【公報種別】特許法第１７条の２の規定による補正の掲載
【部門区分】第６部門第３区分
【発行日】令和6年10月29日(2024.10.29)

【国際公開番号】WO2022/115838
【公表番号】特表2023-550092(P2023-550092A)
【公表日】令和5年11月30日(2023.11.30)
【年通号数】公開公報(特許)2023-225
【出願番号】特願2023-529961(P2023-529961)
【国際特許分類】
G06N 3/063(2023.01)
G10L 25/51(2013.01)
【FI】
G06N 3/063
G10L 25/51

【手続補正書】
【提出日】令和6年10月21日(2024.10.21)
【手続補正１】
【補正対象書類名】特許請求の範囲
【補正対象項目名】全文
【補正方法】変更
【補正の内容】
【特許請求の範囲】
【請求項１】
デバイスであって、
１つまたは複数のプロセッサを備え、前記１つまたは複数のプロセッサは、
サウンドイベント分類モデルにオーディオデータサンプルを与えることと、
前記オーディオデータサンプルに応答した前記サウンドイベント分類モデルの出力に基づいて、前記オーディオデータサンプルのサウンドクラスが前記サウンドイベント分類モデルによって認識されたのかどうかを決定することと、
前記サウンドクラスが認識されなかったという決定に基づいて、前記サウンドイベント分類モデルが前記オーディオデータサンプルに関連するオーディオシーン中のサウンドイベントを認識するようにトレーニングされているのかどうかを決定することと、
前記サウンドイベント分類モデルが前記オーディオデータサンプルに関連する前記オーディオシーン中のサウンドイベントを認識するようにトレーニングされている、および前記サウンドクラスが認識されなかったという決定に基づいて、前記サウンドイベント分類モデルを更新するための前記オーディオデータサンプルに基づいてモデル更新データを記憶することと
を行うように構成された、デバイス。
【請求項２】
前記１つまたは複数のプロセッサに結合され、前記オーディオデータサンプルに対応するオーディオデータをキャプチャするように構成されたマイクロフォンをさらに備える、請求項１に記載のデバイス。
【請求項３】
前記１つまたは複数のプロセッサに結合され、複数のサウンドイベント分類モデルを記憶すること、ここにおいて、前記１つまたは複数のプロセッサは、前記複数のサウンドイベント分類モデルの中から前記サウンドイベント分類モデルを選択するように構成される、を行うように構成されたメモリをさらに備える、請求項１に記載のデバイス。
【請求項４】

前記デバイスは、前記オーディオデータサンプルに関連するセンサデータを生成するように構成された１つまたは複数のセンサ、ここにおいて、前記１つまたは複数のプロセッサは、前記センサデータに基づいて前記サウンドイベント分類モデルを選択するように構成され、前記１つまたは複数のセンサは、カメラおよび／または位置センサとを含む、をさらに備える、または、
前記デバイスは、前記オーディオシーンを識別する入力を受信するように構成された１つまたは複数の入力デバイス、ここにおいて、前記１つまたは複数のプロセッサは、前記オーディオシーンに基づいて前記サウンドイベント分類モデルを選択するように構成される、をさらに備える、または、
前記１つまたは複数のプロセッサは、前記オーディオデータサンプルがいつ受信されたのかに基づいて前記サウンドイベント分類モデルを選択するように構成された、または、
前記メモリは、１つまたは複数のデバイス設定を示す設定データをさらに記憶し、前記１つまたは複数のプロセッサは、前記設定データに基づいて前記サウンドイベント分類モデルを選択するように構成された、請求項３に記載のデバイス。

【請求項５】

前記１つまたは複数のプロセッサは、前記サウンドクラスが認識されたという決定に基づいて、前記オーディオデータサンプルに関連する前記サウンドクラスを示す出力を生成するようにさらに構成された、請求項１に記載のデバイス。

【請求項６】

前記サウンドイベント分類モデルは、前記出力に関連する信頼性メトリックを生成するようにさらに構成され、前記１つまたは複数のプロセッサは、前記サウンドクラスが前記信頼性メトリックに基づいて前記サウンドイベント分類モデルによって認識されたのかどうかを決定するように構成された、請求項１に記載のデバイス。

