ガウシアン尤度(Gaussian Mixture Likelihoods)フレームワーク)及びその変形、RNNベースの再帰的圧縮方法及びその変形のような、任意のDNNベースの画像圧縮方法を使用することができる。

【0131】

別の実施形態では、フレームワークは、JPEG、JPEG2000、BPGのような任意の既存の画像圧縮方法を用いて動作することができる。

【0132】

学習プロセスの実施形態において、予測DNN、符号化DNN、予測残差符号化DNN、符号化DNN、予測残差復号化DNN、デブロッキングDNN、及びポスト・エンハンスメントDNNは、エンド・ツー・エンド(E2E)方式で、上記の逆伝搬勾配に基づいて一緒に更新され得ることは、言及するに価する。また、これらのDNNのうちの何れも個別に学習することが可能である。例えば、予測DNNは、何らかのデータセット(上述の訓練プロセスで使用されたデータセットと同じか又は異なるもの)を用いて個別に訓練されることが可能であり、それらは上述の訓練プロセスでは固定され、その場合においては、残りのDNNのウェイト係数のみが更新される。

【0133】

図17Aに示されるように、ブロック1720(サイズ $w \times h$)を予測するために、以前に再構成されたピクセル1710がコンテキストとして使用されてもよい。コンテキスト・ピクセルは、予測モジュールへの入力として取り込まれ、ブロックの予測サンプルが生成される。以下では、再構成されたピクセル1710からコンテキスト・エリアを形成する方法によ

10

20

30

40

50

って、予測方法が議論される。

【0134】

以下の全ての実施形態では、ラスタ・スキャン順序が使用され、一例として左上ブロックから始まっていることに留意されたい。一般的なケースでは、スキャン開始ブロックは、左上コーナー、右上コーナー、左下コーナー、及び右下コーナーとすることが可能である。そして、スキャン順序は、幅を最初に行うだけでなく、高さを最初に行うことも可能である。

【0135】

HEVC及びVVCでは予測の過去の経験に従って、コンテキスト・エリアは予測されるべきブロックを囲む。実施形態では、再構成されたピクセル1710を、2つの別個のエリアとして考えている(即ち、予測ブロックの上にあるコンテキスト・エリアCaboveと、予測ブロックの左側にあるコンテキスト・エリアCleftであり、ラスタ・スキャン順序で開始点は左上ブロックである)。図17Bは、この予測方法の実施形態を示す。図17Bに示すように、コンテキスト・エリアは‘L’又は‘7’のような形状である。2つのコンテキスト・エリアに基づいてブロックを予測するために、予測モデルは先ずこれら2つのエリア(CaboveとCleft)を潜在表現に変換し、次いで、2つの潜在表現を統合する。このモデルは、統合された潜在表現に基づいて、最終的なブロック予測を生成する。

10

【0136】

図17Bに示されるように、実施形態では、左コンテキスト及び上コンテキストのサイズは、 $2h \times w$ 及び $2w \times h$ (2つの予測ブロックサイズ)とすることが可能である。他の実施形態において、左コンテキスト及び上コンテキストのサイズは異なってもよく、予測ブロック・サイズの整数倍でなくてもよい。

20

【0137】

コンテキスト・エリアをCleftとCaboveにパーティション化する代わりに図17Cに示されるように一実施形態はコンテキスト・エリアを矩形エリアとして考える。それを行うために、パディング処理が必要される可能性がある。図17Cは、パディング・エリアがCaboveとの下とCleftの右側にある例を示す。パディングの後、パディング・エリアとコンテキスト・エリアは、新しい矩形領域を、予測モデルへの入力として形成する。他の実施形態では、パディング・エリアは予測ブロックと同じであることが可能であり、又は、パディング・エリアは余分なブロックを含むことが可能であることに留意されたい(このアイデアは、パディング・エリアとコンテキスト・エリアを統合して矩形エリアを形成するものであり、従ってパディング・エリアは変化する可能性がある)。ツー・コンテキスト・エリア予測法(Two Context Area Prediction Method)とは異なり、パディング法の予測モデルへの入力は1つの矩形領域である。次いで、予測モデルは、入力を取り込み、潜在表現に変換し、そしてブロックの最終的な予測サンプルを生成する。

30

【0138】

以下は、議論しているパディング・エリアを充填するパディング方法の幾つかの実施形態である。

【0139】

一実施形態では、ゼロ・パディングを使用して、上述のエリアを埋めてもよい。

40

【0140】

一実施形態では、平均パディングを使用して、上述のエリアを埋めてもよい。その値は、CaboveとCleftを平均化することにより、又は画像全体を平均化することにより、又は画像データセットを平均化することにより(例えば、ImageNETで使用されるゼロ平均化法のようなものにより)計算することができる。

【0141】

一実施形態では、エリアはCaboveから延長されたピクセルを使用してパディングされてもよい。

【0142】

一実施形態では、エリアはCleftから延長されたピクセルを使用してパディングされてもよ

50

い。

【0143】

他の実施形態では、エリアは、Cleft及びCaboveから延長されたピクセルの組み合わせ(例えば、和、平均)を使用してパディングされてもよい。

【0144】

一実施形態では、従来のイントラ予測方法における予測ブロック・サンプル生成プロセスを使用して、パディング・エリアを埋めることが可能であり、但し、パディング・エリアは予測されるべき現在のブロックであると仮定する。このような方法の例は、DCモード、PLANARモード、不均等に重み付けされた平面モード、PDPC (位置依存予測結合)などを含む。

【0145】

本開示で言及されていない他のパディング方法もまた、パディング予測方法で 사용할ことが可能である。

【0146】

上記の全てのパディング方法において、パディング・エリアは、予測ブロックよりも大きいことが可能であり、又は予測ブロックと同じ大きさであることも可能である。前者の場合、Caboveの幅は予測ブロックより大きいであろう；後者の場合、Caboveの幅は予測ブロックと同じであろう。同様にパディング・エリアの高さも、予測ブロックと同じであるか又はより大きい可能性がある。

【0147】

パディングとは異なる別の方法において、周囲のコンテキスト・エリアは、以下のように、矩形形状に変換されてもよい。

【0148】

図17Dに示されるように、変換方法において、Cleft1730は、水平形状1740に変換され(図示を簡単にするために3ピクセルのみが示されている)、変換されたCleftとCaboveの間の中間エリア(図17D(a)の灰色ピクセル部分1760)をパディングして、矩形コンテキスト・エリア1770を形成する。かくて、コンテキスト・エリアの形状は矩形エリアに(‘L’から‘ー’に変換される。予測モデルへの入力1つの矩形エリアである。次いで、予測モデルは入力を取り込み、潜在表現に変換し、そしてブロックの最終的な予測を生成する。

【0149】

一実施形態では、Caboveは垂直形状に変換される。パディングの後に、矩形領域が得られる(‘L’から‘I’)。

【0150】

一実施形態では、ゼロ・パディングを使用して、中間エリアを埋めてもよい。

【0151】

一実施形態では、平均パディングを使用して、灰色のエリアを埋めて、灰色のエリア全体が同じ値を使用してパディングされてもよい。その値は、CaboveとCleftを平均化することにより、又は画像全体を平均化することにより、又は画像データセットを平均化することにより(例えば、ImageNETで使用するゼロ平均化法のようなものにより)計算することができる。

【0152】

一実施形態では、平均パディングを使用して、中間エリアを埋めてもよい。異なるラインが、異なる値を使用してパディングされてもよい。その値は、Cleft中の同一ライン内のピクセルを平均化することにより、又はCabove中の同一ライン内のピクセルを平均化することにより、又はCleftとCabove双方で同一ライン内のピクセルを組み合わせることにより(例えば、平均、和)、計算することができる。

