

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7196331号
(P7196331)

(45)発行日 令和4年12月26日(2022.12.26)

(24)登録日 令和4年12月16日(2022.12.16)

(51)国際特許分類	F I	
G 0 6 N 3/08 (2006.01)	G 0 6 N 3/08	
G 0 6 N 3/04 (2006.01)	G 0 6 N 3/04	1 5 4
G 0 6 T 9/00 (2006.01)	G 0 6 T 9/00	2 0 0
H 0 4 N 19/46 (2014.01)	H 0 4 N 19/46	
G 1 0 L 19/00 (2013.01)	G 1 0 L 19/00	3 3 0 B
請求項の数 19 (全27頁)		

(21)出願番号	特願2021-549873(P2021-549873)	(73)特許権者	510185767
(86)(22)出願日	令和2年3月5日(2020.3.5)		ドルビー・インターナショナル・アーベ
(65)公表番号	特表2022-522685(P2022-522685 A)		アイランド ディー02 ヴィーケイ6 0 ダブリン グランド カナル ドックラ ンズ サー ジョン ロジャーソンズ キー 77 ブロック シー
(43)公表日	令和4年4月20日(2022.4.20)	(74)代理人	100107766
(86)国際出願番号	PCT/EP2020/055869		弁理士 伊東 忠重
(87)国際公開番号	WO2020/187587	(74)代理人	100070150
(87)国際公開日	令和2年9月24日(2020.9.24)		弁理士 伊東 忠彦
審査請求日	令和3年8月25日(2021.8.25)	(74)代理人	100135079
(31)優先権主張番号	62/818,879		弁理士 宮崎 修
(32)優先日	平成31年3月15日(2019.3.15)	(72)発明者	フェルシュ, クリストフ
(33)優先権主張国・地域又は機関	米国(US)		ドイツ 9 0 4 2 9 ニュルンベルク ド
(31)優先権主張番号	19174542.1		最終頁に続く
(32)優先日	令和1年5月15日(2019.5.15)		最終頁に続く

(54)【発明の名称】 ニューラルネットワークを更新するための方法および装置

(57)【特許請求の範囲】

【請求項1】

デコーダにおいて実装されたニューラルネットワークを更新するためのパラメータを伝送するためのメディアビットストリームを生成する方法であって、前記ニューラルネットワークは少なくとも3つの層を有し、前記少なくとも3つの層のうちの最初の層としてメディアデータ側層をもち、前記少なくとも3つの層のうちの最後の層として出力層をもち、当該方法は：

(a) 前記メディアデータ側層および/または前記出力層の重みを更新するためのパラメータを含む、前記ニューラルネットワークの前記少なくとも3つの層の重みを更新するためのパラメータの少なくとも1つの集合を決定するステップと；

(b) 前記ニューラルネットワークの前記少なくとも3つの層の重みを更新するためのパラメータの前記少なくとも1つの集合のうち、前記メディアデータ側層および/または前記出力層の重みを更新するためのパラメータのみと、オーディオデータおよび/またはビデオデータの一方または複数を含むメディアデータとをエンコードすることによって前記メディアビットストリームを生成するステップと；

(c) 前記メディアデータ側層および/または前記出力層の重みを更新するためのパラメータを用いて前記ニューラルネットワークを更新するために、前記デコーダに前記メディアビットストリームを伝送するステップとを含む、方法。

【請求項2】

パラメータの前記少なくとも1つの集合は、構文要素の集合に基づいてエンコードされる、請求項1に記載の方法。

【請求項3】

ステップ(a)において、前記ニューラルネットワークを更新するためのパラメータの2つ以上の集合が決定され、構文要素の前記集合は、実行される前記ニューラルネットワークのそれぞれの更新のためのパラメータのそれぞれの集合を同定する一つまたは複数の構文要素を含む、請求項2に記載の方法。

【請求項4】

前記デコーダにおいて実装される前記ニューラルネットワークは、メディアデータの処理のために使用され、前記メディアビットストリームにおいて、前記ニューラルネットワークを更新するためのパラメータの前記少なくとも1つの集合は、前記ニューラルネットワークによって処理されるメディアデータと時間整列される、請求項1ないし3のうちいずれか一項に記載の方法。

10

【請求項5】

パラメータの前記少なくとも1つの集合は、コーデック・モード、前記メディアデータの内容、およびエンコード制約条件のうちの一つまたは複数に基づいて決定される、請求項4に記載の方法。

【請求項6】

前記コーデック・モードは、ビットレート、ビデオおよび/またはオーディオ・フレームレート、および使用されるコア・コーデックのうちの一つまたは複数を含む、請求項5に記載の方法。

20

【請求項7】

メディアデータの内容は、発話、音楽、および拍手のうちの一つまたは複数を含む、請求項5または請求項6に記載の方法。

【請求項8】

前記エンコード制約条件は、パフォーマンス・スケーラビリティについての制約条件および適応処理についての制約条件のうちの一つまたは複数を含む、請求項5ないし7のうちいずれか一項に記載の方法。

【請求項9】

パラメータの前記少なくとも1つの集合は、前記メディアビットストリームにおいて、それぞれの更新されたニューラルネットワークによって処理されるべきメディアデータより前に含まれる、請求項5ないし8のうちいずれか一項に記載の方法。

30

【請求項10】

前記メディアデータはMPEG-HオーディオまたはMPEG-Iオーディオ・フォーマットであり、前記メディアビットストリームはMHASフォーマットの packets 化されたメディアビットストリームである、請求項1ないし9のうちいずれか一項に記載の方法。

【請求項11】

パラメータの前記少なくとも1つの集合は、パラメータの前記少なくとも1つの集合を、新規のMHAS packets タイプの一つまたは複数のMHAS packets にカプセル化することによってエンコードされる、請求項10に記載の方法。

40

【請求項12】

前記メディアデータが、AC-4、AC-3、EAC-3フォーマット、MPEG-4またはMPEG-D USACフォーマットである、請求項1ないし11のうちいずれか一項に記載の方法。

【請求項13】

パラメータの前記少なくとも1つの集合は、一つまたは複数のペイロード要素として前記メディアビットストリームにおいてエンコードされる、請求項12に記載の方法。

【請求項14】

パラメータの前記少なくとも1つの集合は、一つまたは複数のペイロード要素または一つまたは複数のデータストリーム要素として、前記メディアビットストリームにおいてエンコードされる、請求項13に記載の方法。

50

【請求項 15】

パラメータの前記少なくとも1つの集合は、重みを更新するための前記パラメータが相対値を表すか絶対値を表すかを識別する識別子を含む、請求項1ないし14のうちいずれか一項に記載の方法。

【請求項 16】

デコーダにおいて実装されるニューラルネットワークを更新する方法であって、前記ニューラルネットワークは少なくとも3つの層を有し、前記少なくとも3つの層のうちの最初の層としてメディアデータ側層をもち、前記少なくとも3つの層のうちの最後の層として出力層をもち、当該方法は：

(a) メディアデータと、前記ニューラルネットワークの前記メディアデータ側層および/または前記出力層の重みを更新するためのパラメータとを含む符号化メディアビットストリームを受領するステップと；

10

(b) 受領されたメディアビットストリームをデコードして、デコードされたメディアデータと、前記ニューラルネットワークの前記メディアデータ側層および/または前記出力層の重みを更新するためのパラメータとを取得するステップと；

(c) 前記デコーダによって、前記ニューラルネットワークの前記メディアデータ側層および/または前記出力層の重みを更新するための受領されたパラメータを用いて前記メディアデータ側層および/または前記出力層を更新するステップとを含む、方法。

【請求項 17】

20

デコーダにおいて実装されるニューラルネットワークを更新するためのパラメータを伝送するためのメディアビットストリームを生成するための装置であって、前記ニューラルネットワークは少なくとも3つの層を有し、前記少なくとも3つの層のうちの最初の層としてメディアデータ側層をもち、前記少なくとも3つの層のうちの最後の層として出力層をもち、当該装置は：

(a) 前記メディアデータ側層および/または前記出力層の重みを更新するためのパラメータを含む、前記ニューラルネットワークの前記少なくとも3つの層の重みを更新するためのパラメータの少なくとも1つの集合を決定するステップと；

(b) 前記ニューラルネットワークの前記少なくとも3つの層の重みを更新するためのパラメータの前記少なくとも1つの集合のうち、前記メディアデータ側層および/または前記出力層の重みを更新するためのパラメータのみと、オーディオデータおよび/またはビデオデータの一つまたは複数を含むメディアデータとをエンコードすることによって前記メディアビットストリームを生成するステップと；

30

(c) 前記メディアデータ側層および/または前記出力層の重みを更新するためのパラメータを用いて前記ニューラルネットワークを更新するために、前記デコーダに前記メディアビットストリームを伝送するステップとを含む方法を実行するように構成されたプロセッサを含む、装置。

【請求項 18】

デコーダにおいて実装されるニューラルネットワークを更新するための装置であって、前記ニューラルネットワークは少なくとも3つの層を有し、前記少なくとも3つの層のうちの最初の層としてメディアデータ側層をもち、前記少なくとも3つの層のうちの最後の層として出力層をもち、当該装置は：

40

(a) メディアデータと、前記ニューラルネットワークの前記メディアデータ側層および/または前記出力層の重みを更新するためのパラメータとを含む符号化メディアビットストリームを受領するように構成された受領器と；

(b) 受領されたメディアビットストリームをデコードして、デコードされたメディアデータと、前記ニューラルネットワークの前記メディアデータ側層および/または前記出力層の重みを更新するためのパラメータとを取得するように構成されたデコーダと；

(c) 前記ニューラルネットワークの前記メディアデータ側層および/または前記出力層の

50

重みを更新するための受領されたパラメータを用いて前記メディアデータ側層および/または前記出力層を更新するように構成された更新器とを含む、装置。

【請求項 19】

処理能力を有する装置に、請求項 1 ないし 15 のうちいずれか一項に記載の方法を実行させるためのコンピュータ・プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願への相互参照

本願は、以下の優先権出願の優先権を主張する：すなわち、2019年3月15日に出願された米国仮出願第62/818,879号（参照番号:D19009USP1）および2019年5月15日に出願された欧州特許出願第19174542.1号（参照番号:D19009EP）であり、これらはここに参照により組み込まれる。

【0002】

技術

本開示は、概括的には、デコーダ内のニューラルネットワークを更新するためのパラメータを伝送するためのメディアビットストリームを生成すること、およびデコーダによって前記ニューラルネットワークを更新することのための方法および装置に関する。