【請求項７】

前記１つまたは複数のプロセッサは、前記モデル更新データに基づいて前記サウンドイベント分類モデルを更新するようにさらに構成された、請求項１に記載のデバイス。

【請求項８】

前記オーディオシーンを識別する入力を受信するように構成された１つまたは複数の入力デバイス、ここにおいて、前記１つまたは複数のプロセッサは、前記入力に基づいて前記サウンドイベント分類モデルが前記オーディオシーン中のサウンドイベントを認識するようにトレーニングされているのかどうかを決定するように構成される、をさらに備える、請求項１に記載のデバイス。

【請求項９】

前記オーディオデータサンプルに関連するセンサデータを生成するように構成された１つまたは複数のセンサ、ここにおいて、前記１つまたは複数のプロセッサは、前記センサデータに基づいて前記サウンドイベント分類モデルが前記オーディオシーン中のサウンドイベントを認識するようにトレーニングされているのかどうかを決定するように構成され、前記１つまたは複数のセンサは、カメラと位置センサとを含む、または、
前記１つまたは複数のプロセッサは、前記オーディオデータサンプルに関連するタイムスタンプに基づいて前記サウンドイベント分類モデルが前記オーディオシーン中のサウンドイベントを認識するようにトレーニングされているのかどうかを決定するようにさらに構成される、
をさらに備える、請求項１に記載のデバイス。

【請求項１０】

前記サウンドイベント分類モデルは、特定のサウンドクラスを認識するようにトレーニングされ、前記モデル更新データは、前記特定のサウンドクラスに対応するものとして前記サウンドイベント分類モデルが認識するようにトレーニングされない前記特定のサウンドクラス内のサウンドの特性の変形形態を表すドリフトデータを含む、請求項１に記載のデバイス。

【請求項１１】

前記１つまたは複数のプロセッサは、モバイルコンピューティングデバイス、車両、ウェアラブルデバイス、拡張現実ヘッドセット、複合現実ヘッドセット、または仮想現実ヘッドセット内に統合される、請求項１に記載のデバイス。
【請求項１２】
方法であって、
１つまたは複数のプロセッサによって、サウンドイベント分類モデルに入力としてオーディオデータサンプルを与えることと、
前記１つまたは複数のプロセッサによって、前記オーディオデータサンプルに応答した前記サウンドイベント分類モデルの出力に基づいて、前記オーディオデータサンプルのサウンドクラスが前記サウンドイベント分類モデルによって認識されたのかどうかを決定することと、
前記サウンドクラスが認識されなかったという決定に基づいて、前記１つまたは複数のプロセッサによって、前記サウンドイベント分類モデルが前記オーディオデータサンプルに関連するオーディオシーン中のサウンドイベントを認識するようにトレーニングされているのかどうかを決定することと、
前記サウンドイベント分類モデルが前記オーディオデータサンプルに関連する前記オーディオシーン中のサウンドイベントを認識するようにトレーニングされている、および前記サウンドクラスが認識されなかったという決定に基づいて、前記１つまたは複数のプロセッサによって、前記サウンドイベント分類モデルを更新するための前記オーディオデータサンプルに基づいてモデル更新データを記憶することと
を備える方法。
【請求項１３】
前記モデル更新データを記憶した後に、
モデル更新データのしきい値量が累積されたのかどうかを決定することと、
モデル更新データの前記しきい値量が累積されたという決定に基づいて、累積モデル更新データを使用して前記サウンドイベント分類モデルの自動更新を開始することと
を行うことをさらに備える、請求項１２に記載の方法。
【請求項１４】
前記自動更新の前に、前記サウンドイベント分類モデルは、特定のサウンドクラスの複数の変形を認識するようにトレーニングされており、前記自動更新は、前記特定のサウンドクラスに対応するものとして前記特定のサウンドクラスの追加の変形を認識することを前記サウンドイベント分類モデルが行うことを可能にするために前記サウンドイベント分類モデルを修正する、請求項１３に記載の方法。
【請求項１５】
非一時的コンピュータ可読媒体であって、プロセッサに、請求項１２乃至１４のいずれか一項に記載の方法を実施させるために前記プロセッサによって実行可能である命令を記憶する非一時的コンピュータ可読媒体。
【手続補正２】
【補正対象書類名】明細書
【補正対象項目名】０２０２
【補正方法】変更
【補正の内容】
【０２０２】
[0217] 開示された態様の上記の説明は、当業者が開示された態様を製作または使用することを可能にするために提供されている。これらの態様への様々な修正が当業者には容易に明らかになり、本明細書で定義された原理が、本開示の範囲から逸脱することなく他の態様に適用され得る。したがって、本開示は、本明細書に示された態様に限定されることを意図されておらず、以下の特許請求の範囲によって定義されるような原理および新規な特徴と一致する可能な最も広い範囲を与えられるべきである。
以下に本願の出願当初の特許請求の範囲に記載された発明を付記する。