【0153】

一実施形態では、中間エリアは、Cabove内の同一ラインから延長されたピクセルを使用してパディングされてもよい。

【0154】

一実施形態では、中間エリアは、Cleft内の同一ラインから延長されたピクセルを使用してパディングされてもよい。

【0155】

一実施形態では、中間エリアは、CleftとCabove双方で同一ラインから延長されたピクセルの組み合わせ(例えば、平均、和)を使用して、パディングされてもよい。

【0156】

一実施形態では、中間エリアは、CleftとCabove双方で同一ラインから延長された最も近いピクセルの組み合わせ(例えば、平均、和)を使用して、パディングされてもよい。即ち、変換後に、同一ラインの隣接する(左右の)ピクセルを使用して、灰色エリアをパディングする。

【0157】

代替的に、図17D(b)の部分に示されるように、コンテキスト・エリア1730の異なるラインに対して、余分な数のピクセルが、各ラインの総数を等しくするために使用される。即ち、変換1780及びシフト1790の後に、中間エリアをパディングすることなく、矩形エリアを確立することができる。左コンテキスト・エリアの右又は下に余分なピクセルがある場合、上記の任意の方法を使用してパディングされてもよい。

【0158】

本開示で言及されていない他のパディング方法もまた、パディング予測方法で 사용할ことが可能である。

【0159】

CaboveとCleftの形状が同一である場合(例えば、予測ブロックが正方形であり、上コンテキスト・エリアの形状が $3 \times w \times 2w$ であり、左コンテキスト・エリアの形状が $3 \times 2w \times w$ である場合；ここで、 w はブロックの幅である)、次の方法を使用して、2つのコンテキスト・エリアを統合してもよい。

【0160】

2つのコンテキスト・エリアは変換されて、 $3 \times 2w \times 2w$ のコンテキスト・エリアに統合され、予測モデルに送られて、最終的な予測を生成する。

【0161】

上コンテキスト・エリアは、 $3 \times 2w \times w$ に変換されてもよい。

【0162】

左コンテキスト・エリアは、 $3 \times w \times 2w$ に変換されてもよい。

【0163】

2つのコンテキスト・エリアは変換され、 $6 \times W \times 2W$ 又は $6 \times 2W \times W$ のエリアに統合され(チャンネル毎の連結)、予測モデルに送られて、最終的な予測を生成する。

【0164】

上コンテキスト・エリアは、 $3 \times 2w \times w$ に変換されてもよい。

【0165】

左コンテキスト・エリアは、 $3 \times w \times 2w$ に変換されてもよい。

【0166】

他の実施形態では、CaboveとCleftが潜在表現に先ず変換され、次いで、上記の方法を使用して、潜在表現を統合し、そして最終的な予測を生成することができる。

【0167】

図18は、実施形態によるスタック化可能なネスト化されたモデル構造を有するマルチ・レート・ニューラル画像圧縮の方法1800のフローチャートである。

【0168】

幾つかの実装形態では、図18の1つ以上のプロセス・ブロックは、プラットフォーム320によって実行されてもよい。幾つかの実装では、図18の1つ以上のプロセス・ブロックは、プラットフォーム320から分離された、又はプラットフォーム320を含む、別のデバイス又はデバイスのグループ、例えばユーザー・デバイス310によって実行されてもよい。

【0169】

図18に示すように、オペレーション1810において、方法1800は、第1ニューラル・ネットワークを使用して、入力画像のブロックを符号化するステップを含み、ここで、符号化されたブロックは、再構成されたブロックを生成するために、第2ニューラル・ネットワークを使用してデコーダによって復号化される。

【0170】

オペレーション1820において、方法1800は、予測ブロックを生成するために、第3ニューラル・ネットワークを使用して、再構成されたブロックに関してイントラ予測を実行するステップを含む。

【0171】

オペレーション1830において、方法1800は、予測残差を生成するために、入力画像のブロックと生成された予測ブロックとの間の差分を決定するステップを含む。

【0172】

オペレーション1840において、方法1800は、第4ニューラル・ネットワークを使用して、生成された予測残差を符号化するステップを含む。ここで、符号化された予測残差は、第5ニューラル・ネットワークを使用してデコーダにより復号化される。

【0173】

オペレーション1850において、方法1800は、復元された予測ブロックを生成するために、復号化された予測残差を、生成された予測ブロックに加えるステップを含む。

【0174】

方法1800は、更に、入力画像のブロックに関してより少ない損失をもたらすものを、再構成されたブロック及び生成され復元された予測ブロックの中から選択するステップ；及び再構成されたブロック及び復元された予測ブロックの中から選択されたものを示す選択信号を、再構成されたブロック及び復元された予測ブロックの中から選択されたものに対応する、符号化されたブロック及び符号化された予測残差のうちのもの、とともにデコーダへ送信するステップを含んでもよい。符号化されたブロック及び符号化された予測残差のうちの送信されたものは、送信された選択信号に基づいてデコーダにより復号化されることが可能である。

【0175】

再構成されたブロックに関してイントラ予測を実行するステップは：複数の再構成されたブロックを1つ以上のグループに、複数の再構成されたブロックの位置に基づいてグループ化するステップ；1つ以上の潜在空間を生成するために、1つ以上のグループの各々に関して個々の変換を実行するステップ；生成された1つ以上の潜在空間を統合するステップ；及び予測ブロックを生成するために、統合された1つ以上の潜在空間に関してイントラ予測を実行するステップを含んでもよい。

【0176】

方法1800は、更に：複数の再構成されたブロックの各々から、アーチファクトを含む境界エリアを取り除くステップ；1つ以上のニューラル・ネットワークを使用して、取り除かれた境界エリアに関してデブロッキングを実行して、アーチファクトを伴わないデブロック化されたエリアを生成するステップ；及び複数の再構成されたブロックの各々を、生成されたデブロック化されたエリアでアップデートするステップを含んでもよい。

【0177】

方法1800は、更に：前記複数の再構成されたブロックの各々から、取り除かれた境界エリアを伴わない残りのブロック・エリアを取り除くステップ；エンハンスド・ブロック・エリアを生成するために、少なくとも1つのニューラル・ネットワークを使用して、取り除かれた残りのブロック・エリアに関してポスト・エンハンスメントを実行するステップ；及び複数の再構成されたブロックの各々を、生成されたエンハンスド・ブロック・エリアでアップデートするステップを含んでもよい。

【0178】

再構成されたブロックに関してイントラ予測を実行するステップは：再構成されたブロックの上にある再構成されたピクセルの第1コンテキスト・エリアと、再構成されたブロッ

10

20

30

40

50

クの左にある再構成されたピクセルの第2コンテキスト・エリアとを取得するステップ；潜在表現を生成するために、取得された第1コンテキスト・エリアと取得された第2コンテキスト・エリアの各々に関して個々の変換を実行するステップ；生成された潜在表現を統合するステップ；及び予測ブロックを生成するために、統合された潜在表現に関してイントラ予測を実行するステップを含んでもよい。

【0179】

再構成されたブロックに関してイントラ予測を実行するステップは：再構成されたブロックの上にある再構成されたピクセルの第1コンテキスト・エリアと、再構成されたブロックの左にある再構成されたピクセルの第2コンテキスト・エリアと、第1コンテキスト・エリアの下及び第2コンテキスト・エリアの右にあるパディング・エリアとを取得するステップ；矩形エリアを取得するために、取得された第1コンテキスト・エリアと、取得された第2コンテキスト・エリアと、取得されたパディング・エリアを統合するステップ；潜在表現を生成するために、取得された矩形エリアに関して変換を実行するステップ；及び予測ブロックを生成するために、生成された潜在表現に関してイントラ予測を実行するステップを含んでもよい。

