



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0154597
(43) 공개일자 2023년11월09일

(51) 국제특허분류(Int. Cl.)
G10L 17/26 (2013.01) G06N 20/00 (2019.01)
G10L 25/24 (2013.01)
(52) CPC특허분류
G10L 17/26 (2013.01)
G06N 20/00 (2021.08)
(21) 출원번호 10-2022-0054210
(22) 출원일자 2022년05월02일
심사청구일자 2022년05월02일

(71) 출원인
계명대학교 산학협력단
대구광역시 달서구 달구벌대로 1095, 계명대학교
산학협력관 201호(신당동)
(72) 발명자
정용주
대구광역시 수성구 청수로 261, 1412동 1401호(황
금동, 캐슬골드파크 4단지)도로명주소 안내도 새
창으로 보기검색결과 주소 복사하기
(74) 대리인
특허법인 무한

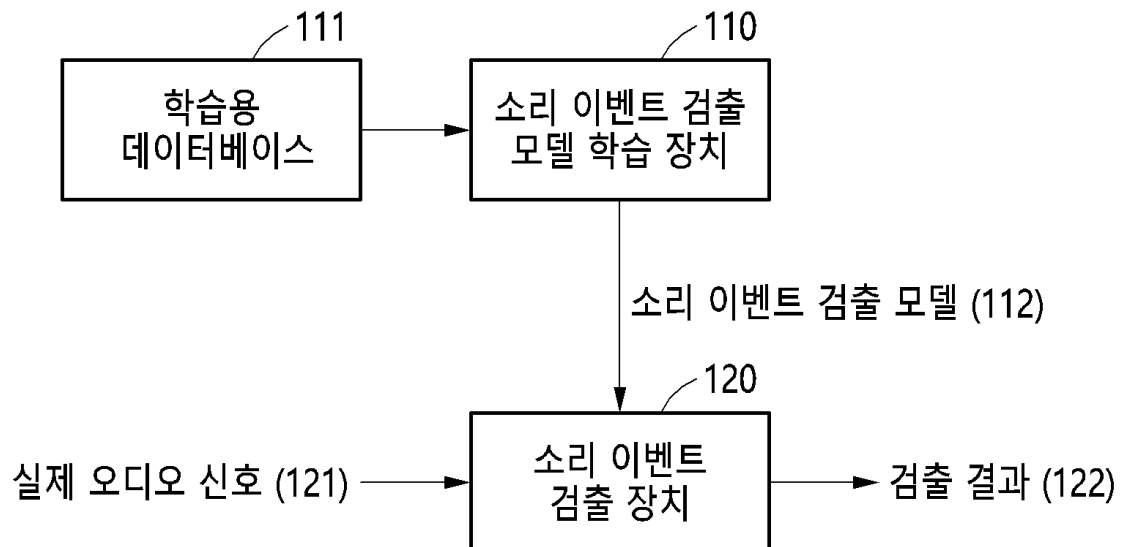
전체 청구항 수 : 총 18 항

(54) 발명의 명칭 차분 특징을 이용한 소리 이벤트 검출 모델 기반의 소리 이벤트 검출 방법 및 장치

(57) 요약

차분 특징을 이용한 소리 이벤트 검출 모델 기반의 소리 이벤트 검출 방법 및 장치가 개시된다. 소리 이벤트 검출 방법은 오디오 신호를 수신하는 단계; 상기 오디오 신호에서 오디오 신호의 원본 특징들을 추출하는 단계; 상기 원본 특징들을 기초로 차분 특징을 추출하는 단계; 상기 원본 특징들 및 상기 차분 특징을 기초로 입력 특징맵을 구성하는 단계; 및 상기 입력 특징맵을 소리 이벤트 검출 모델에 입력하여 검출 결과를 출력하는 단계를 포함할 수 있다.

대표도 - 도1



(52) CPC특허분류

G10L 25/24 (2013.01)

명세서

청구범위

청구항 1

오디오 신호를 수신하는 단계;

상기 오디오 신호에서 오디오 신호의 원본 특징들을 추출하는 단계;

상기 원본 특징들을 기초로 차분 특징을 추출하는 단계;

상기 원본 특징들 및 상기 차분 특징을 기초로 입력 특징맵을 구성하는 단계; 및

상기 입력 특징맵을 소리 이벤트 검출 모델에 입력하여 검출 결과를 출력하는 단계를 포함하는 소리 이벤트 검출 방법.

청구항 2

제1항에 있어서,

상기 원본 특징들을 추출하는 단계는,

상기 오디오 신호를 샘플링하여 STFT(Short-Time Fourier Transform)를 계산하는 단계;

상기 오디오 신호의 STFT값으로부터 멜-필터뱅크 값을 결정하는 단계; 및

상기 멜-필터뱅크 값을 로그 변환한 로그-멜-필터뱅크 값을 상기 원본 특징들로 추출하는 단계를 포함하는 소리 이벤트 검출 방법.

청구항 3

제1항에 있어서,

상기 차분 특징을 추출하는 단계는,

상기 오디오 신호의 프레임들 각각에서 추출한 원본 특징들 간의 차이를 나타내는 1차 차분 특징들을 생성하는 단계; 및

상기 1차 차분 특징들 간의 차이를 나타내는 2차 차분 특징들을 생성하는 단계

를 포함하는 소리 이벤트 검출 방법.

청구항 4

제1항에 있어서,

상기 소리 이벤트 검출 모델은,

학습 데이터의 원본 특징들, 1차 차분 특징들 및 2차 차분 특징들로 구성된 입력 특징맵을 이용하여 학습된 모델인 소리 이벤트 검출 방법.

청구항 5

제1항에 있어서,

상기 소리 이벤트 검출 모델은,

복수의 CNN 블록을 이용하여 상기 입력 특징맵에 대한 차원 축소를 수행한 후, 트랜스포머 인코더를 적용하는 베이스라인 트랜스포머;

컨볼루션 블록과 결합된 트랜스포머 인코더들을 이용하여 학습 데이터의 멀티 스케일 특징을 추출하는 멀티-스

케일 추출 모듈; 및

상기 멀티-스케일 추출 모듈의 출력에 포함된 상기 오디오 신호의 프레임 레벨 데이터로부터 프레임 레벨 예측을 수행하여 상기 오디오 신호의 오디오 클래스를 분류하는 분류 레이어

를 포함하는 소리 이벤트 검출 방법.

청구항 6

학습 데이터를 수신하는 단계;

상기 학습 데이터에서 상기 학습 데이터의 원본 특징들을 추출하는 단계;

상기 원본 특징들을 기초로 차분 특징을 추출하는 단계;

상기 원본 특징들 및 상기 차분 특징을 기초로 입력 특징맵을 구성하는 단계; 및

상기 입력 특징맵을 이용하여 소리 이벤트 검출 모델을 학습하는 단계

를 포함하는 소리 이벤트 검출 모델 학습 방법.

청구항 7

제6항에 있어서,

상기 학습 데이터는,

약전사(Weak label) 레이블, 강전사(Strong label) 레이블, 및 비전사(Un-label) 레이블을 포함하는 소리 이벤트 검출 모델 학습 방법.

청구항 8

제7항에 있어서,

상기 소리 이벤트 검출 모델은,

상기 입력 특징맵의 주파수 영역의 정보를 감소시키는 복수의 컨볼루션 블록들;

쌍방향 GRU(Gated Recurrent Unit);

상기 학습 데이터의 오디오 클래스를 분류하는 분류 레이어(Classification layer); 및

상기 분류 레이어의 출력을 클립(Clip) 전체 구간에 대하여 평균하여 출력하는 글로벌 평균 풀링 레이어(global average pooling layer)

를 포함하고,

상기 학습 데이터가 강전사 레이블인 경우, 상기 분류 레이어의 출력을 소리 이벤트 검출 결과로 출력하며,

상기 학습 데이터가 약전사 레이블인 경우, 상기 분류 레이어의 출력을 입력받은 글로벌 평균 풀링 레이어의 출력을 소리 이벤트 검출 결과로 출력하는 소리 이벤트 검출 모델 학습 방법.