【0003】

いくつかの実施形態が、その開示を特に参照して本明細書に記載されるが、本開示は、そのような使用分野に限定されず、より広い文脈において適用可能であることが理解されるであろう。

【背景技術】

【0004】

本開示を通じた背景技術のいかなる議論も、当該技術が広く知られている、または当該技術分野における技術常識の一部を構成するものであるとの自認とみなされるべきではない。

【0005】

近年、ニューラルネットワークは、継続的に信頼性を改善しているため、ますます注目を集めている。

【0006】

一般に、深層ニューラルネットワークの生成器（Generator）は、いわゆる弁別器（Discriminator）との真贋（real-fake）ベースの対話によってトレーニングされる。特に、トレーニング中に、生成器は、いわゆる偽のデータを生成するために使用され、弁別器は、この偽のデータを本物のデータから区別しようと試みる。この手順を逐次反復することによって、生成器は、最終的には、本物のデータに近い（弁別器がそれを本物のデータと区別できない程度まで）偽のデータを生成するようにトレーニングされる。最近の研究は、主に深層畳み込み敵対的生成ネットワーク（Generative Adversarial Network、GAN）に基づいている。GANは、たとえば、多様なコンピュータビジョンおよび画像処理のタスクにおいて、また発話およびオーディオ関連アプリケーションにおいてもすでに成功裏に使用されている。

【0007】

ひとたび深層ニューラルネットワーク（たとえば、深層ニューラルネットワーク生成器）が特定のタスクでうまくトレーニングされると、それは、たとえば、デコーダにおいて、その特定のタスクのために適用されうる。しかしながら、デコーダにおいて実装され、エンコーダによって提供されるビットストリームから抽出されたメディアコンテンツに作用するトレーニングされたニューラルネットワークについては、ニューラルネットワークは、エンコーダにおいてメディアコンテンツをエンコードするとき存在していた特定のエンコード制約条件に依存して、満足できる結果をもたらすこともあれば、もたらさない

10

20

30

40

50

こともある。これは、エンコード制約条件が時間の経過とともに変化する可能性があるため、一層当てはまる。たとえば、エンコードされたメディアコンテンツのビットレートまたはフレームレートの変化は、ニューラルネットワークによるメディアコンテンツの処理の結果の品質に影響を与えることがある。

【発明の概要】

【発明が解決しようとする課題】

【0008】

このように、デコーダで実装されるニューラルネットワークを特定の要件に適応させる際に、より大きな柔軟性を許容する方法および装置が必要とされている。メディアコンテンツを処理するデコーダで実装されるニューラルネットワークが、メディアコンテンツをエンコードする段階において、可変なエンコード制約条件に対処できるようにする方法および装置が特に必要とされている。

10

【課題を解決するための手段】

【0009】

上記の必要性を考慮して、本開示は、デコーダにおいて実装されたニューラルネットワークを更新するためのパラメータを伝送するためのメディアビットストリームを生成する方法、デコーダにおいて実装されたニューラルネットワークを更新する方法、ならびに対応する装置およびコンピュータ・プログラム・プロダクトであって、それぞれの独立請求項の特徴を有するものを提供する。

【0010】

本開示の第1の側面によれば、デコーダにおいて実装されたニューラルネットワークを更新するためのパラメータを伝送するためのメディアビットストリームを生成する方法（たとえば、そのようなメディアビットストリームを伝送する方法）が提供される。本方法は、（a）ニューラルネットワークを更新するためのパラメータの少なくとも1つの集合を決定するステップを含んでいてもよい。本方法はさらに、（b）パラメータの前記少なくとも1つの集合およびメディアデータをエンコードしてメディアビットストリームを生成するステップを含んでいてもよい。これは、パラメータの前記少なくとも1つの集合およびメディアデータをビットストリーム中にエンコードすることに関わってもよい。本方法はさらに、（c）ニューラルネットワークをパラメータの前記少なくとも1つの集合で（たとえば、それを使用して）更新するために、デコーダにメディアビットストリームを伝送（出力）するステップを含んでいてもよい。メディアビットストリームを出力/伝送することは、有線または無線接続を介してメディアビットストリームを伝送すること、またはメディアビットストリームを有形な媒体上に記憶し、該有形な媒体をデコーダに提供することを含みうる。

20

【0011】

いくつかの実施形態では、メディアデータは、オーディオデータおよび/またはビデオデータの一つまたは複数を含んでいてもよい。

【0012】

いくつかの実施形態では、パラメータの前記少なくとも1つの集合は、構文要素の集合に基づいて（メディアビットストリーム中に）エンコードされてもよい。これには、構文要素を生成（作成）することに関わってもよい。

30

40

【0013】

いくつかの実施形態では、ステップ（a）において、ニューラルネットワークを更新するためのパラメータの2つ以上の集合が決定されてもよく、構文要素の前記集合は、実行されるニューラルネットワークのそれぞれの更新のためのパラメータのそれぞれの集合を同定する一つまたは複数の構文要素を含んでいてもよい。

【0014】

いくつかの実施形態では、デコーダにおいて実装されるニューラルネットワークは、メディアデータの処理のために使用されてもよく、メディアビットストリームにおいて、ニューラルネットワークを更新するためのパラメータの前記少なくとも1つの集合は、（た

50

例えば、ニューラルネットワークを更新するためにパラメータの前記少なくとも1つの集合を使用した後に)ニューラルネットワークによって処理されるメディアデータ(のその部分)と時間整列されてもよい。たとえば、メディアデータは、メディアデータの諸部分(たとえば、メディアデータの諸フレーム)に分割されてもよい。次いで、各部分について、その部分を処理するためにニューラルネットワークを更新するためのパラメータのそれぞれの少なくとも1つの集合が、その部分と時間整列されてもよい。フレームベースのメディアデータについては、各フレームは、メディアデータと、そのメディアデータを処理する際に使用できる対応するパラメータの少なくとも1つの集合とを含んでいてもよい。

【0015】

いくつかの実施形態では、パラメータの前記少なくとも1つの集合は、コーデック・モード、メディアデータの内容、およびエンコード制約条件のうちの一つまたは複数に基づいて決定されてもよい。いくつかの実施形態では、コーデック・モードは、ビットレート、ビデオおよび/またはオーディオ・フレームレート、および使用されるコア・コーデック(たとえば、ASFまたはSSFなど)のうちの一つまたは複数を含んでいてもよい。いくつかの実施形態では、メディアデータの内容は、発話、音楽、および拍手のうちの一つまたは複数を含んでいてもよい。いくつかの実施形態では、エンコード制約条件は、パフォーマンス・スケーラビリティについての制約条件および適応処理についての(たとえば、適応ストリーミングについての)制約条件のうちの一つまたは複数を含んでいてもよい。いくつかの実施形態では、パラメータの前記少なくとも1つの集合が、それぞれの更新されたニューラルネットワークによって処理されるメディアデータに先立って、メディアビットストリームに含まれる。

【0016】

いくつかの実施形態では、メディアデータはMPEG-HオーディオまたはMPEG-Iオーディオ・フォーマットであってもよく、メディアビットストリームはMHASフォーマットの packets 化されたメディアビットストリームであってもよい。パラメータ(更新パラメータ)の前記少なくとも1つの集合および対応するメディアデータは、メディアビットストリーム内に一緒にパッケージングされてもよい。

【0017】

いくつかの実施形態では、パラメータの前記少なくとも1つの集合は、パラメータの前記少なくとも1つの集合を、(可能性としては新しい)MHAS packet タイプの一つまたは複数のMHAS packet にカプセル化することによってエンコードされてもよい。

【0018】

いくつかの実施形態において、メディアデータは、AC-4、AC-3またはEAC-3フォーマットであってもよい。

【0019】

いくつかの実施形態では、パラメータの前記少なくとも1つの集合は、一つまたは複数のペイロード要素としてメディアビットストリームにおいてエンコードされてもよい。

【0020】

いくつかの実施形態では、メディアデータは、MPEG-4またはMPEG-D USACフォーマットであってもよい。

【0021】

いくつかの実施形態では、パラメータの前記少なくとも1つの集合は、一つまたは複数のペイロード要素として、または一つまたは複数のデータストリーム要素として、メディアビットストリームにおいてエンコードされてもよい。

【0022】

いくつかの実施形態では、パラメータの前記少なくとも1つの集合は、重みを更新するためのパラメータが相対値を表すのか絶対値を表すのかを識別する識別子を含む、ニューラルネットワークの一つまたは複数の層の重みを更新するためのパラメータを含む。

【0023】

いくつかの実施形態では、ニューラルネットワークは、層構造を有してもよく、パラメ

10

20

30

40

50

ータの前記少なくとも1つの集合は、ニューラルネットワークのメディアデータ側層（すなわち、複数の層のうちの最初の層）および/または出力層（すなわち、前記複数の層のうちの最後の層）の重みを少なくとも更新するためのパラメータを含んでいてもよい。一例では、ニューラルネットワークは、3つ以上の層を有していてもよく、メディアデータ側層と出力層との間に少なくとも1つの層があってもよい。

【0024】

いくつかの実施形態では、メディアビットストリームは、ニューラルネットワークの複数の層の重みを更新するためのパラメータの前記少なくとも1つの集合のうち、メディアデータ側層および/または出力層の重みを更新するためのパラメータおよびメディアデータのみをエンコードすることによって生成されてもよい。言い換えると、他の層の重みを更新するためのパラメータはビットストリームにおいてエンコードされない。第1の例では、デコーダにおいて実装されるニューラルネットワークのすべての層の重みを更新するためのパラメータが決定されるが、メディアデータ側層および/または出力層の重みを更新するためのパラメータのみがメディアビットストリームにおいてエンコードされる。第2の例では、メディアデータ側層および/または出力層の重みを更新するためのパラメータのみが決定される。すなわち、他の層についてはパラメータが決定されない。

10

【0025】

本開示の第2の側面によれば、デコーダにおいて実装されるニューラルネットワークを更新するための方法が提供される。本方法は、(a)メディアデータと、ニューラルネットワークを更新するためのパラメータの少なくとも1つの集合とを含む符号化メディアビットストリームを受領するステップを含んでいてもよい。本方法はさらに、(b)受領されたメディアビットストリームをデコードして、デコードされたメディアデータと、ニューラルネットワークを更新するためのパラメータの前記少なくとも1つの集合とを取得するステップとを含んでいてもよい。本方法は、さらに、(c)デコーダによって、パラメータの前記少なくとも1つの集合で（たとえば、それを使用して）ニューラルネットワークを更新するステップを含んでいてもよい。