［Ｃ１］
デバイスであって、
１つまたは複数のプロセッサを備え、前記１つまたは複数のプロセッサは、
サウンドイベント分類モデルにオーディオデータサンプルを与えることと、
前記オーディオデータサンプルに応答した前記サウンドイベント分類モデルの出力に基づいて、前記オーディオデータサンプルのサウンドクラスが前記サウンドイベント分類モデルによって認識されたのかどうかを決定することと、
前記サウンドクラスが認識されなかったという決定に基づいて、前記サウンドイベント分類モデルが前記オーディオデータサンプルに関連するオーディオシーンに対応するのかどうかを決定することと、
前記サウンドイベント分類モデルが前記オーディオデータサンプルに関連する前記オーディオシーンに対応するという決定に基づいて、前記オーディオデータサンプルに基づいてモデル更新データを記憶することと
を行うように構成された、デバイス。
［Ｃ２］
前記１つまたは複数のプロセッサに結合され、前記オーディオデータサンプルに対応するオーディオデータをキャプチャするように構成されたマイクロフォンをさらに備える、Ｃ１に記載のデバイス。
［Ｃ３］
前記１つまたは複数のプロセッサに結合され、複数のサウンドイベント分類モデルを記憶すること、ここにおいて、前記１つまたは複数のプロセッサは、前記複数のサウンドイベント分類モデルの中から前記サウンドイベント分類モデルを選択するように構成される、を行うように構成されたメモリをさらに備える、Ｃ１に記載のデバイス。
［Ｃ４］
前記オーディオデータサンプルに関連するセンサデータを生成するように構成された１つまたは複数のセンサ、ここにおいて、前記１つまたは複数のプロセッサは、前記センサデータに基づいて前記サウンドイベント分類モデルを選択するように構成される、をさらに備える、Ｃ３に記載のデバイス。
［Ｃ５］
前記１つまたは複数のセンサはカメラと位置センサとを含む、Ｃ４に記載のデバイス。
［Ｃ６］
前記オーディオシーンを識別する入力を受信するように構成された１つまたは複数の入力デバイス、ここにおいて、前記１つまたは複数のプロセッサは、前記オーディオシーンに基づいて前記サウンドイベント分類モデルを選択するように構成される、をさらに備える、Ｃ３に記載のデバイス。
［Ｃ７］
前記１つまたは複数のプロセッサは、前記オーディオデータサンプルがいつ受信されたのかに基づいて前記サウンドイベント分類モデルを選択するように構成された、Ｃ３に記載のデバイス。
［Ｃ８］
前記メモリは、１つまたは複数のデバイス設定を示す設定データをさらに記憶し、前記１つまたは複数のプロセッサは、前記設定データに基づいて前記サウンドイベント分類モデルを選択するように構成された、Ｃ３に記載のデバイス。
［Ｃ９］
前記１つまたは複数のプロセッサは、前記サウンドクラスが認識されたという決定に基づいて、前記オーディオデータサンプルに関連する前記サウンドクラスを示す出力を生成するようにさらに構成された、Ｃ１に記載のデバイス。
［Ｃ１０］
前記１つまたは複数のプロセッサは、前記サウンドイベント分類モデルが前記オーディオデータサンプルに関連する前記オーディオシーンに対応しないという決定に基づいて、