청구항 9

제6항에 있어서,

상기 소리 이벤트 검출 모델은,

복수의 CNN 블록을 이용하여 상기 입력 특징맵에 대한 차원 축소를 수행한 후, 트랜스포머 인코더를 적용하는 베이스라인 트랜스포머;

컨볼루션 블록과 결합된 트랜스포머 인코더들을 이용하여 상기 학습 데이터의 멀티 스케일 특징을 추출하는 멀티-스케일 추출 모듈; 및

상기 멀티-스케일 추출 모듈의 출력을 이용하여 상기 학습 데이터의 오디오 클래스를 분류하는 분류 레이어

를 포함하는 소리 이벤트 검출 모델 학습 방법.

청구항 10

제9항에 있어서,

상기 분류 레이어는,

상기 학습 데이터의 프레임 레벨 데이터와 클립 레벨 데이터로부터 각각 프레임 레벨 예측과 클립 레벨 예측을 수행하여 상기 학습 데이터의 오디오 클래스를 분류하는 소리 이벤트 검출 모델 학습 방법.

청구항 11

제6항에 있어서,

상기 원본 특징들을 추출하는 단계는,

상기 학습 데이터를 샘플링하여 STFT(Short-Time Fourier Transform)를 계산하는 단계;

상기 학습 데이터의 STFT값으로부터 멜-필터뱅크 값을 결정하는 단계; 및

상기 멜-필터뱅크 값을 로그 변환한 로그-멜-필터뱅크 값을 상기 원본 특징들로 추출하는 단계를 포함하는 소리 이벤트 검출 모델 학습 방법.

청구항 12

제6항에 있어서,

상기 차분 특징을 추출하는 단계는,

상기 학습 데이터의 프레임들 각각에서 추출한 원본 특징들 간의 차이를 나타내는 1차 차분 특징들을 생성하는 단계; 및

상기 1차 차분 특징들 간의 차이를 나타내는 2차 차분 특징들을 생성하는 단계를 포함하는 소리 이벤트 검출 모델 학습 방법.

청구항 13

제1항 내지 제12항 중 어느 한 항의 방법을 실행하기 위한 프로그램이 기록된 컴퓨터에서 판독 가능한 기록 매체.

청구항 14

오디오 신호를 오디오 신호에서 오디오 신호의 원본 특징들을 추출하고,

상기 원본 특징들을 기초로 차분 특징을 추출하며,

상기 원본 특징들 및 상기 차분 특징을 기초로 입력 특징맵을 구성하고,

상기 입력 특징맵을 소리 이벤트 검출 모델에 입력하여 검출 결과를 출력하는 프로세서

를 포함하는 소리 이벤트 검출 장치.

청구항 15

제14항에 있어서,

상기 프로세서는,

상기 오디오 신호의 프레임들 각각에서 추출한 원본 특징들 간의 차이를 나타내는 1차 차분 특징들을 생성하고,

상기 1차 차분 특징들 간의 차이를 나타내는 2차 차분 특징들을 생성하는 소리 이벤트 검출 장치

청구항 16

제14항에 있어서,

상기 소리 이벤트 검출 모델은,

학습 데이터의 원본 특징들, 1차 차분 특징들 및 2차 차분 특징들로 구성된 입력 특징맵을 이용하여 학습된 모델인 소리 이벤트 검출 장치

청구항 17

학습 데이터에서 상기 학습 데이터의 원본 특징들을 추출하고,

상기 원본 특징들을 기초로 차분 특징을 추출하며,

상기 원본 특징들 및 상기 차분 특징을 기초로 입력 특징맵을 구성하고,

상기 입력 특징맵을 이용하여 소리 이벤트 검출 모델을 학습하는 프로세서

를 포함하는 소리 이벤트 검출 모델 학습 장치

청구항 18

제17항에 있어서,

상기 학습 데이터는,

약전사(Weak label) 레이블, 강전사(Strong label) 레이블, 및 비전사(Un-label) 레이블을 포함하는 소리 이벤트 검출 모델 학습 장치

발명의 설명

기술 분야

[0001] 본 발명은 소리 이벤트 검출 방법 및 장치에 관한 것으로, 보다 구체적으로는 차분 특징을 이용한 소리 이벤트 검출 모델을 이용하여 소리 이벤트를 검출하는 방법 및 장치에 관한 것이다.

배경 기술

[0002] 소리 이벤트 검출(Sound event detection)은 특정한 소리가 발생하는지에 여부와 함께 발생 시점을 탐지하는 기법이다. 종래의 소리 이벤트 검출 방법은 GMM(Gaussian Mixture Model) 또는 SVM(Support Vector Machine) 및 딥 뉴럴 네트워크(Deep neural network) 기반 방식을 사용하고 있다.

[0003] 종래의 소리 이벤트 검출 방법에 사용된 FNN(Feed Forward Neural Network)은 그 구조가 간단하다는 장점이 있고 구현도 비교적 쉽게 이루어질 수 있으나, 영상이나 오디오 신호에서 발생하는 신호의 변이나 부가된 잡음신호에 대해서 강인하지 못하다는 한계가 있다.

[0004] 따라서, 종래의 소리 이벤트 검출 방법 보다 성능이 향상된 소리 이벤트 검출 방법이 요청되고 있다.

발명의 내용

해결하려는 과제

[0005] 본 발명은 오디오 신호의 원본 특징들과 원본 특징들로부터 추출한 차분 특징들을 입력 특징맵으로 구성하여 소리 이벤트 검출 모델에 사용함으로써, 소리 이벤트 검출 성능을 향상시키는 방법 및 장치를 제공한다.

[0006] 또한, 본 발명은 분류 레이어의 출력을 클립(Clip) 전체 구간에 대하여 평균하여 출력하는 글로벌 평균 풀링 레이어가 포함된 소리 이벤트 검출 모델을 사용함으로써, 강전사(strong label) 레이블만 아니라 약전사(weak label) 레이블 및 비전사(un-label) 레이블도 학습 데이터로 활용가능한 방법 및 장치를 제공한다.

과제의 해결 수단

[0007] 본 발명의 일실시예에 따른 소리 이벤트 검출 방법은 오디오 신호를 수신하는 단계; 상기 오디오 신호에서 오디오 신호의 원본 특징들을 추출하는 단계; 상기 원본 특징들을 기초로 차분 특징을 추출하는 단계; 상기 원본 특

징들 및 상기 차분 특징을 기초로 입력 특징맵을 구성하는 단계; 및 상기 입력 특징맵을 소리 이벤트 검출 모델에 입력하여 검출 결과를 출력하는 단계를 포함할 수 있다.