【0180】

図18は、方法1800の例示的なブロックを示しているが、幾つかの実装において、方法1800は、図18に示されているものに対して追加のブロック、より少ないブロック、異なるブロック、又は別様に配置されたブロックを含んでもよい。追加的又は代替的に、方法1800のブロックのうちの2つ以上は、並列的に実行されてもよい。

【0181】

図19は、実施形態によるスタック化可能なネスト化されたモデル構造を有するマルチ・レート・ニューラル画像圧縮の装置1900のブロック図である。

【0182】

図19に示されるように、装置1900は、第1符号化コード1910と、第1実行コード1920と、決定コード1930と、第2符号化コード1940と、加算コード1950とを含む。

【0183】

第1符号化コード1910は、第1ニューラル・ネットワークを使用して、入力画像のブロックを符号化することを、少なくとも1つのプロセッサに行わせるように構成されており、符号化されたブロックは、再構成されたブロックを生成するために、第2ニューラル・ネットワークを使用してデコーダによって復号化される。

【0184】

第1実行コード1920は、予測ブロックを生成するために、第3ニューラル・ネットワークを使用して、再構成されたブロックに関してイントラ予測を実行することを、少なくとも1つのプロセッサに行わせるように構成されている。

【0185】

決定コード1930は、予測残差を生成するために、入力画像のブロックと生成された予測ブロックとの間の差分を決定することを、少なくとも1つのプロセッサに行わせるように構成されている。

【0186】

第2符号化コード1940は、第4ニューラル・ネットワークを使用して、生成された予測残差を符号化することを、少なくとも1つのプロセッサに行わせるように構成されており、符号化された予測残差は、第5ニューラル・ネットワークを使用してデコーダにより復号化される。

【0187】

加算コード1950は、復元された予測ブロックを生成するために、復号化された予測残差を、生成された予測ブロックに加えることを、少なくとも1つのプロセッサに行わせるように構成されている。

【0188】

プログラム・コードは、更に：入力画像のブロックに関してより少ない損失をもたらすも

10

20

30

40

50

のを、再構成されたブロック及び生成され復元された予測ブロックの中から選択することを、少なくとも1つのプロセッサに行わせるように構成された選択コード；及び再構成されたブロック及び復元された予測ブロックの中から選択されたものを示す選択信号を、再構成されたブロック及び復元された予測ブロックの中から選択されたものに対応する、符号化されたブロック及び符号化された予測残差のうちのもの、とともにデコーダへ送信することを、少なくとも1つのプロセッサに行わせるように構成された送信コードを含んでもよい。符号化されたブロック及び符号化された予測残差のうちの送信されたものは、送信された選択信号に基づいてデコーダにより復号化されてもよい。

【0189】

第1実行コード1920は、更に：複数の再構成されたブロックを1つ以上のグループに、複数の再構成されたブロックの位置に基づいてグループ化すること；1つ以上の潜在空間を生成するために、1つ以上のグループの各々に関して個々の変換を実行すること；生成された1つ以上の潜在空間を統合すること；及び予測ブロックを生成するために、統合された1つ以上の潜在空間に関してイントラ予測を実行することを、少なくとも1つのプロセッサに行わせるように構成されていてもよい。

【0190】

前記プログラム・コードは、更に：複数の再構成されたブロックの各々から、アーチファクトを含む境界エリアを取り除くことを、少なくとも1つのプロセッサに行わせるように構成された第1除去コード；1つ以上のニューラル・ネットワークを使用して、取り除かれた境界エリアに関してデブロッキングを実行して、アーチファクトを伴わないデブロック化されたエリアを生成することを、少なくとも1つのプロセッサに行わせるように構成された第2実行コード；及び複数の再構成されたブロックの各々を、生成されたデブロック化されたエリアでアップデートすることを、少なくとも1つのプロセッサに行わせるように構成された第1アップデート・コードを含んでもよい。

【0191】

プログラム・コードは、更に：複数の再構成されたブロックの各々から、取り除かれた境界エリアを伴わない残りのブロック・エリアを取り除くことを、少なくとも1つのプロセッサに行わせるように構成された第2除去コード；エンハンスド・ブロック・エリアを生成するために、少なくとも1つのニューラル・ネットワークを使用して、取り除かれた残りのブロック・エリアに関してポスト・エンハンスメントを実行することを、少なくとも1つのプロセッサに行わせるように構成された第3実行コード；及び複数の再構成されたブロックの各々を、生成されたエンハンスド・ブロック・エリアでアップデートすることを、少なくとも1つのプロセッサに行わせるように構成された第2アップデート・コードを含んでもよい。

【0192】

第1実行コード1920は、更に：再構成されたブロックの上にある再構成されたピクセルの第1コンテキスト・エリアと、再構成されたブロックの左にある再構成されたピクセルの第2コンテキスト・エリアとを取得すること；潜在表現を生成するために、取得された第1コンテキスト・エリアと取得された第2コンテキスト・エリアの各々に関して個々の変換を実行すること；生成された潜在表現を統合すること；及び予測ブロックを生成するために、統合された潜在表現に関してイントラ予測を実行することを、少なくとも1つのプロセッサに行わせるように構成されていてもよい。

【0193】

第1実行コード1920は、更に：再構成されたブロックの上にある再構成されたピクセルの第1コンテキスト・エリアと、再構成されたブロックの左にある再構成されたピクセルの第2コンテキスト・エリアと、第1コンテキスト・エリアの下及び第2コンテキスト・エリアの右にあるパディング・エリアとを取得すること；矩形エリアを取得するために、取得された第1コンテキスト・エリアと、取得された第2コンテキスト・エリアと、取得されたパディング・エリアを統合すること；潜在表現を生成するために、取得された矩形エリアに関して変換を実行すること；及び予測ブロックを生成するために、生成された潜在表現

に関してイントラ予測を実行することを、少なくとも1つのプロセッサに行わせるように構成されていてもよい。

【0194】

従来のE2E画像圧縮方法と比較して、本開示は以下の利点を有する。予測機構は、オリジナル・ピクセルを符号化する代わりに、予測ブロックとオリジナル・ブロックの間の残差を符号化することによって、NICコーディング効率を改善するために活用される。柔軟で一般的なフレームワークは、様々なイントラ予測法及びイントラ予測プロセスからの残差に対する様々なニューラル符号化法に適合する。柔軟で一般的なフレームワークは、様々なタイプの品質メトリックに適合する。

【0195】

従来の予測方法と比較して、本開示は以下の利点を有する。実施形態はニューラル・ベースの画像圧縮方法に使用することができる。パディング方法及び変換方法に関し、コンテキスト・エリアをパーティション化する必要はない。パディング方法、変換方法、及び特殊なケースにおける方法に関し、コンテキストは矩形エリアであり、従って2つの分離されたエリアをマージするための追加の演算を減らすことができる。

【0196】

本方法は、別々に又は任意の順序で組み合わせて使用されてもよい。更に、方法(又は実施形態)、エンコーダ、及びデコーダの各々は、処理回路(例えば、1つ以上のプロセッサ、又は1つ以上の集積回路)によって実装されてもよい。一例では、1つ以上のプロセッサは、非一時的なコンピュータ読み取り可能な媒体に格納されたプログラムを実行する。

【0197】

前述の開示は、図示及び説明を提供するが、網羅的であることや、実装を開示された厳密な形態に限定するようなことを意図してはいない。修正及び変形が上記の開示に関して可能であり、又は実装の慣行から把握される可能性がある。