20

【0026】

いくつかの実施形態では、メディアデータは、オーディオデータおよび/またはビデオデータの一つまたは複数を含んでいてもよい。

【0027】

いくつかの実施形態では、本方法は、さらに、構文要素の集合を受領し、デコーダによって、パラメータの前記少なくとも1つの集合で（たとえば、それを使用するして）ニューラルネットワークを更新することを含んでいてもよい。これは、受領された構文要素の集合に基づいてもよい。

30

【0028】

いくつかの実施形態では、ステップ(a)において、ニューラルネットワークを更新するためのパラメータの2つ以上の集合が、受領された符号化メディアビットストリームに含まれていてもよく、受領された構文要素の集合は、実行されるニューラルネットワークのそれぞれの更新のためのパラメータのそれぞれの集合を識別する一つまたは複数の構文要素を含んでいてもよい。

40

【0029】

いくつかの実施形態では、ニューラルネットワークは、メディアデータの処理のために使用されてもよく、受信された符号化メディアビットストリームにおいて、パラメータの前記少なくとも1つの集合は、（たとえば、パラメータの前記少なくとも1つの集合を使用した後で）ニューラルネットワークによって処理されるメディアデータ（のその部分）と時間整列されてもよい。

【0030】

いくつかの実施形態では、ニューラルネットワークは、層構造を有してもよく、ニューラルネットワークの、少なくともメディアデータ側層および/または出力層の重みは、デコーダによって、パラメータの前記少なくとも1つの集合で（たとえば、それを使用して

50

)更新されてもよい。

【0031】

いくつかの実施形態では、受領されたメディアビットストリームは、メディアデータ側層および/または出力層を更新するためのパラメータのみを含む。すなわち、ニューラルネットワークの他の層を更新するためのパラメータはメディアビットストリームに含まれない。デコードのステップは、受領されたメディアビットストリームをデコードして、デコードされたメディアデータと、メディアデータ側層および/または出力層の重みを更新するためのパラメータとを得ることを含むことができる。ニューラルネットワークを更新するステップは、受領されたパラメータを用いてニューラルネットワークのメディアデータ側層および/または出力層の重みを更新することを含んでいてもよい。たとえば、メディアデータ側層および/または出力層のみが更新され、他の層は更新されない。

10

【0032】

本開示の第3の側面によれば、デコーダにおいて実装されるニューラルネットワークを更新するためのパラメータを伝送するためのメディアビットストリームを生成するための装置(たとえば、そのようなメディアビットストリームを伝送するための装置)が提供される。本装置は、(a)ニューラルネットワークを更新するためのパラメータの少なくとも1つの集合を決定するステップを含む方法を実行するように構成されたプロセッサを含んでいてもよい。本方法はさらに、(b)パラメータの前記少なくとも1つの集合およびメディアデータをエンコードして、メディアビットストリームを生成するステップを含んでいてもよい。本方法は、さらに、(c)パラメータの前記少なくとも1つの集合で(たとえばそれを使用して)ニューラルネットワークを更新するために、メディアビットストリームをデコーダに伝送するステップを含んでいてもよい。

20

【0033】

本開示の第4の側面によれば、デコーダにおいて実装されるニューラルネットワークを更新するための装置が提供される。本装置は、(a)メディアデータと、ニューラルネットワークを更新するためのパラメータの少なくとも1つの集合とを含む符号化メディアビットストリームを受領するための受領器を含んでいてもよい。本装置はさらに、(b)受領されたメディアビットストリームをデコードして、デコードされたメディアデータと、ニューラルネットワークを更新するためのパラメータの前記少なくとも1つの集合とを得るためのデコーダを含んでいてもよい。本装置は、さらに、(c)パラメータの前記少なくとも1つの集合で(たとえば、それを使用して)ニューラルネットワークを更新するための更新器を含んでいてもよい。

30

【0034】

いくつかの実施形態では、受領器は、さらに、構文要素の集合を受領してもよく、更新器は、構文要素の集合に基づいて、パラメータの前記少なくとも1つの集合で(たとえば、それを使用して)ニューラルネットワークを更新してもよい。

【0035】

いくつかの実施形態では、ニューラルネットワークを更新するためのパラメータの2つ以上の集合が、受領器によって受領される符号化メディアビットストリームに含まれてもよく、受領された構文要素の集合は、実行されるニューラルネットワークのそれぞれの更新のためのパラメータのそれぞれの集合を識別する一つまたは複数の構文要素を含んでいてもよい。

40

【0036】

本開示の第5の側面によれば、処理能力を有する装置によって実行されたときに、処理能力を有する装置(たとえば、プロセッサ)に、デコーダにおいて実装されたニューラルネットワークを更新するためのパラメータを伝送するためにメディアビットストリームを生成する方法を実行させるように適応された命令を有するコンピュータ読み取り可能な記憶媒体を有するコンピュータ・プログラム・プロダクトが提供される。

【0037】

本開示の第6の側面によれば、処理能力を有する装置によって実行されたときに、処理

50

能力を有する装置（たとえば、プロセッサ）に、デコーダにおいて実装されたニューラルネットワークを更新するための方法を実行させるように適応された命令を有するコンピュータ読み取り可能な記憶媒体を有するコンピュータ・プログラム・プロダクトが提供される。

【0038】

方法ステップおよび装置特徴は、多くの仕方で交換されることが理解されるであろう。特に、当業者が理解するであろうように、開示された方法の詳細は、方法のステップの一部または全部を実行するように適応された装置として実装されることができ、その逆も可能である。特に、本開示による方法は、それぞれの装置を動作させる方法に関することが理解される。さらに、方法に関してなされるそれぞれの陳述は、対応する装置にも同様に適用されることが理解される。さらに、エンコーダ/送信器側に関してなされるそれぞれの陳述は、デコーダ/受信器側にも同様に適用されることが理解される。

10

【図面の簡単な説明】

【0039】

ここで、添付の図面を参照して、単に例として、本開示の例示的な実施形態を記述する。

【図1】デコーダにおいて実装されたニューラルネットワークを更新するためのパラメータを伝送するためのメディアビットストリームを生成する方法の一例の流れ図を示す。

【図2】デコーダにおいて実装されたニューラルネットワークを更新する方法の一例の流れ図を示す。

【図3】エンコーダ段とデコーダ段とを含む多層生成器として実装されるニューラルネットワークの一例を概略的に示す。

20

【発明を実施するための形態】

【0040】

概観

本開示の文脈において、深層ニューラルネットワークは、オーディオおよび/またはビデオメディアデータを処理するために使用されうる。オーディオについては、ニューラルネットワークの可能な応用は、メディア解析、メディア向上（たとえば、符号化オーディオ向上）、メディア分類（発話、映画、音楽、拍手など）、メディア生成（たとえば、欠けているメディアデータを、または完全に新しいメディアデータを生成する深層生成モデル（deep generative model））、またはダイアログ向上を含む。ビデオについては、ニューラルネットワークの可能な応用は、ブロック化解除または動き向上（たとえば、スポーツ番組の場合）を含む。

30

【0041】

ニューラルネットワークは、可能性としては好適な部分（たとえばフレーム）に分割されたメディアコンテンツを入力として受領し、処理された（たとえば、向上された）メディアコンテンツ（オーディオおよび/またはビデオ）、メディアコンテンツの解析結果、メディアコンテンツの分類、または新しく生成されたメディアコンテンツのうちの一つまたは複数を出力しうる。

【0042】

すでに見いだされたように、デコーダ/クライアントにおける既存のニューラルネットワークの重みの小さな変化が、処理されたメディアコンテンツ、たとえば発話コンテンツの品質の著しい改善を可能にすることが多い。敵対的生成ネットワーク（GAN）をニューラルネットワークの例にとると、ビットレートXの符号化オーディオを向上させるために深層ニューラルネットワーク生成器（G）を使用する場合、それを（異なる）ビットレートYで機能させるためには、（ビットレートXのための事前トレーニングされたモデルから始めて）ほんのわずかのトレーニング時間しか必要とされないことが観察されている。すなわち、事前トレーニングされたモデルはすでにビットレートYで機能し、最適な性能を達成するためには、増分的な（incremental）トレーニングだけで十分であると思われる。同様の観察は、異なるタイプのメディアコンテンツについても当てはまる（たとえば、発話で事前トレーニングされて、拍手のために適応される）。

40

50

【 0 0 4 3 】

増分的な更新を決定するために、GANのトレーニングは転移学習に基づいてもよい。転移学習は、特定のタスクを行うようにトレーニングされたニューラルネットワークが異なるタスクに適応される方法である。

【 0 0 4 4 】

本開示の文脈では、デコーダで実装されるニューラルネットワークは、対応するメディアデータと一緒にエンコードされた（標準化された）メディアビットストリームにおいて伝送される更新パラメータに基づいて更新される。更新パラメータは、対応するメディアデータと一緒にメディアビットストリーム中にパッケージングされてもよい。メディアビットストリームはデコーダにおいてデコードされ、更新パラメータおよび対応するメディアデータが抽出され、対応するメディアデータを処理する前に、更新パラメータを使用してニューラルネットワークが更新され、更新されたニューラルネットワークが対応するメディアデータを処理する。

10

【 0 0 4 5 】

言い換えると、受領装置（たとえば、デコーダ）内の既存のニューラルネットワークに適用されるべき更新（更新パラメータ）を、本開示は、少なくとも部分的には、（標準化された）メディアビットストリームを通じて、ニューラルネットワークによって処理されるべき対応するメディアデータと一緒に、伝達することに関する。好ましくは、これらの更新は、更新されたニューラルネットワークがその向上を適用しようとしているメディアデータと一緒にパッケージングされる。さらに好ましくは、これは、メディアデータと一緒に伝送される時間整列された更新（更新パラメータ）に関わり、それはまた、メディアを向上させるニューラルネットワークを、現在のメディアコンテンツに基づいて微調整することができることを意味する。

20

【 0 0 4 6 】

このように、おおまかには、本開示は、ニューラルネットワーク更新を、メディアビットストリームにおいて、エンコードされたメディアと一緒に、デコード・システムに転送することに関する。これは、ビデオ、オーディオ、およびその他のメディア関連コンテンツ（たとえば、触覚コンテンツなど）に適用される。パラメータ更新が種々のタイプのビットストリーム内のどこで伝送されるかに関する非限定的な例のリストが、のちにさらに提供される。