- [0008] 본 발명의 일실시예에 따른 소리 이벤트 검출 방법의 상기 원본 특징들을 추출하는 단계는, 상기 오디오 신호를 샘플링하여 STFT(Short-Time Fourier Transform)를 계산하는 단계; 상기 오디오 신호의 STFT값으로부터 멜-필터뱅크 값을 결정하는 단계; 및 상기 멜-필터뱅크 값을 로그 변환한 로그-멜-필터뱅크 값을 상기 원본 특징들로 추출하는 단계를 포함할 수 있다.
- [0009] 본 발명의 일실시예에 따른 소리 이벤트 검출 방법의 상기 차분 특징을 추출하는 단계는, 상기 오디오 신호의 프레임들 각각에서 추출한 원본 특징들 간의 차이를 나타내는 1차 차분 특징들을 생성하는 단계; 및 상기 1차 차분 특징들 간의 차이를 나타내는 2차 차분 특징들을 생성하는 단계를 포함할 수 있다.
- [0010] 본 발명의 일실시예에 따른 소리 이벤트 검출 방법의 상기 소리 이벤트 검출 모델은, 학습 데이터의 원본 특징들, 1차 차분 특징들 및 2차 차분 특징들로 구성된 입력 특징맵을 이용하여 학습된 모델일 수 있다.
- [0011] 본 발명의 일실시예에 따른 소리 이벤트 검출 방법의 상기 소리 이벤트 검출 모델은, 복수의 CNN 블록을 이용하여 상기 입력 특징맵에 대한 차원 축소를 수행한 후, 트랜스포머 인코더를 적용하는 베이스라인 트랜스포머; 컨볼루션 블록과 결합된 트랜스포머 인코더들을 이용하여 상기 학습 데이터의 멀티 스케일 특징을 추출하는 멀티-스케일 추출 모듈; 및 상기 멀티-스케일 추출 모듈의 출력에 포함된 상기 오디오 신호의 프레임 레벨 데이터로부터 프레임 레벨 예측을 수행하여 상기 오디오 신호의 오디오 클래스를 분류하는 분류 레이어를 포함할 수 있다.
- [0012] 본 발명의 일실시예에 따른 소리 이벤트 검출 모델 학습 방법은 학습 데이터를 수신하는 단계; 상기 학습 데이터에서 상기 학습 데이터의 원본 특징들을 추출하는 단계; 상기 원본 특징들을 기초로 차분 특징을 추출하는 단계; 상기 원본 특징들 및 상기 차분 특징을 기초로 입력 특징맵을 구성하는 단계; 및 상기 입력 특징맵을 이용하여 소리 이벤트 검출 모델을 학습하는 단계를 포함할 수 있다.
- [0013] 본 발명의 일실시예에 따른 소리 이벤트 검출 모델 학습 방법의 상기 학습 데이터는, 약전사(Weak label) 레이블, 강전사(Strong label) 레이블, 및 비전사(Un-label) 레이블을 포함할 수 있다.
- [0014] 본 발명의 일실시예에 따른 소리 이벤트 검출 모델 학습 방법의 상기 소리 이벤트 검출 모델은, 상기 입력 특징맵의 주파수 영역의 정보를 감소시키는 복수의 컨볼루션 블록들; 쌍방향 GRU(Gated Recurrent Unit); 상기 학습 데이터의 오디오 클래스를 분류하는 분류 레이어(Classification layer); 및 상기 분류 레이어의 출력을 클립(Clip) 전체 구간에 대하여 평균하여 출력하는 글로벌 평균 풀링 레이어(global average pooling layer)를 포함하고, 상기 학습 데이터가 강전사 레이블인 경우, 상기 분류 레이어의 출력을 소리 이벤트 검출 결과로 출력하며, 상기 학습 데이터가 약전사 레이블인 경우, 상기 분류 레이어의 출력을 입력받은 글로벌 평균 풀링 레이어의 출력을 소리 이벤트 검출 결과로 출력할 수 있다.
- [0015] 본 발명의 일실시예에 따른 소리 이벤트 검출 모델 학습 방법의 상기 소리 이벤트 검출 모델은, 복수의 CNN 블록을 이용하여 상기 입력 특징맵에 대한 차원 축소를 수행한 후, 트랜스포머 인코더를 적용하는 베이스라인 트랜스포머; 컨볼루션 블록과 결합된 트랜스포머 인코더들을 이용하여 상기 학습 데이터의 멀티 스케일 특징을 추출하는 멀티-스케일 추출 모듈; 및 상기 멀티-스케일 추출 모듈의 출력을 이용하여 상기 학습 데이터의 오디오 클래스를 분류하는 분류 레이어를 포함할 수 있다.
- [0016] 본 발명의 일실시예에 따른 소리 이벤트 검출 모델 학습 방법의 상기 분류 레이어는, 상기 학습 데이터의 프레임 레벨 데이터와 클립 레벨 데이터로부터 각각 프레임 레벨 예측과 클립 레벨 예측을 수행하여 상기 학습 데이터의 오디오 클래스를 분류할 수 있다.
- [0017] 본 발명의 일실시예에 따른 소리 이벤트 검출 모델 학습 방법의 상기 원본 특징들을 추출하는 단계는, 상기 학습 데이터를 샘플링하여 STFT(Short-Time Fourier Transform)를 계산하는 단계; 상기 학습 데이터의 STFT값으로부터 멜-필터뱅크 값을 결정하는 단계; 및 상기 멜-필터뱅크 값을 로그 변환한 로그-멜-필터뱅크 값을 상기 원본 특징들로 추출하는 단계를 포함할 수 있다.
- [0018] 본 발명의 일실시예에 따른 소리 이벤트 검출 모델 학습 방법의 상기 차분 특징을 추출하는 단계는, 상기 학습 데이터의 프레임들 각각에서 추출한 원본 특징들 간의 차이를 나타내는 1차 차분 특징들을 생성하는 단계; 및 상기 1차 차분 특징들 간의 차이를 나타내는 2차 차분 특징들을 생성하는 단계를 포함할 수 있다. 본 발명의 일실시예에 따른 소리 이벤트 검출 장치는 오디오 신호를 오디오 신호에서 오디오 신호의 원본 특징들을

추출하고, 상기 원본 특징들을 기초로 차분 특징을 추출하며, 상기 원본 특징들 및 상기 차분 특징을 기초로 입력 특징맵을 구성하고, 상기 입력 특징맵을 소리 이벤트 검출 모델에 입력하여 검출 결과를 출력하는 프로세서를 포함할 수 있다.

[0019] 본 발명의 일실시예에 따른 소리 이벤트 검출 장치의 상기 프로세서는, 상기 오디오 신호의 프레임들 각각에서 추출한 원본 특징들 간의 차이를 나타내는 1차 차분 특징들을 생성하고, 상기 1차 차분 특징들 간의 차이를 나타내는 2차 차분 특징들을 생성할 수 있다.

[0020] 본 발명의 일실시예에 따른 소리 이벤트 검출 장치의 상기 소리 이벤트 검출 모델은, 학습 데이터의 원본 특징들, 1차 차분 특징들 및 2차 차분 특징들로 구성된 입력 특징맵을 이용하여 학습된 모델일 수 있다.

[0021] 본 발명의 일실시예에 따른 소리 이벤트 검출 모델 학습 장치는 학습 데이터에서 상기 학습 데이터의 원본 특징들을 추출하고, 상기 원본 특징들을 기초로 차분 특징을 추출하며, 상기 원본 특징들 및 상기 차분 특징을 기초로 입력 특징맵을 구성하고, 상기 입력 특징맵을 이용하여 소리 이벤트 검출 모델을 학습하는 프로세서를 포함할 수 있다.

[0022] 본 발명의 일실시예에 따른 소리 이벤트 검출 모델 학습 장치의 상기 학습 데이터는, 약전사(Weak label) 레이블, 강전사(Strong label) 레이블, 및 비전사(Un-label) 레이블을 포함할 수 있다.

발명의 효과

[0023] 본 발명의 일실시예에 의하면, 오디오 신호의 원본 특징들과 원본 특징들로부터 추출한 차분 특징들을 입력 특징맵으로 구성하여 소리 이벤트 검출 모델에 사용함으로써, 소리 이벤트 검출 성능을 향상시킬 수 있다.

[0024] 또한, 본 발명의 일실시예에 의하면, 분류 레이어의 출력을 클립(Clip) 전체 구간에 대하여 평균하여 출력하는 글로벌 평균 풀링 레이어가 포함된 소리 이벤트 검출 모델을 사용함으로써, 강전사(strong label) 레이블만 아니라 약전사(weak label) 레이블 및 비전사(un-label) 레이블도 학습 데이터로 활용할 수 있다.

도면의 간단한 설명

[0025] 도 1은 본 발명의 일실시예에 따른 소리 이벤트 검출 시스템을 도시한 도면이다.

도 2는 본 발명의 일실시예에 따른 소리 이벤트 검출 모델 학습 방법을 도시한 플로우차트이다.

도 3은 본 발명의 일실시예에 따른 원본 특징 추출 과정을 도시한 도면이다.

도 4는 본 발명의 일실시예에 따른 차분 특징 추출 과정을 도시한 도면이다.

도 5는 본 발명의 일실시예에 따른 입력 특징맵의 일례이다.