【0198】

本件で使用される場合に、コンポーネントという用語は、ハードウェア、ファームウェア、又はハードウェアとソフトウェアの組み合わせとして広く解釈されるように意図されている。

【0199】

本件で説明されるシステム及び／又は方法は、ハードウェア、ファームウェア、又はハードウェアとソフトウェアの組み合わせの様々な形態で実施されてもよいことは、明らかであろう。これらのシステム及び／又は方法を実装するために使用される実際の特化された制御ハードウェア又はソフトウェア・コードは、実装を制限するものではない。従って、システム及び／又は方法の動作及び挙動は、特定のソフトウェア・コードを参照することなく、本件で説明されており—ソフトウェア及びハードウェアは、本件における説明に基づいてシステム及び／又は方法を実装するように設計されてもよいことが理解される。

【0200】

特徴の組み合わせがクレームに記載され及び／又は明細書で開示されてたとしても、これらの組み合わせは、可能な実装の開示を限定するようには意図されていない。実際、これらの特徴の多くは、クレームに具体的には記載されていない、及び/又は明細書に開示されていない方法で組み合わせられる可能性がある。以下に列挙される各従属クレームは、1つのクレームのみに直接的に従属しているかもしれないが、可能な実装の開示は、各々の従属クレームを、クレーム・セット中の他の全てのクレームとの組み合わせで含んでいる。

【0201】

本件で使用される要素、動作、又は命令は、しかるべく明示的に記述されていない限り、重要又は必須と解釈されてはならない。また、本件で使用される場合、“ある”及び“或る”(“a” and “an”)という不定冠詞的な語は、1つ以上のアイテムを含むように意図されており、“1つ以上”と可換に使用されてもよい。更に、本件で使用される場合、“セット”という用語は、1つ以上のアイテム(例えば、関連するアイテム、非関連のアイテム、関連するアイテムと非関連のアイテムの組み合わせ等)を含むように意図されており、“1つ以上”と可

10

20

30

40

50

換に使用されてもよい。1つのアイテムのみが意図される場合、用語“1つ”又は類似の言葉が使用される。また、本件で使用される場合、“有する”、“含む”、“有している”等の用語は、オープン・エンドの用語であるように意図されている。更に、“基づいて”という言い回しは、明示的に別段の記載がない限り、“少なくとも部分的に基づいて”を意味するように意図されている。

【0202】

(付記1)

ポスト・フィルタリングを用いるブロック・ワイズ・ニューラル画像圧縮の方法であって：

第1ニューラル・ネットワークを使用して、入力画像のブロックを符号化するステップであって、符号化されたブロックは、再構成されたブロックを生成するために、第2ニューラル・ネットワークを使用してデコーダによって復号化される、ステップ；

予測ブロックを生成するために、第3ニューラル・ネットワークを使用して、前記再構成されたブロックに関してイントラ予測を実行するステップ；

予測残差を生成するために、前記入力画像のブロックと生成された予測ブロックとの間の差分を決定するステップ；

第4ニューラル・ネットワークを使用して、生成された予測残差を符号化するステップであって、符号化された予測残差は、第5ニューラル・ネットワークを使用して前記デコーダにより復号化される、ステップ；及び

復元された予測ブロックを生成するために、復号化された予測残差を、前記生成された予測ブロックに加えるステップ；

を含む方法。

(付記2)

付記1に記載の方法において、更に：

前記入力画像のブロックに関してより少ない損失をもたらすものを、前記再構成されたブロック及び生成され復元された予測ブロックの中から選択するステップ；及び

前記再構成されたブロック及び復元された予測ブロックの中から選択されたものを示す選択信号を、前記再構成されたブロック及び復元された予測ブロックの中から選択されたものに対応する、前記符号化されたブロック及び符号化された予測残差のうちのもの、とともに前記デコーダへ送信するステップ；

を含み、前記符号化されたブロック及び符号化された予測残差のうちの送信されたものは、送信された選択信号に基づいて前記デコーダにより復号化される、方法。

(付記3)

付記1に記載の方法において、前記再構成されたブロックに関してイントラ予測を実行するステップは：

複数の再構成されたブロックを1つ以上のグループに、前記複数の再構成されたブロックの位置に基づいてグループ化するステップ；

1つ以上の潜在空間を生成するために、前記1つ以上のグループの各々に関して個々の変換を実行するステップ；

生成された1つ以上の潜在空間を統合するステップ；及び

前記予測ブロックを生成するために、統合された1つ以上の潜在空間に関して前記イントラ予測を実行するステップ；

を含む、方法。

(付記4)

付記1に記載の方法において、更に：

複数の再構成されたブロックの各々から、アーチファクトを含む境界エリアを取り除くステップ；

1つ以上のニューラル・ネットワークを使用して、取り除かれた境界エリアに関してデブロッキングを実行して、前記アーチファクトを伴わないデブロック化されたエリアを生成するステップ； 及び

前記複数の再構成されたブロックの各々を、生成されたデブロック化されたエリアでアップデートするステップ；

を含む、方法。

(付記5)

付記4に記載の方法において、更に：

前記複数の再構成されたブロックの各々から、取り除かれた境界エリアを伴わない残りのブロック・エリアを取り除くステップ；

エンハンスト・ブロック・エリアを生成するために、少なくとも1つのニューラル・ネットワークを使用して、取り除かれた残りのブロック・エリアに関してポスト・エンハンスメントを実行するステップ；及び

前記複数の再構成されたブロックの各々を、生成されたエンハンスト・ブロック・エリアでアップデートするステップ；

を含む、方法。

(付記6)

付記1に記載の方法において、前記再構成されたブロックに関してイントラ予測を実行するステップは：

前記再構成されたブロックの上にある再構成されたピクセルの第1コンテキスト・エリアと、前記再構成されたブロックの左にある再構成されたピクセルの第2コンテキスト・エリアとを取得するステップ；

潜在表現を生成するために、取得された前記第1コンテキスト・エリアと取得された前記第2コンテキスト・エリアの各々に関して個々の変換を実行するステップ；

生成された潜在表現を統合するステップ；及び

前記予測ブロックを生成するために、統合された潜在表現に関して前記イントラ予測を実行するステップ；

を含む、方法。

(付記7)

付記1に記載の方法において、前記再構成されたブロックに関してイントラ予測を実行するステップは：

前記再構成されたブロックの上にある再構成されたピクセルの第1コンテキスト・エリアと、前記再構成されたブロックの左にある再構成されたピクセルの第2コンテキスト・エリアと、前記第1コンテキスト・エリアの下及び前記第2コンテキスト・エリアの右にあるパディング・エリアとを取得するステップ；

矩形エリアを取得するために、取得された前記第1コンテキスト・エリアと、取得された前記第2コンテキスト・エリアと、取得された前記パディング・エリアを統合するステップ；

潜在表現を生成するために、取得された前記矩形エリアに関して変換を実行するステップ；及び

前記予測ブロックを生成するために、生成された潜在表現に関して前記イントラ予測を実行するステップ；

を含む、方法。

(付記8)

ポスト・フィルタリングを用いるブロック・ワイズ・ニューラル画像圧縮のための装置であって、前記装置はエンコーダとして実装され、前記装置は：

プログラム・コードを記憶するように構成された少なくとも1つのメモリ；及び

前記プログラム・コードを読み込み、前記プログラム・コードにより指示されるとおりに動作するように構成された少なくとも1つのプロセッサ；

を含み、前記プログラム・コードは：

第1ニューラル・ネットワークを使用して、入力画像のブロックを符号化することを、前記少なくとも1つのプロセッサに行わせるように構成された第1符号化コードであって、符号化されたブロックは、再構成されたブロックを生成するために、第2ニューラル・ネ