30

【 0 0 4 7 】

また、以下により詳細に記載されるように、本開示に記載される方法および装置、およびニューラルネットワークを更新するためのパラメータのための基礎となる転送機構は、ETSI AC-4、ETSI AC-3/EAC-3、ISO/IEC23008-3（MPEG-H Audio）、ISO/IEC14496-3（MPEG-4 Audio）およびすべての関連規格（たとえば、3GPPなど）、ISO/IEC23003-3（USAC）、およびISO/IEC23090-4（MPEG-I Audio）にとって有意でありうる。

【 0 0 4 8 】

まず、図1を参照する。図1では、デコーダに実装されたニューラルネットワークを更新するためのパラメータを伝送するためのメディアビットストリームを生成する方法が示されている。ステップS101では、デコーダにおいて実装されるニューラルネットワークを更新するために、パラメータ（更新パラメータ）の少なくとも1つの集合が決定される。デコーダは、ニューラルネットワークを更新するためのパラメータの前記少なくとも1つの集合のための受領装置として理解されてもよい。パラメータの前記少なくとも1つの集合をデコーダに送信するために、ステップ102において、パラメータの前記少なくとも1つの集合およびメディアデータがエンコードされて、メディアビットストリームを生成する。すなわち、パラメータの前記少なくとも1つの集合はメディアビットストリームにエンコードされる。好適に更新されたニューラルネットワーク（すなわち、パラメータの前記少なくとも1つの集合を用いて更新されている）によって処理されるべき対応するメディアデータは、パラメータの前記少なくとも1つの集合と一緒にメディアビットストリーム

40

50

にエンコードされる。パケット化されたメディアビットストリームについては、メディアデータとパラメータの前記少なくとも1つの集合とが一緒にパッケージングされてもよい。一般に、パラメータの前記少なくとも1つの集合は、好ましくは、メディアビットストリームにおいて、対応するメディアデータと時間整列される。次いで、ステップ103において、生成されたメディアビットストリームが、ニューラルネットワークを更新するために、デコーダに提供される（たとえば、出力される、または伝送される）。パラメータの前記少なくとも1つの集合およびメディアデータを含むメディアビットストリームを生成することによって、ニューラルネットワークを更新するためのパラメータの前記少なくとも1つの集合が、メディアデータと一緒にデコーダに伝達されることができる。いくつかの実施形態では、メディアデータは、オーディオデータおよび/またはビデオデータの一つまたは複数を含んでいてもよい。

10

【0049】

次に、図2を参照する。図2では、デコーダに実装されたニューラルネットワークを更新する方法が示されている。ステップS201では、メディアデータと、ニューラルネットワークを更新するためのパラメータの少なくとも1つの集合とを含む符号化メディアビットストリームが受領される。ステップS202では、符号化メディアビットストリームがデコードされ、デコードされたメディアデータおよびパラメータの前記少なくとも1つの集合が得られる。次いで、ステップS203において、デコーダによって、パラメータの前記少なくとも1つの集合が、ニューラルネットワークを更新するために使用される。

【0050】

20

いくつかの実施形態では、ニューラルネットワークはメディアデータの処理のために使用されてもよく、受領された符号化メディアビットストリームにおいて、パラメータの前記少なくとも1つの集合が、ニューラルネットワークによって処理されるメディアデータと時間整列されてもよい。図2の例におけるステップS204に示されるように、パラメータの前記少なくとも1つの集合がメディアデータと時間整列されている場合、ニューラルネットワークは、デコードされたメディアデータをその後処理するために、パラメータの前記少なくとも1つの集合を用いてデコーダによって更新されてもよい。いくつかの実施形態では、メディアデータは、オーディオデータおよび/またはビデオデータの一つまたは複数を含んでいてもよい。

【0051】

30

上記の諸方法は、それぞれの装置において実装されてもよい。上記の諸方法は、それぞれのコンピュータ・プログラム・プロダクトによってさらに実装されてもよい。

【0052】

メディアビットストリームにおける更新パラメータの伝送

いくつかの実施形態において、パラメータの前記少なくとも1つの集合は、構文要素の集合に基づいてエンコードされてもよい。構文要素は、生成されたメディアビットストリーム内のパラメータの前記少なくとも1つの集合を伝送（転送）することを許容でき、さらに、デコーダがニューラルネットワークを更新することを許容できる。いくつかの実施形態では、構文要素の集合は、デコーダによって受領され、デコーダによってパラメータの前記少なくとも1つの集合を用いてニューラルネットワークを更新することは、受領された構文要素の集合に基づいていてもよい。

40

【0053】

たとえば、以下の構文要素が、メディアビットストリームにおいてパラメータの前記少なくとも1つの集合を伝送するために使用されてもよい。

【0054】

表1：ニューラルネットワークを更新するためのパラメータの少なくとも1つの集合をメディアビットストリームにおいて伝送するための構文要素の例

50

【表 1】

構文	ビット数
neural_network_update() {	
while (neural_network_id != 0) {	5
if (neural_network_id == 31) {	
neural_network_id += variable_bits(5);	
}	
nn_update_config();	
nn_update_size = variable_bits(8);	
for (i = 0; i < nn_update_size; i++) {	
nn_update_byte[i];	8
}	
}	
byte_align;	0...7
}	
nn_update_config() {	
upd_type	1
if (update_id_present) {	1
update_id	14
}	8
number_of_upd_packages	
if (number_of_upd_packages > 0){	8
upd_sequence_counter	
} else {	
upd_sequence_counter = 0	
}	
}	

【0055】

表1の例を参照すると、そこに記載された構文要素は、ニューラルネットワーク更新のために使用されるパラメータの前記少なくとも1つの集合のための必要とされるビットをメディアビットストリーム中にパッケージングするために使用されうる。しかしながら、これらの構文要素は、パラメータの前記少なくとも1つの集合を伝送するために使用されるそれぞれのメディアビットストリームのコーデックに依存して修正されてもよい。たとえば、ISO/IEC 23003-3、USACで規定されているようなescaped_value()構文が、ETSI 103 190で規定されているようなvariable_bits()構文の代わりに使用されてもよい。

【0056】

再び表1の例を参照すると、これらの構文要素で使用される意味内容は次のとおりでありうる。

neural_network_idは、更新されなければならないデコーダ内のニューラルネットワークを識別するために使用されてもよい。

nn_update_byteは、特定のフォーマットで、ニューラルネットワークを更新するためのパラメータの前記少なくとも1つの集合を担持するために使用されてもよい。

upd_typeは、パラメータの前記少なくとも1つの集合を用いて実行される更新のタイプを識別するために使用されてもよく、これは、たとえば、以下の可能性がある：

0 相対更新（伝送された値が、勾配を含み、もとの重みから加算または減算される）

1 絶対更新（伝送された値がもとの重みを置き換えてもよい）

注意すべきことに、相対更新および絶対更新に対する0および1の割り当ては、限定しない例であり、実装によっては、割り当ては逆でもよい。現在のコンテキストで重要なのは、upd_typeが何らかの仕方で相対的な更新と絶対的な更新を示しうることである。

update_idは、実行される更新のためのパラメータの集合を識別するために、すなわち、

更新パラメータの異なる集合の間の区別をするために使用されてもよい。これにより、更新パラメータの異なる集合からのパラメータの混同を避けることができる。一般に、これらの構文要素は、実行されるべきニューラルネットワークのそれぞれの更新のためのパラメータのそれぞれの集合を識別する一つまたは複数の構文要素を含んでいてもよい。

number_of_upd_packagesは、関連するneural_network_idについての更新パッケージの総数を信号伝達する。パラメータの前記少なくとも1つの集合を用いた更新は、デコーダ（クライアント）がその更新のための関連するすべてのパッケージを受領した場合にのみ適用されうる。

upd_sequence_counterは、関連する更新のための特定のパッケージを識別するために使用されてもよい。

10

nn_update_byteは、パラメータ（たとえば、重みの更新）の前記少なくとも1つの集合を担持するために使用されてもよい。これらのバイトのフォーマットはneural_network_idと、upd_typeのような他の値とに依存してもよい。

【0057】

更新されるべきニューラルネットワークの構造は限定されないが、いくつかの実施形態では、ニューラルネットワークは層構造（多層構造）を有してもよく、パラメータの前記少なくとも1つの集合は、少なくともニューラルネットワークのメディアデータ側層および/または出力層の重みを更新するためのパラメータを含んでいてもよい。いくつかの実施形態では、よって、少なくともニューラルネットワークのメディアデータ側層および/または出力層の重みは、パラメータの前記少なくとも1つの集合を用いてデコーダによって更新されてもよい。いくつかの実施形態では、ニューラルネットワークのメディアデータ側層および/または出力層のみが更新されてもよい。たとえば、弁別モデル（すなわち、入力メディアコンテンツを分類するニューラルネットワーク）の場合、出力層のみが更新されてもよい。

20

【0058】

たとえば、表2に示される以下の構文要素が、パラメータの前記少なくとも1つの集合を用いてニューラルネットワークを更新するために使用されてもよい。

【0059】

表2：パラメータの前記少なくとも1つの集合を用いてニューラルネットワークを更新するための構文要素の例

30

【表2】

構文	ビット数
nn_update_example()	
{	
number_of_upd_weights = variable_bits(8);	
for (i=0; i<number_of_upd_weights; i++) {	
layer_id	5
weight_num_in_layer	11
weight_value	8
}	
byte_align;	0...7
}	

40

【0060】

表2の例を参照すると、これらの構文要素で使用される意味内容は次のとおりでありうる。

layer_idは、重みが更新されなければならない層を識別するために使用されてもよい。

weight_num_in_layerは、層内のどの重みが更新される必要があるかを識別するために使用されてもよい。

weight_valueは、更新されなければならない重みの値を担持するために使用されてもよい。それは、upd_typeの値、すなわち、相対更新か絶対更新かに依存して解釈されても

50

よい。

upd_type = 0であれば、符号付バイト

upd_type = 1であれば、符号なしバイト

ここでもまた、0および1の割り当ては、限定しない例であると理解される。

【0061】

いくつかの実施形態では、デコーダにおいて実装されるニューラルネットワークは、メディアデータの処理（たとえば、向上、解析、分類など）に使用されてもよい。その際、メディアビットストリームにおいて、ニューラルネットワークを更新するためのパラメータ（更新パラメータ）の前記少なくとも1つの集合は、パラメータの前記少なくとも1つの集合を用いて更新されたときにニューラルネットワークによって処理されるメディアデータと時間整列されてもよい。言い換えると、処理されるべきメディアデータがデコーダに到着する場合、デコーダで実装されたニューラルネットワークは、メディアデータのその特定の部分のために意図されたパラメータによって更新されなければならない。よって、それぞれの更新パラメータは、それについて更新パラメータが必要とされる処理のための対応するメディアデータと少なくとも同時に、可能性としてはそれよりも前に、到着しなければならない。