도 6은 본 발명의 일실시예에 따른 소리 이벤트 검출 모델의 일례이다.

도 7은 본 발명의 일실시예에 따른 소리 이벤트 검출 모델의 다른 일례이다.

도 8은 본 발명의 일실시예에 따른 소리 이벤트 검출 방법을 도시한 플로우차트이다.

발명을 실시하기 위한 구체적인 내용

[0026] 이하에서, 첨부된 도면을 참조하여 실시예들을 상세하게 설명한다. 그러나, 실시예들에는 다양한 변경이 가해질 수 있어서 특허출원의 권리 범위가 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 실시예들에 대한 모든 변경, 균등물 내지 대체물이 권리 범위에 포함되는 것으로 이해되어야 한다.

[0027] 실시예에서 사용한 용어는 단지 설명을 목적으로 사용된 것으로, 한정하려는 의도로 해석되어서는 안된다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 명세서 상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0028] 또한, 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일한 참조부호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다. 실시예를 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 실시예의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.

- [0029] 이하, 본 발명의 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.
- [0030] 도 1은 본 발명의 일실시예에 따른 소리 이벤트 검출 시스템을 도시한 도면이다.
- [0031] 본 발명의 일실시예에 따른 소리 이벤트 검출 시스템은 도 1에 도시된 바와 같이 소리 이벤트 검출 모델 학습 장치(110), 및 소리 이벤트 검출 장치(120)를 포함할 수 있다. 예를 들어, 소리 이벤트 검출 모델 학습 장치(110), 및 소리 이벤트 검출 장치(120)은 PC, 태블릿 단말, 스마트 단말, 또는 이동 단말과 같이 데이터를 입력 받는 입력 인터페이스, 입력 받은 데이터를 처리하는 프로세서 및 처리 결과를 출력하는 출력 인터페이스를 포함하는 장치일 수 있다.
- [0032] 소리 이벤트 검출 모델 학습 장치(110) 학습용 데이터베이스(101)로부터 학습 데이터를 획득할 수 있다. 예를 들어, 학습 데이터는 약전사(Weak label) 레이블, 강전사(Strong label) 레이블, 및 비전사(Un-label) 레이블 중 적어도 하나를 포함할 수 있다. 이때, 소리 이벤트 검출 모델 학습 장치(110)는 학습 데이터의 원본 특징들을 추출할 수 있다. 예를 들어, 원본 특징들은 64차원의 로그-멜-필터뱅크 값일 수 있다.
- [0033] 그리고, 소리 이벤트 검출 모델 학습 장치(110)는 원본 특징들을 기초로 차분 특징을 추출할 수 있다. 다음으로, 소리 이벤트 검출 모델 장치(110)는 원본 특징들 및 차분 특징을 기초로 입력 특징맵을 구성하여 소리 이벤트 검출 모델(112)을 학습할 수 있다. 또한, 소리 이벤트 검출 모델 장치(110)는 학습이 완료된 소리 이벤트 검출 모델(112)을 소리 이벤트 검출 장치(120)로 전송할 수 있다.
- [0034] 소리 이벤트 검출 장치(120)는 실제 오디오 신호를 수신하고, 수신한 실제 오디오 신호에서 실제 오디오 신호의 원본 특징들을 추출할 수 있다. 다음으로, 소리 이벤트 검출 장치(120)는 추출한 원본 특징들을 기초로 차분 특징을 추출하고, 원본 특징들 및 차분 특징을 기초로 입력 특징맵을 구성할 수 있다. 마지막으로 소리 이벤트 검출 장치(120)는 입력 특징맵을 소리 이벤트 검출 모델 학습 장치(110)로부터 수신한 소리 이벤트 검출 모델(112)에 입력하여 검출 결과(122)를 출력할 수 있다.
- [0035] 본 발명의 일실시예에 따른 소리 이벤트 검출 시스템은 오디오 신호의 원본 특징들과 원본 특징들로부터 추출한 차분 특징들을 입력 특징맵으로 구성하여 소리 이벤트 검출 모델에 사용함으로써, 소리 이벤트 검출 성능을 향상시킬 수 있다.
- [0036] 또한, 본 발명의 일실시예에 따른 소리 이벤트 검출 시스템은 분류 레이어의 출력을 클립(Clip) 전체 구간에 대하여 평균하여 출력하는 글로벌 평균 풀링 레이어(global average pooling layer)가 포함된 소리 이벤트 검출 모델을 사용함으로써, 강전사(strong label) 레이블만 아니라 약전사(weak label) 레이블 및 비전사(un-label) 레이블도 학습 데이터로 활용할 수 있다.
- [0037] 도 2는 본 발명의 일실시예에 따른 소리 이벤트 검출 모델 학습 방법을 도시한 플로우차트이다.
- [0038] 단계(210)에서 소리 이벤트 검출 모델 학습 장치(110)는 학습용 데이터베이스(111)로부터 학습 데이터를 수신할 수 있다. 예를 들어, 학습 데이터는 표 1에 도시된 바와 같이 약전사(Weak label) 레이블, 강전사(Strong label) 레이블, 및 비전사(Un-label) 레이블을 포함하도록 구성될 수 있다.

표 1

Label type	Weak label	Strong label	Un-label
No. of clips	1578	2045	14412
Label properties	Clip-level	Frame-level	None
Clip length	10 sec.		
Classes(10)	Speech, Dog, Cat, Alarm bell ring, Dishes, Frying, Blender, Running water, Vacuum cleaner, Electric shaver toothbrush		

- [0039]
- [0040] 단계(220)에서 소리 이벤트 검출 모델 학습 장치(110)는 단계(210)에서 수신한 학습 데이터에서 학습 데이터의 원본 특징들을 추출할 수 있다. 이때, 소리 이벤트 검출 모델 학습 장치(110)는 학습 데이터를 샘플링하여 STFT(Short-Time Fourier Transform)를 계산할 수 있다. 다음으로, 소리 이벤트 검출 모델 학습 장치(110)는 학습 데이터의 STFT값으로부터 멜-필터뱅크 값을 결정할 수 있다. 그 다음으로, 소리 이벤트 검출 모델 학

습 장치(110)는 멜-필터뱅크 값을 로그 변환한 로그-멜-필터뱅크 값을 원본 특징들로 추출할 수 있다.