10

20

30

40

50

ットワークを使用してデコーダによって復号化される、第1符号化コード；

予測ブロックを生成するために、第3ニューラル・ネットワークを使用して、前記再構成されたブロックに関してイントラ予測を実行することを、前記少なくとも1つのプロセッサに行わせるように構成された第1実行コード；

予測残差を生成するために、前記入力画像のブロックと生成された予測ブロックとの間の差分を決定することを、前記少なくとも1つのプロセッサに行わせるように構成された決定コード；

第4ニューラル・ネットワークを使用して、生成された予測残差を符号化することを、前記少なくとも1つのプロセッサに行わせるように構成された第2符号化コードであって、符号化された予測残差は、第5ニューラル・ネットワークを使用して前記デコーダにより復号化される、第2符号化コード；及び

復元された予測ブロックを生成するために、復号化された予測残差を、前記生成された予測ブロックに加えることを、前記少なくとも1つのプロセッサに行わせるように構成された加算コード；

を含む、装置。

(付記9)

付記8に記載の装置において、前記プログラム・コードは、更に：

前記入力画像のブロックに関してより少ない損失をもたらすものを、前記再構成されたブロック及び生成され復元された予測ブロックの中から選択することを、前記少なくとも1つのプロセッサに行わせるように構成された選択コード；及び

前記再構成されたブロック及び復元された予測ブロックの中から選択されたものを示す選択信号を、前記再構成されたブロック及び復元された予測ブロックの中から選択されたものに対応する、前記符号化されたブロック及び符号化された予測残差のうちのもの、とともに前記デコーダへ送信することを、前記少なくとも1つのプロセッサに行わせるように構成された送信コード；

を含み、前記符号化されたブロック及び符号化された予測残差のうちの送信されたものは、送信された選択信号に基づいて前記デコーダにより復号化される、装置。

(付記10)

付記8に記載の装置において、前記第1実行コードは、更に：

複数の再構成されたブロックを1つ以上のグループに、前記複数の再構成されたブロックの位置に基づいてグループ化すること；

1つ以上の潜在空間を生成するために、前記1つ以上のグループの各々に関して個々の変換を実行すること；

生成された1つ以上の潜在空間を統合すること；及び

前記予測ブロックを生成するために、統合された1つ以上の潜在空間に関して前記イントラ予測を実行すること；

を、前記少なくとも1つのプロセッサに行わせるように構成されている、装置。

(付記11)

付記8に記載の装置において、前記プログラム・コードは、更に：

複数の再構成されたブロックの各々から、アーチファクトを含む境界エリアを取り除くことを、前記少なくとも1つのプロセッサに行わせるように構成された第1除去コード；

1つ以上のニューラル・ネットワークを使用して、取り除かれた境界エリアに関してデブロッキングを実行して、前記アーチファクトを伴わないデブロック化されたエリアを生成することを、前記少なくとも1つのプロセッサに行わせるように構成された第2実行コード；及び

前記複数の再構成されたブロックの各々を、生成されたデブロック化されたエリアでアップデートすることを、前記少なくとも1つのプロセッサに行わせるように構成された第1アップデート・コード；

を含む、装置。

(付記12)

10

20

30

40

50

付記11に記載の装置において、前記プログラム・コードは、更に：

前記複数の再構成されたブロックの各々から、取り除かれた境界エリアを伴わない残りのブロック・エリアを取り除くことを、前記少なくとも1つのプロセッサに行わせるように構成された第2除去コード；

エンハンスド・ブロック・エリアを生成するために、少なくとも1つのニューラル・ネットワークを使用して、取り除かれた残りのブロック・エリアに関してポスト・エンハンスメントを実行することを、前記少なくとも1つのプロセッサに行わせるように構成された第3実行コード；及び

前記複数の再構成されたブロックの各々を、生成されたエンハンスド・ブロック・エリアでアップデートすることを、前記少なくとも1つのプロセッサに行わせるように構成された第2アップデート・コード；

を含む、装置。

(付記13)

付記8に記載の装置において、前記第1実行コードは、更に：

前記再構成されたブロックの上にある再構成されたピクセルの第1コンテキスト・エリアと、前記再構成されたブロックの左にある再構成されたピクセルの第2コンテキスト・エリアとを取得すること；

潜在表現を生成するために、取得された前記第1コンテキスト・エリアと取得された前記第2コンテキスト・エリアの各々に関して個々の変換を実行すること；

生成された潜在表現を統合すること；及び

前記予測ブロックを生成するために、統合された潜在表現に関して前記イントラ予測を実行すること；

を、前記少なくとも1つのプロセッサに行わせるように構成されている、装置。

(付記14)

付記8に記載の装置において、前記第1実行コードは、更に：

前記再構成されたブロックの上にある再構成されたピクセルの第1コンテキスト・エリアと、前記再構成されたブロックの左にある再構成されたピクセルの第2コンテキスト・エリアと、前記第1コンテキスト・エリアの下及び前記第2コンテキスト・エリアの右にあるパディング・エリアとを取得すること；

矩形エリアを取得するために、取得された前記第1コンテキスト・エリアと、取得された前記第2コンテキスト・エリアと、取得された前記パディング・エリアを統合すること；

潜在表現を生成するために、取得された前記矩形エリアに関して変換を実行すること；及び

前記予測ブロックを生成するために、生成された潜在表現に関して前記イントラ予測を実行すること；

を、前記少なくとも1つのプロセッサに行わせるように構成されている、装置。

(付記15)

命令を記憶する非一時的なコンピュータ読み取り可能な媒体であって、前記命令は、エンコーダのポスト・フィルタリングを用いるブロック・ワイズ・ニューラル画像圧縮のために、少なくとも1つのプロセッサにより実行されると、前記少なくとも1つのプロセッサに：

第1ニューラル・ネットワークを使用して、入力画像のブロックを符号化するステップであって、符号化されたブロックは、再構成されたブロックを生成するために、第2ニューラル・ネットワークを使用してデコーダによって復号化される、ステップ；

予測ブロックを生成するために、第3ニューラル・ネットワークを使用して、前記再構成されたブロックに関してイントラ予測を実行するステップ；

予測残差を生成するために、前記入力画像のブロックと生成された予測ブロックとの間の差分を決定するステップ；

第4ニューラル・ネットワークを使用して、生成された予測残差を符号化するステップであって、符号化された予測残差は、第5ニューラル・ネットワークを使用して前記デコ

10

20

30

40

50

ーダにより復号化される、ステップ；及び

復元された予測ブロックを生成するために、復号化された予測残差を、前記生成された予測ブロックに加えるステップ；

を行わせる、非一時的なコンピュータ読み取り可能な媒体。

(付記 16)

付記15に記載の非一時的なコンピュータ読み取り可能な媒体において、前記命令は、前記少なくとも1つのプロセッサにより実行されると、前記少なくとも1つのプロセッサに、更に：

前記入力画像のブロックに関してより少ない損失をもたらすものを、前記再構成されたブロック及び生成され復元された予測ブロックの中から選択するステップ；及び

前記再構成されたブロック及び復元された予測ブロックの中から選択されたものを示す選択信号を、前記再構成されたブロック及び復元された予測ブロックの中から選択されたものに対応する、前記符号化されたブロック及び符号化された予測残差のうちのもの、とともに前記デコーダへ送信するステップ；

を行わせ、前記符号化されたブロック及び符号化された予測残差のうちの送信されたものは、送信された選択信号に基づいて前記デコーダにより復号化される、非一時的なコンピュータ読み取り可能な媒体。