10

【0062】

この文脈では、時間整列とは、パラメータの前記少なくとも1つの集合およびメディアデータをメディアビットストリームにエンコードする順序を指しうる。たとえば、パラメータの前記少なくとも1つの集合は、それぞれの更新されたニューラルネットワークによって処理されるべきメディアデータに先立って、メディアビットストリームにおいてエンコードされてもよい。メディアデータが拍手を表すオーディオデータである場合、拍手を表すメディアデータを処理するためのニューラルネットワークを更新するために決定されるパラメータの前記少なくとも1つの集合は、前記メディアデータに先立って、ビットストリーム内でエンコードされる。代替的または追加的に、パラメータの前記少なくとも1つの集合は、パラメータの前記少なくとも1つの集合を用いて更新されるニューラルネットワークによって処理されるべきそれぞれのメディアデータが、パラメータのその少なくとも1つの集合に関して遅延されるように、メディアビットストリームにおいてエンコードされてもよい。それにより、ニューラルネットワークによるメディアデータの処理時に更新が完了しないことによって生じるいかなる遅延も回避することができる。必要なら、更新パラメータは、デコーダでバッファリングされ、それぞれのメディアデータがデコーダに到着したときにニューラルネットワークを更新するために使用されてもよい。

20

30

【0063】

たとえば、フレームベースのフォーマットについては、パラメータの前記少なくとも1つの集合は、ニューラルネットワークによって処理されるべきメディアデータと同じフレーム内でエンコードされ、パラメータのこの少なくとも1つの集合を用いて更新されてもよい。あるいはまた、パラメータの前記少なくとも1つの集合は、より早いフレームにおいてエンコードされてもよく、パラメータの前記少なくとも1つの集合を含むフレームと、パラメータの前記少なくとも1つの集合を用いて更新されたニューラルネットワークによって処理されるべきメディアデータを含むフレームとの間には、フレーム単位で固定されたオフセットがある。

40

【0064】

いくつかの実施形態では、パラメータの前記少なくとも1つの集合は、コーデック・モード、メディアデータの内容、およびエンコード制約条件のうちの一つまたは複数に基づいて決定されてもよい。

【0065】

いくつかの実施形態では、コーデック・モードは、ビットレート、ビデオおよび/またはオーディオ・フレームレート、および使用されるコア・コーデックのうちの一つまたは複数を含んでいてもよい。パラメータの前記少なくとも1つの集合がたとえばビットレートに基づいて決定される場合、これは、メディアビットストリーム内のパラメータの前記

50

少なくとも1つの集合を、そのビットレートを有するそれぞれのメディアデータを参照して、時間整列させるために利用されてもよい。これは、ビデオおよび/またはオーディオ・フレームレートおよび使用されるコア・コーデックにも当てはまりうる。

【0066】

いくつかの実施形態では、メディアデータの内容は、オーディオデータのための発話、音楽、および拍手のうちの一つまたは複数であってもよい。同様に、メディアデータの内容はビデオデータであってもよい。たとえば、オーディオおよび/またはビデオデータのそれぞれの内容に基づいて、パラメータの前記少なくとも1つの集合が決定される場合、これは、メディアビットストリーム内のパラメータの前記少なくとも1つの集合を、その内容を有するそれぞれのメディアデータを参照して、時間整列させるために利用されてもよい。

10

【0067】

いくつかの実施形態では、エンコード制約条件は、スケーラビリティのパフォーマンスについての制約条件および適応処理についての制約条件のうちの一つまたは複数を含んでもよい。パラメータの前記少なくとも1つの集合が、たとえば、それぞれのエンコード制約条件に基づいて決定される場合、これも、メディアビットストリーム内のパラメータの前記少なくとも1つの集合を、エンコード制約条件が当てはまるそれぞれのメディアデータを参照して、時間整列させるために利用されてもよい。

【0068】

メディアデータのフォーマットは限定されないが、いくつかの実施形態では、メディアデータはMPEG-Hオーディオ(ISO/IEC23008-3)またはMPEG-Iオーディオ・フォーマットであってもよく、メディアビットストリームはMHASフォーマットでのパケット化されたビットストリームであってもよい。いくつかの実施形態では、その際、パラメータの前記少なくとも1つの集合は、パラメータの前記少なくとも1つの集合を、新しいMHASパケットタイプの一つまたは複数のMHASパケットにカプセル化することによってエンコードされてもよい。たとえば、必要とされるペイロードを伝送するための1つの可能な方法は、新しいMHASパケットタイプPACTYP_NN_UPDATEを割り当てることでありうる。

20

【0069】

デコーダは、MHASパケットタイプを解釈することによって、それぞれの(あらかじめ定義された)ニューラルネットワークを更新するためのペイロードを識別してもよい。この目的のために、MPEG-Hオーディオ規格では、構文要素は次の表3に示されるように修正されてもよい。

30

【0070】

表3: メディアビットストリーム中のパラメータの前記少なくとも一つの集合を伝送するためのMPEG-Hオーディオ規格における修正構文要素の例

【表3】

構文	ビット数	ニーマニック
MHASPacketPayload(MHASPacketType)		
{		
switch (MHASPacketType) {		
...		
case PACTYP_NN_UPDATE:		
for (i=0; i<MHASPacketLength; i++) {		
neural_network_update_byte[i];	8	bslbf
}		
...		
}		
...		
}		

40

【0071】

表3の例を参照すると、neural_network_update_byteは、以前に説明した構文要素n

50

neural_network_update()のバイトを担持してもよい。さらに、更新されるべきニューラルネットワーク(の性質)に依存して、ビットストリーム拡張機構usacExtElementが使用されてもよい。

【0072】

いくつかの実施形態では、メディアデータは、AC-4、AC-3、EAC-3、MPEG-4またはMPEG-D USACフォーマットであってもよい。さらに、いくつかの実施形態では、その際、パラメータの前記少なくとも1つの集合は、メディアビットストリームにおいて、一つまたは複数のペイロード要素として、または一つまたは複数のデータストリーム要素(DSE)としてエンコードされてもよい。

【0073】

AC-4(ETSI 103 190)の場合、メディアビットストリームにおいてパラメータの前記少なくとも1つの集合を伝送することの例は、neural_network_update()要素(表1参照)を、emdf_payloads_substream()要素を介して伝送(または転送)することであってもよい。これを可能にするために、このEMDFペイロードの識別を可能にするために、フリーなemdf_payload_idが使用されてもよい。その際、neural_network_update()要素に含まれるすべてのバイトは、emdf_payload_byteとして伝送(転送)される。

【0074】

AC-3およびEAC-3(ETSI 102 366)の場合、メディアビットストリーム中のパラメータの前記少なくとも1つの集合を伝送することの例は、ETSI 102 366、Annex Hに記述されているような、neural_network_update()要素をemdf_container()要素を介して伝送(または転送)することでありうる。これを可能にするために、EMDFペイロードの識別を可能にするためにフリーなemdf_payload_idが使用されてもよい。このIDは、EMDFフレームワークの原則を満たすために、AC-4の場合と同じであってもよい。neural_network_update()要素に含まれるすべてのバイトは、emdf_payload_byteとして伝送(転送)されてもよい。

【0075】

MPEG-4(AAC、HE-AACv1、HE-AACv2)の場合、MPEG-4によって定義される少なくとも2つの独立した拡張機構が、ニューラルネットワークを更新するためのパラメータの前記少なくとも1つの集合をメディアビットストリームにおいて伝送するために使用されてもよい。第1の拡張機構を使用する場合、パラメータの前記少なくとも1つの集合は、以下の例に示されるようなデータストリーム要素(DSE)内のデータストリームとして伝送されてもよい。

【表4】

```

data_stream_element() {
    element_instance_tag; 4 uimsbf
    data_byte_align_flag; 1 uimsbf
    cnt = count; 8 uimsbf
    if (cnt == 255)
        cnt += esc_count; 8 uimsbf
    if (data_byte_align_flag)
        byte_alignment();
    for (i = 0; i < cnt; i++)
        data_stream_byte[element_instance_tag][i]; 8 uimsbf
}

```

element_instance_tag	データストリーム要素を識別するための番号。複数のDSEが存在する場合は扱いに注意が必要。
cnt	neural_network_update()のバイト単位でのサイズ
data_stream_byte	neural_network_update()ペイロード、上記参照

【0076】

第2の拡張機構を使用する場合、MPEG-4が(特に非耐性ペイロード(non-resilience payloads)について)

拡張機構を定義していることが活用されてもよい。以下の例に示されるように、充填要素 (Fill Element) ID_FIL内に含まれる最上位の構文 (シンタックス) 要素extension_payloadがある。

【表 5】

```

fill_element() {
    cnt = count; 4 uimsbf
    if (cnt == 15)
        cnt += esc_count - 1; 8 uimsbf
    while (cnt > 0) {
        cnt -= extension_payload(cnt);
    }
}

```

10

【 0 0 7 7 】

代替的または追加的に、MPEG-4のextension_type [拡張タイプ] EXT_DATA_ELEMENTが使用されてもよい。この場合、neural_network_update()構文要素のバイトはextension_payload()として転送されてもよい。

【 0 0 7 8 】

MPEG-D USACの場合、ニューラルネットワークを更新するためのパラメータの前記少なくとも1つの集合は、23003-3に定義されるUSAC拡張要素に基づいて伝送されてもよい。新しいusacExtElementTypeがID_EXT_ELE_NN_UPDATEと名前を付けられてもよく、23003-3のテーブル17「Syntax of UsacExtElementConfig()」における対応する構文要素が次の例のように修正されてもよい。

20

【表 6】

構文	ビット数	ニーモニック
UsacExtElementConfig()		
{		
...		
case ID_EXT_ELE_NN_UPDATE:		
neural_network_update()		
...		
}		
...		
}		

30

【 0 0 7 9 】

ニューラルネットワークの構造

ニューラルネットワークの構造は限定されないが、ニューラルネットワークは層構造 (多層構造) を有してもよい。そのような層構造は、以下の限定しない例に基づいて説明される。