- [0041] 단계(230)에서 소리 이벤트 검출 모델 학습 장치(110)는 단계(220)에서 추출한 원본 특징들을 기초로 차분 특징을 추출할 수 있다. 이때, 소리 이벤트 검출 모델 학습 장치(110)는 학습 데이터의 프레임들 각각에서 추출한 원본 특징들 간의 차이를 나타내는 1차 차분 특징들을 생성할 수 있다. 다음으로, 소리 이벤트 검출 모델 학습 장치(110)는 1차 차분 특징들 간의 차이를 나타내는 2차 차분 특징들을 생성할 수 있다.
- [0042] 단계(240)에서 소리 이벤트 검출 모델 학습 장치(110)는 단계(220)에서 추출한 원본 특징들 및 단계(230)에서 추출한 차분 특징을 기초로 입력 특징맵을 구성할 수 있다. 이때, 입력 특징맵은 원본 특징들, 1차 차분 특징들 및 2차 차분 특징들을 포함할 수 있다.
- [0043] 단계(250)에서 소리 이벤트 검출 모델 학습 장치(110)는 입력 특징맵을 이용하여 소리 이벤트 검출 모델을 학습할 수 있다.
- [0044] 이때, 소리 이벤트 검출 모델은 입력 특징맵의 주파수 영역의 정보를 감소시키는 복수의 컨볼루션 블록들; 쌍방향 GRU(Gated Recurrent Unit); 학습 데이터의 오디오 클래스를 분류하는 분류 레이어(Classification layer); 및 분류 레이어의 출력을 클립(Clip) 전체 구간에 대하여 평균하여 출력하는 글로벌 평균 풀링 레이어(global average pooling layer)를 포함하는 CRNN (Convolutional Recurrent Neural Network)일 수 있다. 그리고, 소리 이벤트 검출 모델은 학습 데이터가 강전사 레이블인 경우, 분류 레이어의 출력을 소리 이벤트 검출 결과로 출력하며, 학습 데이터가 약전사 레이블인 경우, 분류 레이어의 출력을 입력받은 글로벌 평균 풀링 레이어의 출력을 소리 이벤트 검출 결과로 출력할 수 있다.
- [0045] 그리고, 소리 이벤트 검출 모델은 복수의 CNN 블록을 이용하여 입력 특징맵에 대한 차원 축소를 수행한 후, 트랜스포머 인코더를 적용하는 베이스라인 트랜스포머; 컨볼루션 블록과 결합된 트랜스포머 인코더들을 이용하여 학습 데이터의 멀티 스케일 특징을 추출하는 멀티-스케일 추출 모듈; 및 멀티-스케일 추출 모듈의 출력을 이용하여 학습 데이터의 오디오 클래스를 분류하는 분류 레이어를 포함하는 트랜스포머 모델일 수도 있다. 이때, 분류 레이어는, 학습 데이터의 프레임 레벨 데이터와 클립 레벨 데이터로부터 각각 프레임 레벨 예측과 클립 레벨 예측을 수행하여 학습 데이터의 오디오 클래스를 분류할 수 있다.
- [0046] 단계(260)에서 소리 이벤트 검출 모델 학습 장치(110)는 단계(250)에서 학습된 소리 이벤트 검출 모델을 소리 이벤트 검출 장치(120)로 전송할 수 있다.
- [0047] 도 3은 본 발명의 일실시예에 따른 원본 특징 추출 과정을 도시한 도면이다.
- [0048] 단계(310)에서 소리 이벤트 검출 모델 학습 장치(110)는 학습 데이터에 포함된 오디오 웨이브(wave)를 하밍 윈도우(Hamming windowing)을 이용하여 16 kHz의 비율로 샘플링하고, 41.5ms의 프레임 간격을 설정할 수 있다.
- [0049] 단계(320)에서 소리 이벤트 검출 모델 학습 장치(110)는 단계(310)에서 설정된 프레임 간격마다 STFT(Short-Time Fourier Transform)을 수행할 수 있다.
- [0050] 단계(330)에서 소리 이벤트 검출 모델 학습 장치(110)는 단계(320)에서 계산된 STFT 값으로부터 전체 0에서 8,000Hz의 주파수 구간에서 64 band의 멜-필터뱅크 값을 구할 수 있다.
- [0051] 단계(340)에서 소리 이벤트 검출 모델 학습 장치(110)는 단계(330)에서 결정된 멜-필터뱅크 값을 로그(LOG) 변환 함으로써 64차원의 로그-멜-필터뱅크 값을 학습 데이터에 포함된 오디오 웨이브의 원본 특징들로 추출할 수 있다.
- [0052] 또한, CRNN 학습시, 소리 이벤트 검출 모델 학습 장치(110)는 로그-멜-필터뱅크 값을 정규화 한 후, CRNN에 사용할 수 있다. 이때, 소리 이벤트 검출 모델 학습 장치(110)는 로그-멜-필터뱅크 값에서 학습 데이터 전체의 평균값을 감산한 후, 표준편차 값으로 나누는 동작을 수행하여 로그-멜-필터뱅크 값을 정규화 할 수 있다.
- [0053] 도 4는 본 발명의 일실시예에 따른 차분 특징 추출 과정을 도시한 도면이다.
- [0054] 소리 이벤트 검출 모델 학습 장치(110)는 도 3에 기재된 동작에 따라 학습 데이터에 포함된 오디오 웨이브의 제1 프레임(411), 제2 프레임(412), 제3 프레임(413), 내지 제n 프레임(414)로부터 각각 제1 프레임의 원본 특징(421), 제2 프레임의 원본 특징(422), 제3 프레임의 원본 특징(423) 내지 제n 프레임의 원본 특징(424)을 추출할 수 있다.
- [0055] 다음으로, 소리 이벤트 검출 모델 학습 장치(110)는 제1 프레임의 원본 특징(421)과 제2 프레임의 원본 특징

(422) 간의 차이를 나타내는 1차 차분 특징 1(431), 제2 프레임의 원본 특징(422)과 제3 프레임의 원본 특징(423) 간의 차이를 나타내는 1차 차분 특징 2(432) 및 그 외의 1차 차분 특징들을 추출할 수 있다.

[0056] 그 다음으로, 소리 이벤트 검출 모델 학습 장치(110)는 1차 차분 특징 1(431)과, 1차 차분 특징 2(432) 간의 차이를 나타내는 2차 차분 특징 2(441) 및 그 외의 2차 차분 특징들을 추출할 수 있다.

[0057] 예를 들어, 소리 이벤트 검출 모델 학습 장치(110)는 수학식 1을 사용하여 1차 차분 특징들 및 2차 차분 특징들을 추출할 수 있다.

수학식 1

$$d_t = \frac{\sum_{n=1}^N n(c_{t+n} - c_{t-n})}{2 \sum_{n=1}^N n^2}$$

[0058]

[0059] 이때, d_t 는 추출하는 차분 특징이고, c_t 는 t 번째 프레임에서의 정적 특징이며, N 은 차분특징을 계산할 때 사용되는 앞뒤 프레임의 개수이다. 예를 들어, 도 4의 실시예에서 N 은 2일 수 있다. 또한, N 이 3인 경우, 소리 이벤트 검출 모델 학습 장치(110)는 3개의 프레임들 각각에서 추출한 특징들 간의 차분 특징을 추출할 수 있다.

[0060] 도 5는 본 발명의 일실시예에 따른 입력 특징맵의 일례이다.

[0061] 소리 이벤트 검출 모델 학습 장치(110)는 도 5에 도시된 바와 같이 원본 특징들(510), 1차 차분 특징들(520) 및 2차 차분 특징들(530)로 입력 특징맵(500)을 구성할 수 있다. 그리고, 소리 이벤트 검출 모델 학습 장치(110)는 구성된 입력 특징맵(500)을 소리 이벤트 검출 모델(501)에 입력하여 학습시킬 수 있다. 이때, 소리 이벤트 검출 모델(501)은 학습되기 전의 소리 이벤트 검출 모델이므로, 학습이 완료된 소리 이벤트 검출 모델(112)와는 차이가 있을 수 있다.

[0062] 도 6은 본 발명의 일실시예에 따른 소리 이벤트 검출 모델의 일례이다.

[0063] 소리 이벤트 검출 모델이 CRNN인 경우, 소리 이벤트 검출 모델은 입력 특징맵의 주파수 영역의 정보를 감소시키는 복수의 컨볼루션 블록(Conv. block)들(610, 620, 630); 쌍방향(bidirectional) GRU(640); 학습 데이터의 오디오 클래스를 분류하는 분류 레이어(650); 및 분류 레이어의 출력을 클립 전체 구간에 대하여 평균하여 출력하는 글로벌 평균 풀링 레이어(global average pooling layer)(660)를 포함할 수 있다.

[0064] 이때, CRNN은 학습 데이터가 강전사 레이블인 경우, 분류 레이어의 출력을 소리 이벤트 검출 결과로 출력하며, 학습 데이터가 약전사 레이블인 경우, 분류 레이어의 출력을 입력받은 글로벌 평균 풀링 레이어의 출력을 소리 이벤트 검출 결과로 출력할 수 있다.

[0065] 구체적으로, 64차원의 로그-멜 필터뱅크 값과 그들의 1차 차분 특징 및 2차 차분 특징으로 구성된 입력 특징맵(500)이 CRNN에 입력될 수 있다. 예를 들어, CRNN으로 입력되는 모든 오디오 클립의 길이가 10초이며 41.6ms의 프레임 간격마다 로그-멜 필터뱅크 값이 생성된 경우, 입력 특징맵(500)은 각각 독립된 240 프레임 길이의 특징맵으로 구성될 수 있다.