(付記 17)

付記15に記載の非一時的なコンピュータ読み取り可能な媒体において、前記命令は、前記少なくとも1つのプロセッサにより実行されると、前記少なくとも1つのプロセッサに、更に：

複数の再構成されたブロックを1つ以上のグループに、前記複数の再構成されたブロックの位置に基づいてグループ化するステップ；

1つ以上の潜在空間を生成するために、前記1つ以上のグループの各々に関して個々の変換を実行するステップ；

生成された1つ以上の潜在空間を統合するステップ；及び

前記予測ブロックを生成するために、統合された1つ以上の潜在空間に関して前記エンコードを実行するステップ；

を行わせる、非一時的なコンピュータ読み取り可能な媒体。

(付記 18)

付記15に記載の非一時的なコンピュータ読み取り可能な媒体において、前記命令は、前記少なくとも1つのプロセッサにより実行されると、前記少なくとも1つのプロセッサに、更に：

複数の再構成されたブロックの各々から、アーチファクトを含む境界エリアを取り除くステップ；

1つ以上のニューラル・ネットワークを使用して、取り除かれた境界エリアに関してデブロッキングを実行して、前記アーチファクトを伴わないデブロック化されたエリアを生成するステップ； 及び

前記複数の再構成されたブロックの各々を、生成されたデブロック化されたエリアでアップデートするステップ；

を行わせる、非一時的なコンピュータ読み取り可能な媒体。

(付記 19)

付記18に記載の非一時的なコンピュータ読み取り可能な媒体において、前記命令は、前記少なくとも1つのプロセッサにより実行されると、前記少なくとも1つのプロセッサに、更に：

前記複数の再構成されたブロックの各々から、取り除かれた境界エリアを伴わない残りのブロック・エリアを取り除くステップ；

エンハンスド・ブロック・エリアを生成するために、少なくとも1つのニューラル・ネットワークを使用して、取り除かれた残りのブロック・エリアに関してポスト・エンハンスメントを実行するステップ；及び

10

20

30

40

50

前記複数の再構成されたブロックの各々を、生成されたエンハンスド・ブロック・エリアでアップデートするステップ；

を行わせる、非一時的なコンピュータ読み取り可能な媒体。

(付記 20)

付記15に記載の非一時的なコンピュータ読み取り可能な媒体において、前記命令は、前記少なくとも1つのプロセッサにより実行されると、前記少なくとも1つのプロセッサに、更に：

前記再構成されたブロックの上にある再構成されたピクセルの第1コンテキスト・エリアと、前記再構成されたブロックの左にある再構成されたピクセルの第2コンテキスト・エリアとを取得するステップ；