【 0 0 8 0 】

図3の例を参照すると、ニューラルネットワークは、たとえば、オーディオデータを処理するために敵対的生成ネットワークの場面でトレーニングされた、いわゆる生成器であってもよい。そのような生成器の構造は一般に限定されないが、生成器はエンコーダ段およびデコーダ段を含んでいてもよい。生成器のエンコーダ段およびデコーダ段は、全層畳み込み式であってもよい。デコーダ段は、エンコーダ段をミラーしてもよく、エンコーダ段もデコーダ段も、それぞれいくつかのL層を含み、各層LはN個のフィルタをもつ。L層は自然数 > 1 であってもよく、Nは自然数 > 1 であってもよい。N個のフィルタのサイズ (カーネルサイズとしても知られる) は限定されず、用途の要件に従って選択することができる。しかしながら、フィルタサイズは、L層のそれぞれにおいて同じであってもよい。

40

【 0 0 8 1 】

50

図3の例では、処理される生のオーディオデータが、第1ステップ1において生成器に入力されてもよい。よって、第1の図示されたエンコーダ層3は、メディアデータ側層 (media data facing layer) であってもよい。エンコーダ段のL層のそれぞれにおいて、フィルタの数は増加してもよい。各フィルタは、ストライド2で、各エンコーダ層に入力されたオーディオデータに対して作用してもよい。この場合、幅 (時間的な信号の持続時間) が狭くなるほど深さは大きくなる。よって、因子2による学習可能なダウンサンプリングが実行されてもよい。あるいはまた、フィルタは、各エンコーダ層において、ストライド1で動作してもよく、その後、(既知の信号処理におけるように) 因子2によるダウンサンプリングが続く。

【0082】

少なくとも1つのエンコーダ層および少なくとも1つのデコーダ層において、さらに、非線形演算が、活性化として実行されてもよい。非線形演算は、パラメトリック整流線形ユニット (parametric rectified linear unit、PReLU)、整流線形ユニット (rectified linear unit、ReLU)、漏洩整流線形ユニット (leaky rectified linear unit、LReLU)、指数線形ユニット (exponential linear unit、eLU) およびスケールされた指数線形ユニット (scaled exponential linear unit、SeLU) のうちの一つまたは複数を含みうる。しかしながら、生成器の用途によっては、他の活性化も考えられる。

【0083】

図3の例に概略的に示されるように、それぞれのデコーダ層7、8、および9は、エンコーダ層5、4、および3をミラーしてもよい。各層におけるフィルタの数および各層におけるフィルタ幅は、デコーダ段において、エンコーダ段におけるのと同じであってもよいが、狭い幅 (時間的な信号の持続時間) から始まるオーディオ信号のアップサンプリングは、2つの代替アプローチによって実行されうる。端数ストライド畳み込み (fractionally-strided convolution) (転置畳み込み (transposed convolution) としても知られる) が、デコーダ段の諸層において、オーディオ信号の幅を全持続時間、すなわち、生成器に入力されたオーディオ信号のフレームまで増加させるために、使用されてもよい。

【0084】

あるいはまた、デコーダ段の各層において、アップサンプリングおよび補間が、アップサンプリング因子2で従来の信号処理におけるように実行された後に、フィルタは、ストライド1で各層に入力されたオーディオデータに対して作用してもよい。

【0085】

処理されたオーディオデータが最終ステップ11において出力される前に、出力層 (畳み込み層) 10がデコーダ段に続いてもよい。出力層においては、活性化は、エンコーダ層のうちの前記少なくとも1つおよびデコーダ層のうちの前記少なくとも1つにおいて実行される活性化とは異なる場合がある。活性化は、生成器に入力されるオーディオ信号と同じ範囲に制限される任意の非線形関数であってもよい。たとえば、向上されるべき時間信号は ± 1 の間に制限されてもよい。その際、活性化は、たとえば \tanh 演算に基づいていてもよい。

【0086】

エンコーダ段とデコーダ段との間で、オーディオデータが修正されてもよい。修正は、符号化されたオーディオ特徴空間 (ボトルネック層としても知られる) 12に基づいてもよい。符号化されたオーディオ特徴空間における修正は、たとえば、ランダムノイズベクトル (z) を、エンコーダ段の最終層から出力される生のオーディオデータのベクトル表現 (c) と連結することによって行うことができる。ランダムノイズベクトルは、 $z = 0$ に設定されてもよい。

【0087】

符号化されたオーディオの時間構造またはテクスチャを維持するため、エンコーダ段とデコーダ段の相同層の間にスキップ接続2が存在してもよい。それにより、上述の符号化されたオーディオ特徴空間がバイパスされ、情報の損失を防止しうるのである。スキップ接続は、連結および信号加算の一つまたは複数をを使用して実装されてもよい。スキップ接

10

20

30

40

50

続の実装のため、フィルタ出力の数は「実質的に」倍になることがありうる。

【 0 0 8 8 】

図3の例を参照すると、ニューラルネットワークについての限定しない例としての生成器の構造は、以下のものであってもよい：

1/入力：生のオーディオデータ

3/エンコーダ層L = 1：フィルタ数N = 16、フィルタサイズ = 31、活性化 = PreLU

4/エンコーダ層L = 2：フィルタ数N = 32、フィルタサイズ = 31、活性化 = PreLU

。

。

5/エンコーダ層L = 11：フィルタ数N = 512、フィルタサイズ = 31

6/エンコーダ層L = 12：フィルタ数N = 1024、フィルタサイズ = 31

12/符号化されたオーディオ特徴空間

7/デコーダ層L = 1：フィルタ数N = 512、フィルタサイズ = 31

。

。

8/デコーダ層L = 10：フィルタ数N = 32、フィルタサイズ = 31、活性化PreLU

9/デコーダ層L = 11：フィルタ数N = 16、フィルタサイズ = 31、活性化PreLU

10/出力層：フィルタ数N = 1、フィルタサイズ = 31、活性化tanh

11/出力：向上されたオーディオデータ

2/スキップ接続

。

【 0 0 8 9 】

注意しておくが、上述の構造は、単に限定しない例を表すものである。そのような生成器の用途に依存して、エンコーダ段およびデコーダ段における層の数は、それぞれスケールダウンまたはスケールアップされてもよい。さらに、本開示の文脈におけるニューラルネットワークは、GAN関連モデルに限定されないことが理解される。たとえば、上記の例は、生成ネットワークモデルを参照するが、ニューラルネットワークは、弁別ネットワークモデルに基づくニューラルネットワークであってもよい。一般に、本開示は、メディアコンテンツ（たとえば、オーディオおよび/またはビデオコンテンツ）を処理するのに好適な任意のタイプのニューラルネットワークに適用可能である。

【 0 0 9 0 】

図3に示された生成器の例に戻ると、生成器のトレーニングは、弁別器との真偽対話に基づく敵対的生成ネットワークの場面において促進されうる。そのようなニューラルネットワークを更新するためのパラメータの少なくとも1つの集合を決定するために、本発明者らは、パラメータの前記少なくとも1つの集合が、事前トレーニングされた生成器およびトレーニングされていない弁別器を含む敵対的生成ネットワークの場面に基づいて決定されうることを見出した。この場面では、生成器は、非常に限定された（異なる構成の）トレーニングデータおよび時間を用いて段階的に更新され、新しい（更新された）構成についてのパフォーマンスを改善しうる。このアプローチでは、生成器内のすべての重みが段階的に更新されうる。しかしながら、新しいドメインに対して作用するニューラルネットワークのための最も重要な層は、信号に面する（たとえば、メディアデータに面する）層でありうる。よって、更新パラメータを少数の重みのみに限定する（よって、結果としてビットレート・オーバーヘッドを低減する）ために、事前トレーニングされた生成器のすべての重みが凍結され、生成器のメディアデータ（信号）側層および/または出力層のみが更新されてもよい。

【 0 0 9 1 】

解釈

特に断りのない限り、以下の議論から明らかなように、本開示を通じて、「処理」、「

10

20

30

40

50

コンピューティング」、「計算」、「決定」、「解析」などの用語を使用する議論は、コンピュータまたは計算システム、または同様の電子計算装置のアクションおよび/またはプロセスであって、電子的な量のような物理的な量として表わされるデータを操作および/または変換して、物理的な量として同様に表わされる他のデータにするものを指すことが理解される。

【0092】

同様に、用語「プロセッサ」は、たとえばレジスタおよび/またはメモリからの電子データを、該電子データを、たとえばレジスタおよび/またはメモリに記憶されうる他の電子データに変換するために処理する任意の装置または装置の一部を指してもよい。

【0093】

本明細書に記載される方法論は、ある例示的实施形態では、命令の集合を含むコンピュータ読み取り可能な（機械読み取り可能な、ともいう）コードを受け入れる一つまたは複数のプロセッサによって実行可能である。該命令は、前記プロセッサの一つまたは複数によって実行されると、本明細書に記載の方法の少なくとも一つを実行する。実行されるべきアクションを指定する命令（逐次的またはそれ以外）の集合を実行することができる任意のプロセッサが含まれる。よって、一例は一つまたは複数のプロセッサを含む典型的な処理システムである。各プロセッサは、CPU、グラフィック処理ユニット、プログラマブルDSP装置の一つまたは複数を含んでいてもよい。処理システムはさらに、メインRAMおよび/またはスタティックRAMおよび/またはROMを含むメモリ・サブシステムを含んでいてもよい。構成要素間で通信するためのバスサブシステムが含まれてもよい。処理システムはさらに、たとえばネットワークによって結合された諸プロセッサをもつ分散式の処理システムであってもよい。処理システムがディスプレイを必要とする場合、そのようなディスプレイが含まれてもよい。たとえば、液晶ディスプレイ（LCD）または陰極線管（CRT）ディスプレイである。主導のデータ入力が必要とされる場合、処理システムは、キーボードのような英数字入力ユニット、マウスのようなポインティング制御装置などの一つまたは複数として入力装置を含んでいてもよい。処理システムは、ディスクドライブユニットのような記憶システムをも包含してもよい。いくつかの構成の処理システムは、音出力装置およびネットワーク・インターフェース装置を含んでいてもよい。よって、メモリ・サブシステムは、コンピュータ読み取り可能な担体媒体を含み、該媒体は、一つまたは複数のプロセッサによって実行されると、本明細書に記載された方法の一つまたは複数を実行させる一組の命令を含んでいる。方法がいくつかの要素、たとえば、いくつかのステップを含む場合、特に断わりのない限り、そのような要素の順序付けは含意されないことに留意されたい。ソフトウェアは、ハードディスク内に存在してもよく、あるいは、コンピュータ・システムによるその実行中、完全にまたは少なくとも部分的に、RAM内および/またはプロセッサ内に存在してもよい。よって、メモリおよびプロセッサはまた、コンピュータ読み取り可能なコードを担持するコンピュータ読み取り可能な担体媒体を構成する。