[0066] 컨볼루션 블록(610, 620, 630)들은 각각 64개의 특징맵들을 입력받을 수 있다. 그리고, 컨볼루션 블록(610, 620, 630)들은 각 특징맵마다, 3×3 의 필터링 후 배치 정규화(batch normalization)를 적용할 수 있다. 다음으로, 컨볼루션 블록(610, 620, 630)들은 배치 정규화를 적용한 특징맵들에게 ReLU(Rectified Linear Unit) 활성화 함수를 사용한 후, 드롭 아웃(dropout)을 적용하여 출력할 수 있다. 또한, 컨볼루션 블록(610, 620, 630)들은 드롭 아웃이 적용된 후, 1×4 의 맥스 풀링(max pooling)을 적용하여 주파수 영역에서의 정보의 양은 감소시키며 시간 영역에서의 정보는 보전되도록 함으로써, 오디오 웨이브의 시간 관계 정보를 후단의 RNN에서 모델링 가능하도록 할 수 있다.

- [0067] 쌍방향(bidirectional) GRU(640)는 64개의 유니트(unit)를 포함하며, 컨볼루션 블록(630)의 출력을 매 프레임마다, 10개의 유니트(unit)로 구성된 분류 레이어(650)의 입력에 대응되도록 변환하여 출력할 수 있다.
- [0068] 분류 레이어(650)의 10개의 유니트(unit)은 각각 분류하고자 하는 특정 오디오 클래스를 의미하며, 시그모이드(sigmoid) 활성화함수를 포함할 수 있다. 그리고, 분류 레이어(650)의 유니트(unit) 각각의 출력값은 10개 클래스별로의 매 프레임 별 사후확률(Posterior probability) 값을 나타낼 수 있다.
- [0069] 또한, 학습 데이터가 약전사 레이블을 가진 경우, 분류 레이어(650)층의 출력들은 글로벌 평균 풀링 레이어(660)에 입력되고, 글로벌 평균 풀링 레이어(660)는 오디오 클립 전체 길이(10초) 동안의 클래스 별 사후 확률의 평균값을 출력할 수 있다.
- [0070] 도 7은 본 발명의 일실시예에 따른 소리 이벤트 검출 모델의 다른 일례이다.
- [0071] 소리 이벤트 검출 모델이 트랜스포머 모델인 경우, 소리 이벤트 검출 모델은 복수의 CNN 블록을 이용하여 입력 특징맵에 대한 차원 축소를 수행한 후, 트랜스포머 인코더를 적용하는 베이스라인 트랜스포머(730); 컨볼루션 블록과 결합된 트랜스포머 인코더들을 이용하여 학습 데이터의 멀티 스케일 특징을 추출하는 멀티-스케일 추출 모듈(740); 및 멀티-스케일 추출 모듈의 출력을 이용하여 학습 데이터의 오디오 클래스를 분류하는 분류 레이어(750)를 포함할 수 있다.
- [0072] 베이스라인 트랜스포머(730)는 7개의 컨볼루션 블록들과 하나의 트랜스포머 인코더로 구성될 수 있다.
- [0073] 베이스라인 트랜스포머(730)는 학습 데이터의 오디오 웨이브(710)에 기초하여 생성된 입력 특징맵(720)을 입력 받을 수 있다. 예를 들어 입력 특징맵(720)은 도 5의 입력 특징맵(500)과 동일한 정보일 수 있다.
- [0074] 또한, 컨볼루션 블록에서는 3x3의 컨볼루션 필터가 context window로 적용되며, 출력의 평균을 0, 분산을 1로 정규화 하기 위해 배치 정규화가 사용될 수 있다. 배치 정규화 이후 GLU(Gated Linear Unit)가 활성화 함수로 적용되고, 훈련 중 과적합을 줄이기 위해 드랍 아웃이 적용된다. 이때, 마지막 컨볼루션 블록에서의 입력 데이터를 $128 \times 62 \times 1$ (특징맵 수 $\times$ 시간 프레임 수 $\times$ 주파수 대역 수)로 줄이기 위해 컨볼루션 블록들은 주파수와 시간 영역에 각각 $\{2 \times 2, 1 \times 2, 1 \times 1\}$ 의 평균 풀링(Average pooling)을 선택적으로 적용할 수 있다. 그에 따라 주파수 영역은 1프레임으로 축소되었지만, 시간 영역은 트랜스포머 인코더에서 소리 신호의 시간 상관 정보를 사용하기 위해 62프레임으로 보존될 수 있다.
- [0075] 컨볼루션 블록들 중 마지막 컨볼루션 블록의 출력은 트랜스포머 인코더의 입력으로 사용될 수 있다. 이때, 트랜스포머 인코더는 128의 어텐션 유닛(=128), 16개의 병렬헤드(=16), 512개의 내부 position-wise FFN의 심층망(=512)이 하이퍼 파라미터로 설정된 인코더일 수 있다.
- [0076] 멀티-스케일 추출 모듈(740)은 도 7에 도시된 바와 같이 각각 1개의 컨볼루션 블록과 결합된 트랜스포머 인코더 2개를 포함할 수 있다. 멀티-스케일 추출 모듈(740)은 첫 번째 트랜스포머 인코더의 출력을 업샘플링(Upsampling) 한 후, 두 번째 트랜스포머 인코더의 출력과 연결한 연결 특징맵을 생성할 수 있다. 다음으로, 멀티-스케일 추출 모듈(740)은 연결 특징맵에 1x1 컨볼루션을 적용하여 차원을 감소시킬 수 있다. 그 다음으로, 멀티-스케일 추출 모듈(740)은 차원이 감소된 연결 특징맵을 업샘플링 한 후, 베이스라인 트랜스포머(730)의 트랜스포머 인코더의 출력과 연결할 수 있다. 다음으로, 멀티-스케일 추출 모듈(740)은 베이스라인 트랜스포머(730)의 트랜스포머 인코더의 출력과 연결된 특징맵에 1x1 컨볼루션을 적용하여 멀티-스케일 특징을 출력할 수 있다.
- [0077] 분류 레이어(750)의 10개의 유니트(unit)은 각각 분류하고자 하는 특정 오디오 클래스를 의미하며, 시그모이드(sigmoid) 활성화함수를 포함할 수 있다. 그리고, 분류 레이어(650)의 유니트(unit) 각각의 출력값은 10개 클래스별로의 매 프레임 별 사후확률(Posterior probability) 값을 나타낼 수 있다.
- [0078] 또한, 분류 레이어(750)는 학습 데이터의 프레임 레벨 데이터와 클립 레벨 데이터로부터 각각 프레임 레벨 예측과 클립 레벨 예측을 수행하여 학습 데이터의 오디오 클래스를 분류할 수 있다. 이때, 분류 레이어(750)는 softmax 활성화 함수를 가지는 어텐션(attention) 레이어들로 각각 프레임 레벨 예측과 클립 레벨 예측을 수행하고, 레이어들의 출력 값을 결합하여 가중 평균을 계산하는 것으로 오디오 클립 전체 길이(10초) 동안의 클래스별 사후 확률 값을 결정할 수 있다.
- [0079] 도 8은 본 발명의 일실시예에 따른 소리 이벤트 검출 방법을 도시한 플로우차트이다.
- [0080] 단계(810)에서 소리 이벤트 검출 장치(120)는 소리 이벤트 검출 모델 학습 장치(110)로부터 소리 이벤트 검출

모델(112)를 수신할 수 있다.