潜在表現を生成するために、取得された前記第1コンテキスト・エリアと取得された前記第2コンテキスト・エリアの各々に関して個々の変換を実行するステップ；

生成された潜在表現を統合するステップ；及び

前記予測ブロックを生成するために、統合された潜在表現に関して前記イントラ予測を実行するステップ；

を行わせる、非一時的なコンピュータ読み取り可能な媒体。

【図面】

【図 1】

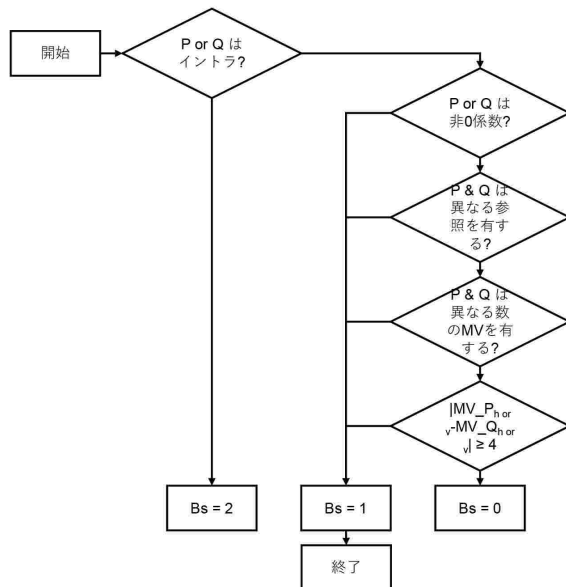


FIG. 1

【図 2】

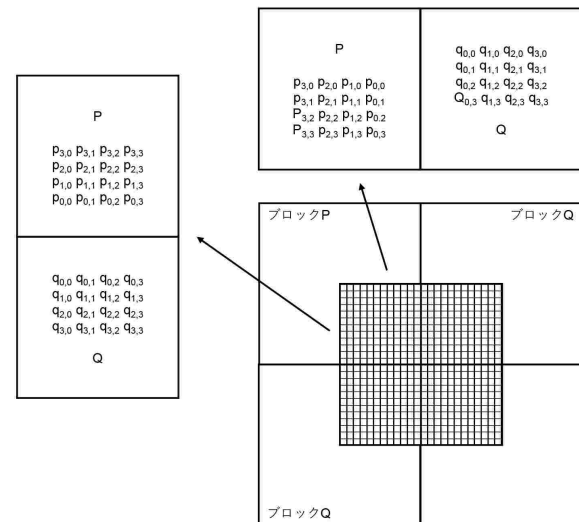


FIG. 2

10

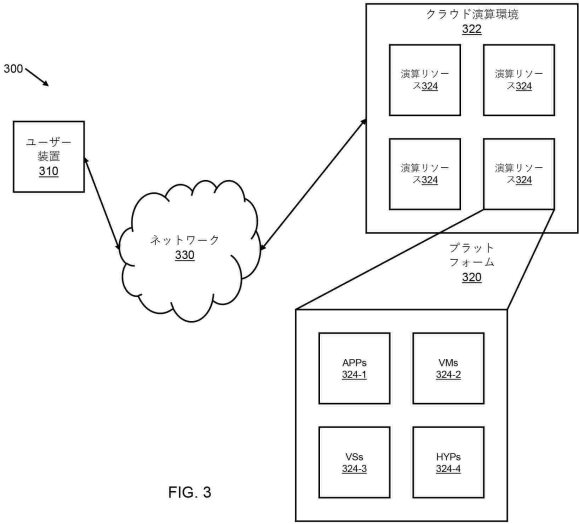
20

30

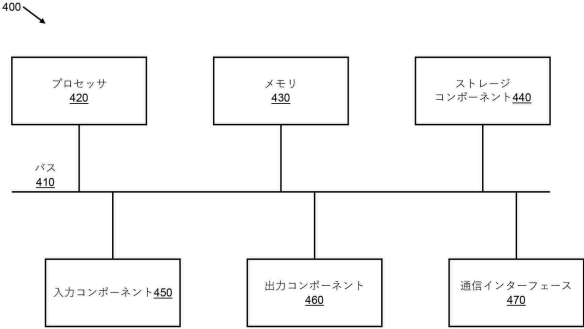
40

50

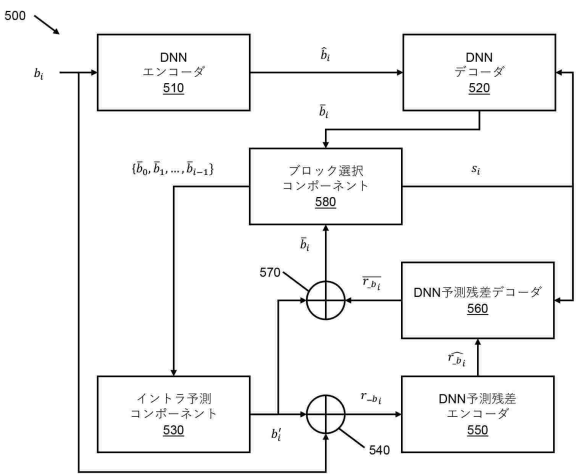
【図 3】



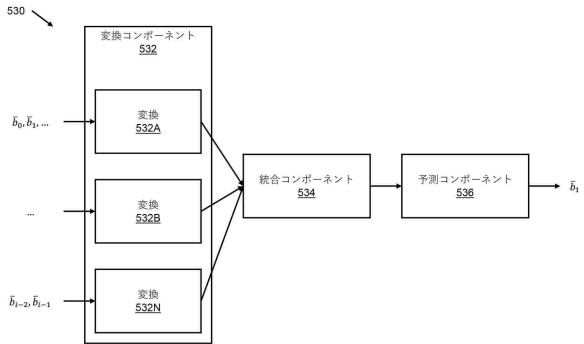
【図 4】



【図 5 A】



【図 5 B】



10

20

30

40

50

【図 6】

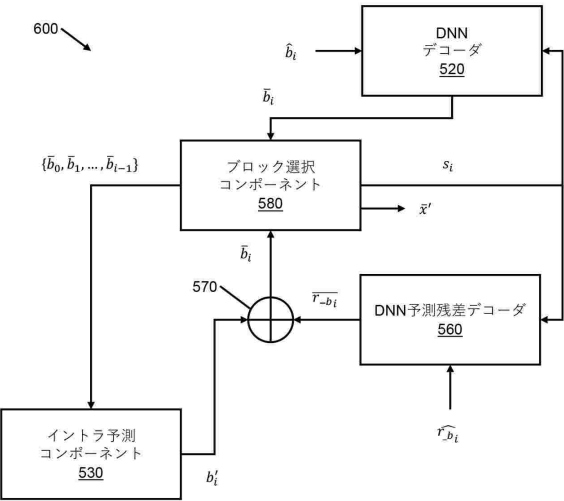


FIG. 6

【図 7】

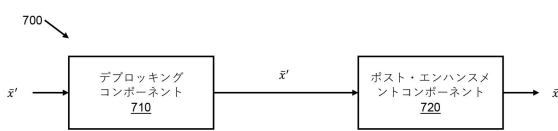


FIG. 7

【図 8】

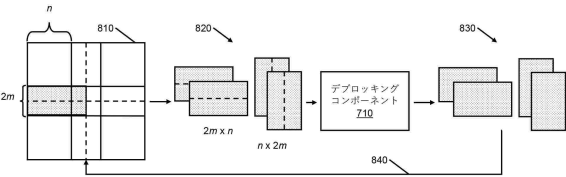


FIG. 8

【図 9】

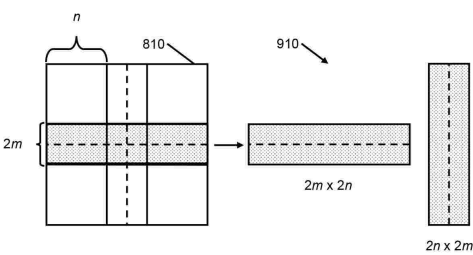


FIG. 9

10

20

30

40

50

【図 10】

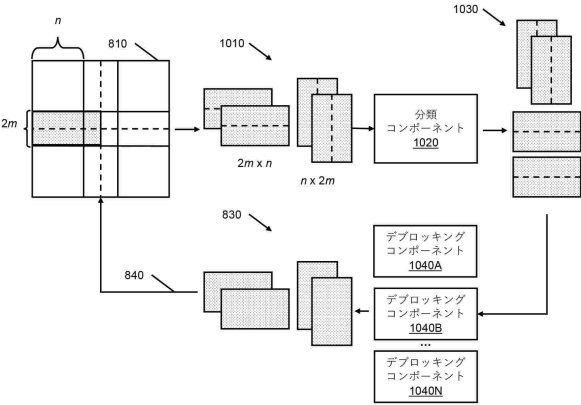


FIG. 10

【図 11】

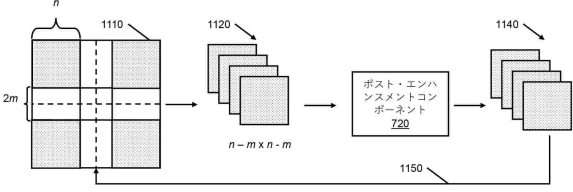


FIG. 11

【図 12】

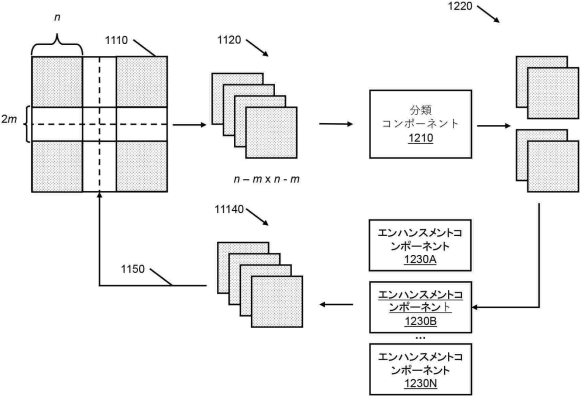


FIG. 12

【図 13】

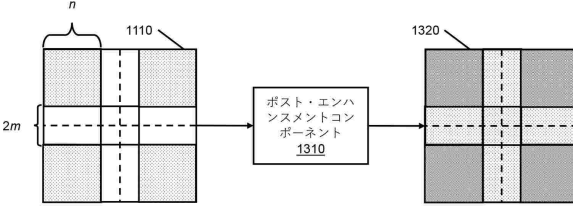


FIG. 13

10

20

30

40

50

【図 1 4】

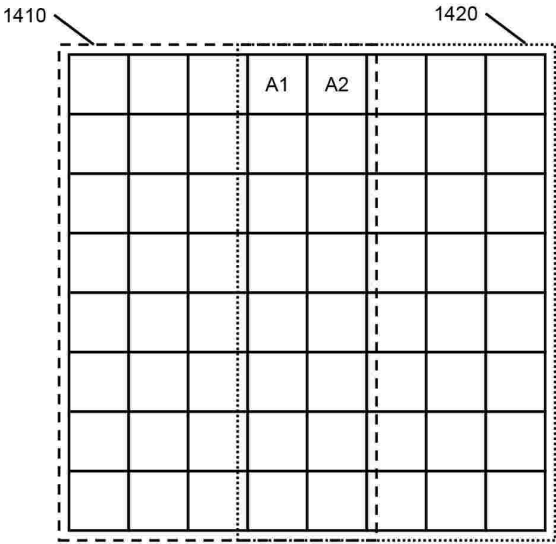


FIG. 14

【図 1 5】

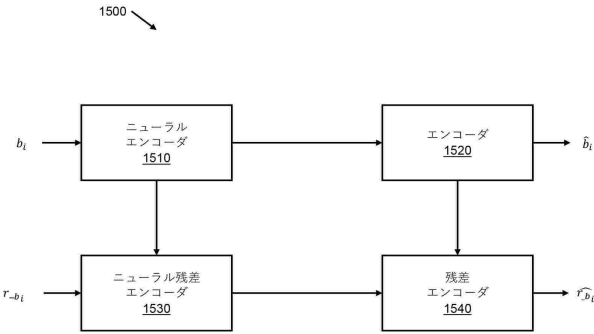


FIG. 15

【図 1 6】

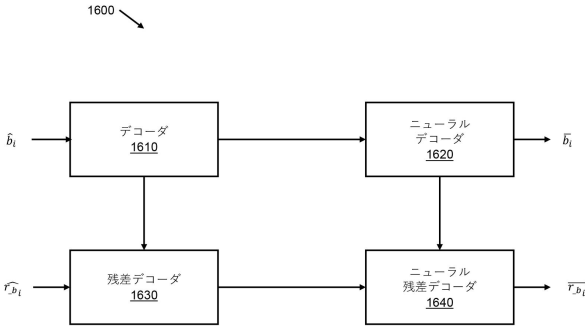


FIG. 16

【図 1 7 A】

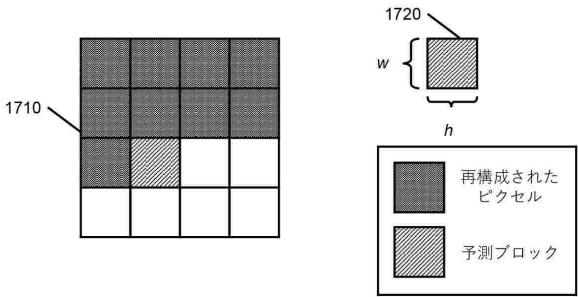


FIG. 17A

10

20

30

40

50

【図 17 B】

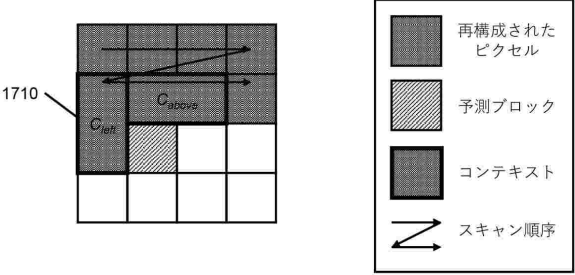


FIG. 17B

【図 17 C】

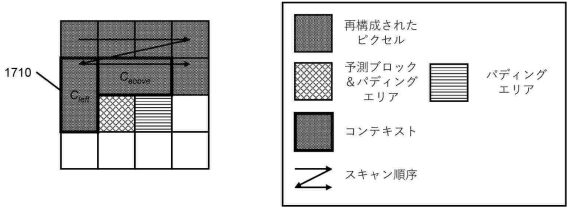


FIG. 17C

【図 17 D】

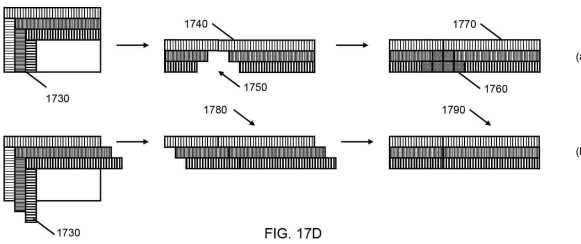


FIG. 17D

【図 18】

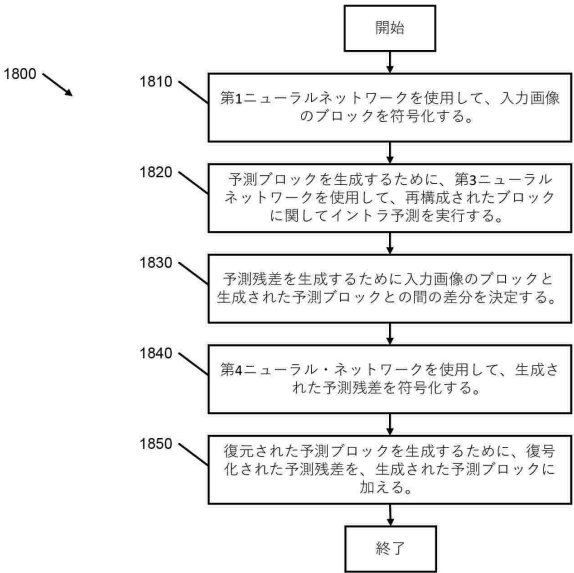


FIG. 18

【図 19】

1900 ↘

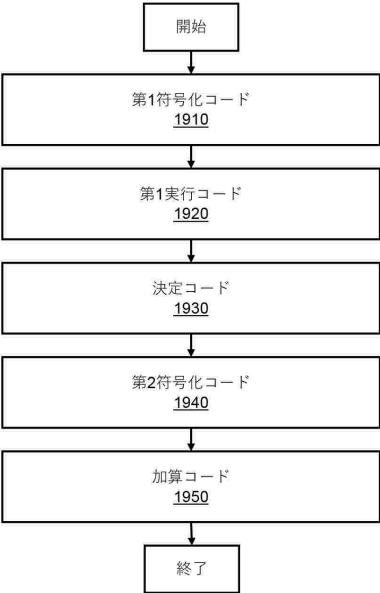


FIG. 19

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)

(31)優先権主張番号 63/088,675

(32)優先日 令和2年10月7日(2020.10.7)