【0094】

代替的な例示的实施形態では、前記一つまたは複数のプロセッサは、スタンドアロン装置として動作するか、あるいは、他のプロセッサに接続、たとえばネットワーク接続されてもよく、ネットワーク接続された展開では、前記一つまたは複数のプロセッサは、サーバー ユーザー・ネットワーク環境におけるサーバーまたはユーザーマシンの資格で動作してもよく、または、ピアツーピアまたは分散ネットワーク環境におけるピアマシンとして動作してもよい。前記一つまたは複数のプロセッサは、パーソナルコンピュータ（PC）、タブレットPC、パーソナルデジタルアシスタント（PDA）、セルラー電話、ウェブブラウザ、ネットワークルータ、スイッチまたはブリッジ、または、そのマシンによって実行されるべきアクションを指定する一組の命令（逐次的またはその他）を実行することができる任意のマシンを形成することができる。

【0095】

「機械/マシン」という用語は、本明細書で議論される方法論のいずれか一つまたは複

10

20

30

40

50

数を実行するための命令の集合（または複数の集合）を個別にまたは合同して実行する機械の任意の集まりを含むとも解釈されることに留意されたい。

【0096】

このように、本明細書に記載される各方法のある例示的な実施形態は、命令の集合、たとえば、一つまたは複数のプロセッサ、たとえば、ウェブサーバー構成の一部である一つまたは複数のプロセッサ上で実行するためのコンピュータ・プログラムを担持するコンピュータ読み取り可能な担体媒体の形である。よって、当業者には理解されるであろうように、本開示の例示的な実施形態は、方法、特殊目的装置のような装置、データ処理システムのような装置、またはコンピュータ読み取り可能な担体媒体、たとえば、コンピュータ・プログラム・プロダクトとして具現されうる。コンピュータ読み取り可能な担体媒体は、一つまたは複数のプロセッサ上で実行されたときに該プロセッサに方法を実施させる命令の集合を含むコンピュータ読み取り可能なコードを担持する。よって、本開示の諸側面は、方法、完全にハードウェアの例示的な実施形態、完全にソフトウェアの例示的な実施形態、またはソフトウェアとハードウェアの側面を組み合わせた例示的な実施形態の形をとることができる。さらに、本開示は、媒体内に具現されたコンピュータ読み取り可能なプログラムコードを担持する担体媒体（たとえば、コンピュータ読み取り可能な記憶媒体上のコンピュータ・プログラム・プロダクト）の形をとることができる。

10

【0097】

ソフトウェアは、さらに、ネットワーク・インターフェース装置を介してネットワークを通じて送受信されてもよい。担体媒体は、ある例示的な実施形態では単一の媒体であるが、用語「担体媒体」は、命令の一つまたは複数の集合を記憶する単一の媒体または複数の媒体（たとえば、中央集中式のまたは分散式のデータベース、および/または関連するキャッシュおよびサーバー）を含むと解されるべきである。用語「担体媒体」はまた、前記プロセッサの一つまたは複数による実行のための、前記一つまたは複数のプロセッサに本開示の方法論のいずれか一つまたは複数を実行させる命令の集合を記憶、エンコードまたは担持することができる任意の媒体を含むとも解されるべきである。担体媒体は、限定されるものではないが、不揮発性媒体、揮発性媒体、および伝送媒体を含む多くの形をとることができる。不揮発性媒体は、たとえば、光ディスク、磁気ディスク、および磁気光学ディスクを含む。揮発性媒体は、メインメモリのようなダイナミックメモリを含む。伝送媒体は、バスサブシステムを構成するワイヤを含む、同軸ケーブル、銅線および光ファイバーを含む。伝送媒体は、電波および赤外線データ通信の間に発生されるような、音波または光波の形をとることもできる。たとえば、「担体媒体」という用語は、固体メモリ、光学および磁気媒体において具現されたコンピュータ・プロダクト；少なくとも1つのプロセッサまたは一つまたは複数のプロセッサによって検出可能な伝搬信号を担持し、実行時に方法を実装する命令の集合を表す媒体；および前記一つまたは複数のプロセッサの少なくとも1つのプロセッサによって検出可能な、命令の集合を表す伝搬信号を担う、ネットワーク内の伝送媒体を含むが、これらに限定されない。

20

30

【0098】

論じられた諸方法のステップは、ある例示的な実施形態では、記憶装置に記憶された命令（コンピュータ読み取り可能なコード）を実行する処理（たとえばコンピュータ）システムの適切なプロセッサ（単数または複数）によって実行されることが理解されよう。本開示は、いかなる特定の实装やプログラミング技法にも限定されず、本開示は、本明細書に記載される機能性を実施するための任意の適切な技法を用いて実装されうることも理解されるであろう。本開示は、特定のプログラミング言語またはオペレーティングシステムに限定されない。

40

【0099】

本開示を通じ、「一つの例示的な実施形態」、「いくつかの例示的な実施形態」、または「ある例示的な実施形態」への言及は、その例示的な実施形態に関連して記載された特定の特徴、構造または特性が本開示の少なくとも一つの例示的な実施形態に含まれることを意味する。よって、本開示を通じたさまざまな箇所における「一つの例示的な実施形態

50

において」、「いくつかの例示的な実施形態において」または「ある例示的な実施形態において」という句の出現は、必ずしもすべてが同じ実施形態を指すものではない。さらに、特定の特徴、構造または特性は、本開示から当業者に明らかなように、一つまたは複数の例示的な実施形態において、任意の好適な仕方でも組み合わせられてもよい。

【0100】

本明細書中で使用される場所では、共通の対象を記述するための順序形容詞「第1の」、「第2の」、「第3の」などの使用は、単に、同様の対象の異なるインスタンスが言及されていることを示すのであり、そのように記述される対象が、時間的に、空間的に、リンク付けにおいて、または他の任意の仕方でも、所与のシーケンスになければならないことを意味することは意図されていない。

10

【0101】

下記の特許請求の範囲および本明細書中の説明において、含む、からなる、または有するといった用語は、記載される要素/特徴を少なくとも含むが他のものを除外しないことを意味するオープンな用語である。よって、特許請求の範囲において使用される場合、含む/有するの用語は、列挙される手段または要素またはステップに限定されるものとして解釈されるべきではない。たとえば、AおよびBを有する装置という表現の範囲は、要素AおよびBのみからなる装置に限定されるべきではない。本明細書で使用される含む、または含んでいる、または包含するといった用語もいずれも、記載される要素/特徴を少なくとも含むが他のものを除外しないことを意味するオープンな用語である。よって、含むは、有すると同義であり、有するを意味する。

20

【0102】

本開示の例示的な実施形態の上述の説明では、開示の流れをよくし、さまざまな発明的側面の一つまたは複数の理解を助ける目的で、本開示のさまざまな特徴が、単一の例示的な実施形態、図、またはその説明にまとめられることがあることが理解されるべきである。しかしながら、この開示方法は、クレームが各クレームにおいて明示的に記載されているよりも多くの特徴を必要とするという意図を反映するものと解釈されるべきではない。むしろ、以下の特許請求の範囲が反映するように、本発明の側面は、単一の前述の開示された例示的な実施形態のすべての特徴よりも少ないものにある。よって、本説明に続くクレームは、本明細書に明示的に組み込まれ、各クレームが本開示の別個の例示的な実施形態としてそれ自体で自立する。

30

【0103】

さらに、本明細書に記載されるいくつかの例示的な実施形態は、他の例示的な実施形態に含まれるいくつかの特徴を含み、他の例示的な実施形態に含まれる他の特徴は含まないが、異なる例示的な実施形態の特徴の組み合わせは、当業者に理解されるように、本開示の範囲内であり、異なる例示的な実施形態を形成することが意図されている。たとえば、以下の請求項では、請求項に記載された例示的な実施形態の任意のものが、任意の組み合わせで使用できる。

【0104】

本明細書に提供される説明において、多数の個別的な詳細が記載されている。しかしながら、本開示の例示的な実施形態は、これらの個別的な詳細なしで実施されてもよいことが理解される。他の例では、周知の方法、構造および技術は、本稿の理解を不明瞭にしないために、詳細には示されていない。

40

【0105】

よって、開示の最良の態様であると考えられるものが記載されてきたが、当業者は、開示の精神から逸脱することなく、それに他のさらなる修正がなされてもよく、本開示の範囲にはいるすべてのそのような変更および修正を特許請求することが意図されていることを認識するであろう。たとえば、上記で与えられた公式があったとすれば、それは単に使用される手順を表すにすぎない。ブロック図から機能が追加または削除されてもよく、機能性ブロック間で動作が交換されてもよい。本開示の範囲内で、記載された方法にステップが追加または削除されてもよい。

50

【 0 1 0 6 】

本発明のさまざまな側面は、以下の箇条書き例示的实施形態 (enumerated example embodiment、EEE) から理解されうる。

【 0 1 0 7 】

〔 EEE 1 〕

デコーダにおいて実装されたニューラルネットワークを更新するためのパラメータを伝送するためのメディアビットストリームを生成する方法であって、当該方法は：

(a) 前記ニューラルネットワークを更新するためのパラメータの少なくとも1つの集合を決定するステップと；

(b) パラメータの前記少なくとも1つの集合およびメディアデータをエンコードして前記メディアビットストリームを生成するステップと；

(c) 前記ニューラルネットワークをパラメータの前記少なくとも1つの集合を用いて更新するために、前記デコーダに前記メディアビットストリームを伝送するステップとを含む、方法。

10

〔 EEE 2 〕

前記メディアデータは、オーディオデータおよび/またはビデオデータの一つまたは複数を含む、EEE 1 に記載の方法。

〔 EEE 3 〕

パラメータの前記少なくとも1つの集合は、構文要素の集合に基づいてエンコードされる、EEE 1 またはEEE 2 に記載の方法。

20

〔 EEE 4 〕

ステップ (a) において、前記ニューラルネットワークを更新するためのパラメータの2つ以上の集合が決定され、構文要素の前記集合は、実行される前記ニューラルネットワークのそれぞれの更新のためのパラメータのそれぞれの集合を同定する一つまたは複数の構文要素を含む、EEE 3 に記載の方法。