新しいサウンドイベント分類モデルのためのトレーニングデータとして前記オーディオデータサンプルに対応するオーディオデータを記憶するようにさらに構成された、Ｃ１に記載のデバイス。

［Ｃ１１］

前記サウンドイベント分類モデルは、前記出力に関連する信頼性メトリックを生成するようにさらに構成され、前記１つまたは複数のプロセッサは、前記サウンドクラスが前記信頼性メトリックに基づいて前記サウンドイベント分類モデルによって認識されたのかどうかを決定するように構成された、Ｃ１に記載のデバイス。

［Ｃ１２］

前記１つまたは複数のプロセッサは、前記モデル更新データに基づいて前記サウンドイベント分類モデルを更新するようにさらに構成された、Ｃ１に記載のデバイス。

［Ｃ１３］

前記オーディオシーンを識別する入力を受信するように構成された１つまたは複数の入力デバイス、ここにおいて、前記１つまたは複数のプロセッサは、前記入力に基づいて前記サウンドイベント分類モデルが前記オーディオシーンに対応するのかどうかを決定するように構成される、をさらに備える、Ｃ１に記載のデバイス。

［Ｃ１４］

前記オーディオデータサンプルに関連するセンサデータを生成するように構成された１つまたは複数のセンサ、ここにおいて、前記１つまたは複数のプロセッサは、前記センサデータに基づいて前記サウンドイベント分類モデルが前記オーディオシーンに対応するのかどうかを決定するように構成される、をさらに備える、Ｃ１に記載のデバイス。

［Ｃ１５］

前記１つまたは複数のセンサはカメラと位置センサとを含む、Ｃ１４に記載のデバイス。

［Ｃ１６］

前記１つまたは複数のプロセッサは、前記オーディオデータサンプルに関連するタイムスタンプに基づいて前記サウンドイベント分類モデルが前記オーディオシーンに対応するのかどうかを決定するようにさらに構成された、Ｃ１４に記載のデバイス。

［Ｃ１７］

前記サウンドイベント分類モデルは、特定のサウンドクラスを認識するようにトレーニングされ、前記モデル更新データは、前記特定のサウンドクラスに対応するものとして前記サウンドイベント分類モデルが認識するようにトレーニングされない前記特定のサウンドクラス内のサウンドの特性の変形形態を表すドリフトデータを含む、Ｃ１に記載のデバイス。

［Ｃ１８］

前記１つまたは複数のプロセッサは、モバイルコンピューティングデバイス、車両、ウェアラブルデバイス、拡張現実ヘッドセット、複合現実ヘッドセット、または仮想現実ヘッドセット内に統合される、Ｃ１に記載のデバイス。

［Ｃ１９］

前記１つまたは複数のプロセッサは、集積回路中に含まれる、Ｃ１に記載のデバイス。

［Ｃ２０］

方法であって、

１つまたは複数のプロセッサによって、サウンドイベント分類モデルに入力としてオーディオデータサンプルを与えることと、

前記１つまたは複数のプロセッサによって、前記オーディオデータサンプルに応答した前記サウンドイベント分類モデルの出力に基づいて、前記オーディオデータサンプルのサウンドクラスが前記サウンドイベント分類モデルによって認識されたのかどうかを決定することと、

前記サウンドクラスが認識されなかったという決定に基づいて、前記１つまたは複数のプロセッサによって、前記サウンドイベント分類モデルが前記オーディオデータサンプルに関連するオーディオシーンに対応するのかどうかを決定することと、