- [0081] 단계(820)에서 소리 이벤트 검출 장치(120)는 소리 이벤트를 검출할 실제 오디오 신호를 수신할 수 있다.
- [0082] 단계(830)에서 소리 이벤트 검출 장치(120)는 오디오 신호에서 오디오 신호의 원본 특징들을 추출할 수 있다. 이때, 소리 이벤트 검출 장치(120)는 오디오 신호를 샘플링하여 STFT를 계산할 수 있다. 다음으로, 소리 이벤트 검출 장치(120)는 오디오 신호의 STFT값으로부터 멜-필터뱅크 값을 결정할 수 있다. 그 다음으로, 소리 이벤트 검출 장치(120)는 멜-필터뱅크 값을 로그 변환한 로그-멜-필터뱅크 값을 원본 특징들로 추출할 수 있다.
- [0083] 단계(840)에서 소리 이벤트 검출 장치(120)는 원본 특징들을 기초로 차분 특징을 추출할 수 있다. 이때, 소리 이벤트 검출 장치(120)는 오디오 신호의 프레임들 각각에서 추출한 원본 특징들 간의 차이를 나타내는 1차 차분 특징을 생성할 수 있다. 그리고, 소리 이벤트 검출 장치(120)는 1차 차분 특징들 간의 차이를 나타내는 2차 차분 특징을 생성할 수 있다.
- [0084] 단계(850)에서 소리 이벤트 검출 장치(120)는 원본 특징들 및 차분 특징을 기초로 입력 특징맵을 구성할 수 있다. 이때, 소리 이벤트 검출 장치(120)는 학습 데이터의 원본 특징, 1차 차분 특징 및 2차 차분 특징으로 입력 특징맵을 구성할 수 있다.
- [0085] 단계(860)에서 소리 이벤트 검출 장치(120)는 입력 특징맵을 소리 이벤트 검출 모델(112)에 입력하여 검출 결과를 출력할 수 있다.
- [0086] 본 발명은 오디오 신호의 원본 특징들과 원본 특징들로부터 추출한 차분 특징들을 입력 특징맵으로 구성하여 소리 이벤트 검출 모델에 사용함으로써, 소리 이벤트 검출 성능을 향상시킬 수 있다.
- [0087] 또한, 본 발명은 분류 레이어의 출력을 클립(Clip) 전체 구간에 대하여 평균하여 출력하는 글로벌 평균 풀링 레이어가 포함된 소리 이벤트 검출 모델을 사용함으로써, 강전사(strong label) 레이블만 아니라 약전사(weak label) 레이블 및 비전사(un-label) 레이블도 학습 데이터로 활용할 수 있다.
- [0089] 한편, 본 발명에 따른 소리 이벤트 검출 모델에 기초한 소리 이벤트 검출 장치 또는 소리 이벤트 검출 모델에 기초한 소리 이벤트 검출 방법은 컴퓨터에서 실행될 수 있는 프로그램으로 작성되어 마그네틱 저장매체, 광학적 판독매체, 디지털 저장매체 등 다양한 기록 매체로도 구현될 수 있다.
- [0090] 본 명세서에 설명된 각종 기술들의 구현들은 디지털 전자 회로조직으로, 또는 컴퓨터 하드웨어, 펌웨어, 소프트웨어로, 또는 그들의 조합들로 구현될 수 있다. 구현들은 데이터 처리 장치, 예를 들어 프로그램가능 프로세서, 컴퓨터, 또는 다수의 컴퓨터들의 동작에 의한 처리를 위해, 또는 이 동작을 제어하기 위해, 컴퓨터 프로그램 제품, 예를 들어 기계 판독가능 저장 장치(컴퓨터 판독가능 매체)에서 유형적으로 구체화된 컴퓨터 프로그램으로서 구현될 수 있다. 상술한 컴퓨터 프로그램(들)과 같은 컴퓨터 프로그램은 컴파일된 또는 인터프리트된 언어들을 포함하는 임의의 형태의 프로그래밍 언어로 기록될 수 있고, 독립형 프로그램으로서 또는 모듈, 구성요소, 서브루틴, 또는 컴퓨팅 환경에서의 사용에 적절한 다른 유닛으로서 포함하는 임의의 형태로 전개될 수 있다. 컴퓨터 프로그램은 하나의 사이트에서 하나의 컴퓨터 또는 다수의 컴퓨터들 상에서 처리되도록 또는 다수의 사이트들에 걸쳐 분배되고 통신 네트워크에 의해 상호 연결되도록 전개될 수 있다.
- [0091] 컴퓨터 프로그램의 처리에 적절한 프로세서들은 예로서, 범용 및 특수 목적 마이크로프로세서들 둘 다, 및 임의의 종류의 디지털 컴퓨터의 임의의 하나 이상의 프로세서들을 포함한다. 일반적으로, 프로세서는 판독 전용 메모리 또는 랜덤 액세스 메모리 또는 둘 다로부터 명령어들 및 데이터를 수신할 것이다. 컴퓨터의 요소들은 명령어들을 실행하는 적어도 하나의 프로세서 및 명령어들 및 데이터를 저장하는 하나 이상의 메모리 장치들을 포함할 수 있다. 일반적으로, 컴퓨터는 데이터를 저장하는 하나 이상의 대량 저장 장치들, 예를 들어 자기, 자기-광 디스크들, 또는 광 디스크들을 포함할 수 있거나, 이것들로부터 데이터를 수신하거나 이것들에 데이터를 송신하거나 또는 양쪽으로 되도록 결합될 수도 있다. 컴퓨터 프로그램 명령어들 및 데이터를 구체화하는데 적절한 정보 캐리어들은 예로서 반도체 메모리 장치들, 예를 들어, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(Magnetic Media), CD-ROM(Compact Disk Read Only Memory), DVD(Digital Video Disk)와 같은 광 기록 매체(Optical Media), 플롭티컬 디스크(Floptical Disk)와 같은 자기-광 매체(Magneto-Optical Media), 롬(ROM, Read Only Memory), 램(RAM, Random Access Memory), 플래시 메모리, EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM) 등을 포함한다. 프로세서 및 메모리는 특수 목적 논리 회로조직에 의해 보충되거나, 이에 포함될 수 있다.

[0092] 또한, 컴퓨터 판독가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용매체일 수 있고, 컴퓨터 저장매체를 모두 포함할 수 있다.

[0093] 본 명세서는 다수의 특정한 구현물의 세부사항들을 포함하지만, 이들은 어떠한 발명이나 청구 가능한 것의 범위에 대해서도 제한적인 것으로서 이해되어서는 안되며, 오히려 특정한 발명의 특정한 실시형태에 특유할 수 있는 특징들에 대한 설명으로서 이해되어야 한다. 개별적인 실시형태의 문맥에서 본 명세서에 기술된 특정한 특징들은 단일 실시형태에서 조합하여 구현될 수도 있다. 반대로, 단일 실시형태의 문맥에서 기술한 다양한 특징들 역시 개별적으로 혹은 어떠한 적절한 하위 조합으로도 복수의 실시형태에서 구현 가능하다. 나아가, 특징들이 특정한 조합으로 동작하고 초기에 그와 같이 청구된 바와 같이 묘사될 수 있지만, 청구된 조합으로부터의 하나 이상의 특징들은 일부 경우에 그 조합으로부터 배제될 수 있으며, 그 청구된 조합은 하위 조합이나 하위 조합의 변형물로 변경될 수 있다.

[0094] 마찬가지로, 특정한 순서로 도면에서 동작들을 묘사하고 있지만, 이는 바람직한 결과를 얻기 위하여 도시된 그 특정한 순서나 순차적인 순서대로 그러한 동작들을 수행하여야 한다거나 모든 도시된 동작들이 수행되어야 하는 것으로 이해되어서는 안 된다. 특정한 경우, 멀티태스킹과 병렬 프로세싱이 유리할 수 있다. 또한, 상술한 실시형태의 다양한 장치 컴포넌트의 분리는 그러한 분리를 모든 실시형태에서 요구하는 것으로 이해되어서는 안되며, 설명한 프로그램 컴포넌트와 장치들은 일반적으로 단일의 소프트웨어 제품으로 함께 통합되거나 다중 소프트웨어 제품에 패키징 될 수 있다는 점을 이해하여야 한다.

[0095] 한편, 본 명세서와 도면에 개시된 본 발명의 실시 예들은 이해를 돕기 위해 특정 예를 제시한 것에 지나지 않으며, 본 발명의 범위를 한정하고자 하는 것은 아니다. 여기에 개시된 실시 예들 이외에도 본 발명의 기술적 사상에 바탕을 둔 다른 변형 예들이 실시 가능하다는 것은, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 자명한 것이다.