〔 EEE 5 〕

前記デコーダにおいて実装される前記ニューラルネットワークは、メディアデータの処理のために使用され、前記メディアビットストリームにおいて、前記ニューラルネットワークを更新するためのパラメータの前記少なくとも1つの集合は、前記ニューラルネットワークによって処理される前記メディアデータと時間整列される、EEE 1 ないし 4 のうちいずれか一項に記載の方法。

30

〔 EEE 6 〕

パラメータの前記少なくとも1つの集合は、コーデック・モード、前記メディアデータの内容、およびエンコード制約条件のうちの一つまたは複数に基づいて決定される、EEE 5 に記載の方法。

〔 EEE 7 〕

前記コーデック・モードは、ビットレート、ビデオおよび/またはオーディオ・フレームレート、および使用されるコア・コーデックのうちの一つまたは複数を含む、EEE 6 に記載の方法。

〔 EEE 8 〕

メディアデータの内容は、発話、音楽、および拍手のうちの一つまたは複数を含む、EEE 6 またはEEE 7 に記載の方法。

40

〔 EEE 9 〕

前記エンコード制約条件は、パフォーマンス・スケーラビリティについての制約条件および適応処理についての制約条件のうちの一つまたは複数を含む、EEE 6 ないし 8 のうちいずれか一項に記載の方法。

〔 EEE 1 0 〕

前記メディアデータはMPEG-HオーディオまたはMPEG-Iオーディオ・フォーマットであり、前記メディアビットストリームはMHASフォーマットの packets 化されたメディアビットストリームである、EEE 1 ないし 9 のうちいずれか一項に記載の方法。

50

〔 EEE 1 1 〕

パラメータの前記少なくとも1つの集合は、パラメータの前記少なくとも1つの集合を、新規のMHASパケットタイプの一つまたは複数のMHASパケットにカプセル化することによってエンコードされる、EEE 1 0に記載の方法。

〔 EEE 1 2 〕

前記メディアデータが、AC-4、AC-3またはEAC-3フォーマットである、EEE 1 ないし 1 1のうちいずれか一項に記載の方法。

〔 EEE 1 3 〕

パラメータの前記少なくとも1つの集合は、一つまたは複数のペイロード要素として前記メディアビットストリームにおいてエンコードされる、EEE 1 2に記載の方法。

〔 EEE 1 4 〕

前記メディアデータは、MPEG-4またはMPEG-D USACフォーマットである、EEE 1 ないし 1 3のうちいずれか一項に記載の方法。

〔 EEE 1 5 〕

パラメータの前記少なくとも1つの集合は、一つまたは複数のペイロード要素として、または一つまたは複数のデータストリーム要素として、前記メディアビットストリームにおいてエンコードされる、EEE 1 4に記載の方法。

〔 EEE 1 6 〕

前記ニューラルネットワークは層構造を有しており、パラメータの前記少なくとも1つの集合は、前記ニューラルネットワークの、少なくともメディアデータ側層および/または出力層の重みを更新するためのパラメータを含む、EEE 1 ないし 1 5のうちいずれか一項に記載の方法。

〔 EEE 1 7 〕

デコーダにおいて実装されるニューラルネットワークを更新する方法であって、当該方法は：

(a) メディアデータと、前記ニューラルネットワークを更新するためのパラメータの少なくとも1つの集合とを含む符号化メディアビットストリームを受領するステップと；

(b) 受領されたメディアビットストリームをデコードして、デコードされたメディアデータと、前記ニューラルネットワークを更新するためのパラメータの前記少なくとも1つの集合とを取得するステップと；

(c) 前記デコーダによって、パラメータの前記少なくとも1つの集合を用いて前記ニューラルネットワークを更新するステップとを含む、方法。

〔 EEE 1 8 〕

前記メディアデータは、オーディオデータおよび/またはビデオデータの一つまたは複数を含む、EEE 1 7に記載の方法。

〔 EEE 1 9 〕

当該方法は、さらに、構文要素の集合を受領し、前記デコーダによって、パラメータの前記少なくとも1つの集合を用いて前記ニューラルネットワークを更新することを含み、前記デコーダによって、パラメータの前記少なくとも1つの集合を用いて前記ニューラルネットワークを更新することは、受領された構文要素の集合に基づく、EEE 1 7またはEEE 1 8に記載の方法。

〔 EEE 2 0 〕

ステップ(a)において、前記ニューラルネットワークを更新するためのパラメータの2つ以上の集合が、受領された符号化メディアビットストリームに含まれ、受領された構文要素の集合は、実行される前記ニューラルネットワークのそれぞれの更新のためのパラメータのそれぞれの集合を識別する一つまたは複数の構文要素を含む、EEE 1 9に記載の方法。

〔 EEE 2 1 〕

前記ニューラルネットワークは、メディアデータの処理のために使用され、受領された

10

20

30

40

50

符号化メディアビットストリームにおいて、パラメータの前記少なくとも1つの集合は、前記ニューラルネットワークによって処理されるメディアデータと時間整列される、EEE 17ないし20のうちいずれか一項に記載の方法。

〔EEE 22〕

前記ニューラルネットワークは、層構造を有しており、前記ニューラルネットワークの、少なくともメディアデータ側層および/または出力層の重みは、前記デコーダによって、パラメータの前記少なくとも1つの集合を用いて更新される、EEE 17ないし21のうちいずれか一項に記載の方法。

〔EEE 23〕

デコーダにおいて実装されるニューラルネットワークを更新するためのパラメータを伝送するためのメディアビットストリームを生成するための装置であって、当該装置は：

(a) 前記ニューラルネットワークを更新するためのパラメータの少なくとも1つの集合を決定するステップと；

(b) パラメータの前記少なくとも1つの集合およびメディアデータをエンコードして、前記メディアビットストリームを生成するステップと；

(c) パラメータの前記少なくとも1つの集合を用いて前記ニューラルネットワークを更新するために、前記メディアビットストリームを前記デコーダに伝送するステップを含む方法を実行するように構成されたプロセッサを含む、装置。

〔EEE 24〕

デコーダにおいて実装されるニューラルネットワークを更新するための装置であって、当該装置は、

(a) メディアデータと、前記ニューラルネットワークを更新するためのパラメータの少なくとも1つの集合とを含む符号化メディアビットストリームを受領するための受領器と；

(b) 受領されたメディアビットストリームをデコードして、デコードされたメディアデータと、前記ニューラルネットワークを更新するためのパラメータの前記少なくとも1つの集合とを得るためのデコーダと；

(c) パラメータの前記少なくとも1つの集合を用いて前記ニューラルネットワークを更新するための更新器とを含む、装置。

〔EEE 25〕

前記受領器は、さらに、構文要素の集合を受領し、前記更新器は、前記構文要素の集合に基づいて、パラメータの前記少なくとも1つの集合を用いて前記ニューラルネットワークを更新する、EEE 24に記載の装置。

〔EEE 26〕

前記ニューラルネットワークを更新するためのパラメータの2つ以上の集合が、前記受領器によって受領される前記符号化メディアビットストリームに含まれており、前記受領された構文要素の集合は、実行される前記ニューラルネットワークのそれぞれの更新のためのパラメータのそれぞれの集合を識別する一つまたは複数の構文要素を含む、EEE 25に記載の装置。

〔EEE 27〕

処理能力を有する装置によって実行されたときに、該装置に、EEE 1ないし16のうちいずれか一項に記載の方法を実行させるように適応された命令を有するコンピュータ読み取り可能な記憶媒体を有するコンピュータ・プログラム・プロダクト。

〔EEE 28〕

処理能力を有する装置によって実行されたときに、該装置に、EEE 17ないし22のうちいずれか一項に記載の方法を実行させるように適応された命令を有するコンピュータ読み取り可能な記憶媒体を有するコンピュータ・プログラム・プロダクト。

10

20

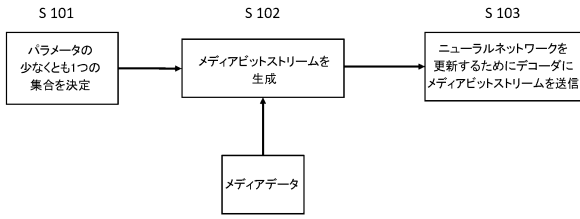
30

40

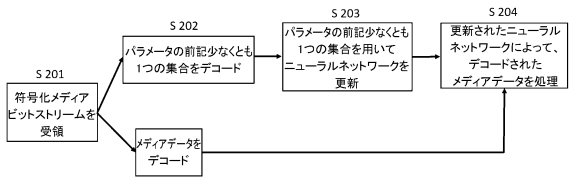
50

【図面】

【図 1】

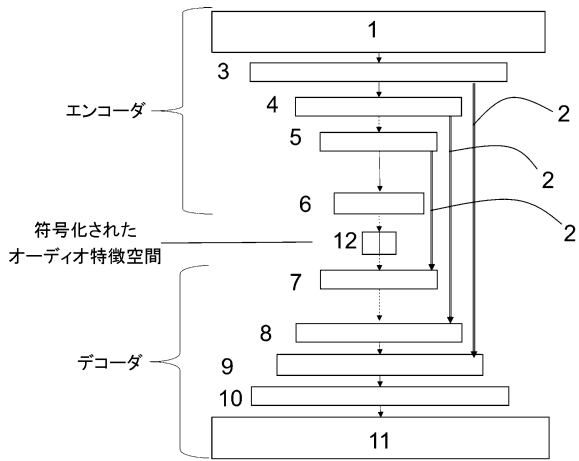


【図 2】



10

【図 3】



20

30

40

50

フロントページの続き

(33)優先権主張国・地域又は機関

欧州特許庁(EP)

イッチュヘルンシュトラッセ 15 - 19 ドルビー ジャーマニー ゲーエムベーハー 内

(72)発明者 ビスワス, アリジート

ドイツ 90429 ニュルンベルク ドイッチュヘルンシュトラッセ 15 - 19 ドルビー ジャ
ーマニー ゲーエムベーハー 内

審査官 多賀 実

(56)参考文献 国際公開第2016/199330(WO, A1)

国際公開第2018/150083(WO, A1)

国際公開第2018/163011(WO, A1)

特開平02-032679(JP, A)

(58)調査した分野 (Int.Cl., DB名)

G06N 3/00 - 99/00

G06T 7/00

G06T 9/00

G10L 13/00 - 99/00

H04N 19/00 - 21/858