前記サウンドイベント分類モデルが前記オーディオデータサンプルに関連する前記オーディオシーンに対応するという決定に基づいて、前記１つまたは複数のプロセッサによって、前記オーディオデータサンプルに基づいてモデル更新データを記憶することと
を備える方法。
［Ｃ２１］
前記１つまたは複数のプロセッサに結合されたメモリに記憶された複数のサウンドイベント分類モデルの中から前記サウンドイベント分類モデルを選択することをさらに備える、Ｃ２０に記載の方法。
［Ｃ２２］
前記サウンドイベント分類モデルは、ユーザ入力、設定データ、ロケーションデータ、画像データ、ビデオデータ、前記オーディオデータサンプルに関連するタイムスタンプ、またはそれらの組合せに基づいて選択される、Ｃ２１に記載の方法。
［Ｃ２３］
前記サウンドイベント分類モデルが前記オーディオシーンに対応するのかどうかの決定は、前記サウンドイベント分類モデル、ユーザ入力、設定データ、ロケーションデータ、画像データ、ビデオデータ、前記オーディオデータサンプルに関連するタイムスタンプ、またはそれらの組合せにより生成される信頼性メトリックに基づく、Ｃ２０に記載の方法。
［Ｃ２４］
前記モデル更新データを記憶した後に、
モデル更新データのしきい値量が累積されたのかどうかを決定することと、
モデル更新データの前記しきい値量が累積されたという決定に基づいて、累積モデル更新データを使用して前記サウンドイベント分類モデルの自動更新を開始することと
を行うことをさらに備える、Ｃ２０に記載の方法。
［Ｃ２５］
前記自動更新の前に、前記サウンドイベント分類モデルは、特定のサウンドクラスの複数の変形を認識するようにトレーニングされており、前記自動更新は、前記特定のサウンドクラスに対応するものとして前記特定のサウンドクラスの追加の変形を認識することを前記サウンドイベント分類モデルが行うことを可能にするために前記サウンドイベント分類モデルを修正する、Ｃ２４に記載の方法。
［Ｃ２６］
デバイスであって、
サウンドイベント分類モデルにオーディオデータサンプルを与えるための手段と、
前記サウンドイベント分類モデルの出力に基づいて、前記オーディオデータサンプルのサウンドクラスが前記サウンドイベント分類モデルによって認識されたのかどうかを決定するための手段と、
前記サウンドクラスが認識されなかったという決定に応答して、前記サウンドイベント分類モデルが前記オーディオデータサンプルに関連するオーディオシーンに対応するのかどうかを決定するための手段と、
前記サウンドイベント分類モデルが前記オーディオデータサンプルに関連する前記オーディオシーンに対応するという決定に応答して、前記オーディオデータサンプルに基づいてモデル更新データを記憶するための手段と
を備えるデバイス。
［Ｃ２７］
選択基準に基づいて複数のサウンドイベント分類モデルの中から前記サウンドイベント分類モデルを選択するための手段をさらに備える、Ｃ２６に記載のデバイス。
［Ｃ２８］
前記モデル更新データに基づいて前記サウンドイベント分類モデルを更新するための手段をさらに備える、Ｃ２６に記載のデバイス。
［Ｃ２９］

非一時的コンピュータ可読媒体であって、プロセッサに、
サウンドイベント分類モデルに入力としてオーディオデータサンプルを与えることと、前記オーディオデータサンプルに応答した前記サウンドイベント分類モデルの出力に基づいて、前記オーディオデータサンプルのサウンドクラスが前記サウンドイベント分類モデルによって認識されたのかどうかを決定することと、
前記サウンドクラスが認識されなかったという決定に基づいて、前記サウンドイベント分類モデルが前記オーディオデータサンプルに関連するオーディオシーンに対応するのかどうかを決定することと、
前記サウンドイベント分類モデルが前記オーディオデータサンプルに関連する前記オーディオシーンに対応するという決定に基づいて、前記オーディオデータサンプルに基づいてモデル更新データを記憶することと
を行わせるために前記プロセッサによって実行可能である命令を記憶する非一時的コンピュータ可読媒体。
［Ｃ３０］
前記命令は、前記プロセッサに、前記モデル更新データに基づいて前記サウンドイベント分類モデルを更新することをさらに行わせる、Ｃ２９に記載の非一時的コンピュータ可読媒体。