(33)優先権主張国・地域又は機関

米国(US)

(31)優先権主張番号 17/362,003

(32)優先日 令和3年6月29日(2021.6.29)

(33)優先権主張国・地域又は機関

米国(US)

エルシー

(72)発明者 ジアン, ウエイ

アメリカ合衆国 9 4 3 0 6 カリフォルニア州 パロアルト パーク・ブールバード 2 7 4 7 テンセント アメリカ エルエルシー

(72)発明者 ワン, ウエイ

アメリカ合衆国 9 4 3 0 6 カリフォルニア州 パロアルト パーク・ブールバード 2 7 4 7 テンセント アメリカ エルエルシー

(72)発明者 リウ, シャン

アメリカ合衆国 9 4 3 0 6 カリフォルニア州 パロアルト パーク・ブールバード 2 7 4 7 テンセント アメリカ エルエルシー

(72)発明者 シュイ, シアオジョオン

アメリカ合衆国 9 4 3 0 6 カリフォルニア州 パロアルト パーク・ブールバード 2 7 4 7 テンセント アメリカ エルエルシー

審査官 間宮 嘉誉

(56)参考文献 特開平10-210474 (JP, A)

特開平5-308629 (JP, A)

特開2018-125713 (JP, A)

特開2019-201256 (JP, A)

国際公開第2019/093234 (WO, A1)

米国特許出願公開第2020/0244997 (US, A1)

国際公開第2019/115865 (WO, A1)

米国特許出願公開第2019/0306526 (US, A1)

CHEN, Tong et al., DeepCoder: A Deep Neural Network Based Video Compression, Proc. IEEE Visual Communications and Image Processing 2017, 米国, IEEE, 2018年03月01日, pp.1-4, [online], [retrieved on 2023-08-25], Retrieved from the Internet: <URL: <https://ieeexplore.ieee.org/document/8305033>>, <https://doi.org/10.1109/VCIP.2017.8305033>HUSSAIN, Farhan et al., Exploiting Deep Neural Networks for Digital Image Compression, Proc. 2nd World Symposium on Web Applications and Networking 2015, TN, IEEE, 2015年08月20日, pp.1-6, [online], [retrieved on 2023-08-25], Retrieved from the Internet: <URL: <https://ieeexplore.ieee.org/document/7210294>>, <https://doi.org/10.1109/WSWAN.2015.7210294>PFAFF, Jonathan et al., Intra Prediction Modes based on Neural Networks, JVET-J0037 (version 2), ITU, 2018年04月13日, pp.1-14, [online], [retrieved on 2023-08-25], Retrieved from the Internet: <URL: https://jvet-experts.org/doc_end_user/documents/10_San%20Diego/wg11/JVET-J0037-v2.zip>, JVET-J0037.docx

(58)調査した分野 (Int.Cl., DB名)

H04N 7/12

H04N 19/00-19/98

IEEE Xplore