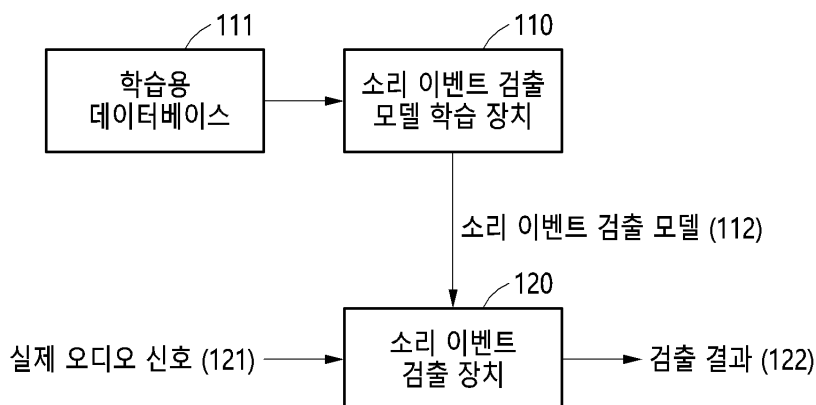
부호의 설명

[0096] 110: 소리 이벤트 검출 모델 학습 장치

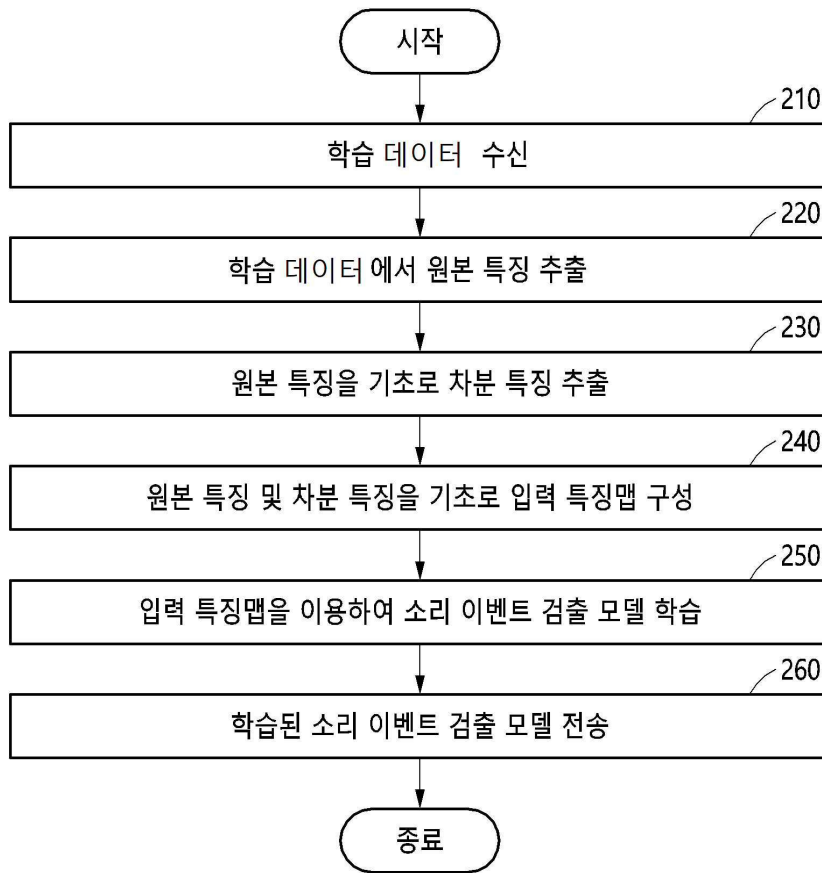
120: 소리 이벤트 검출 장치

도면

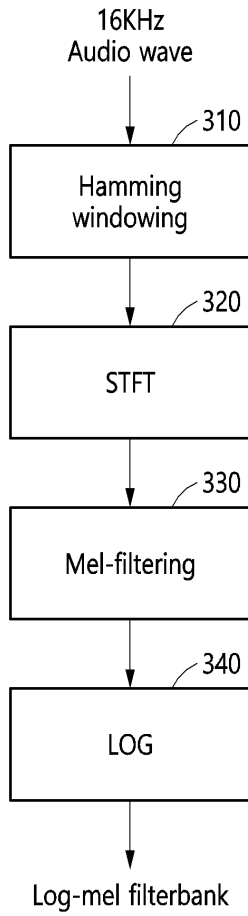
도면1



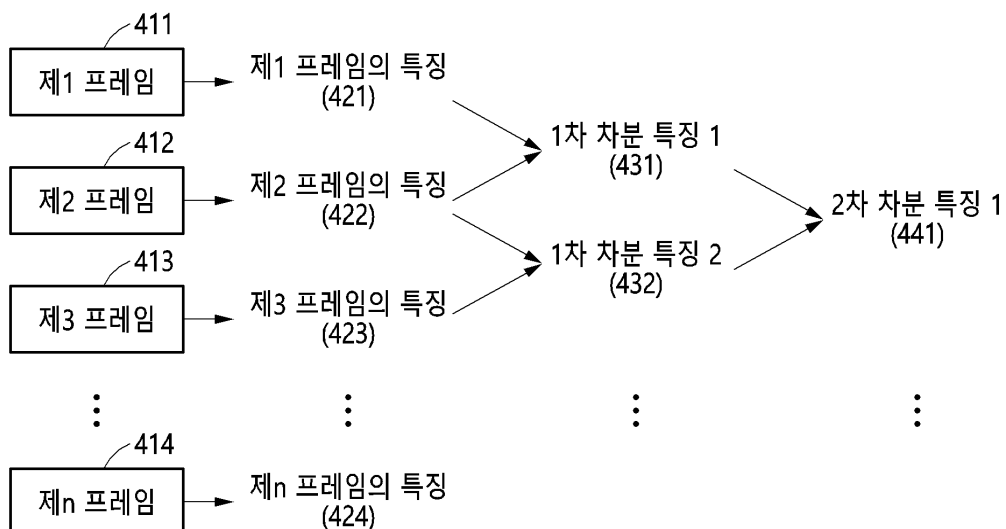
도면2



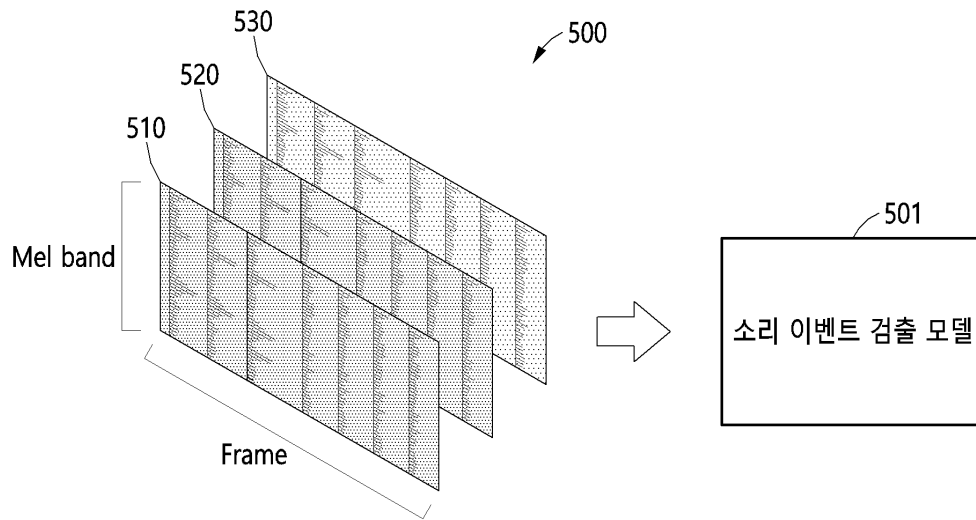
도면3



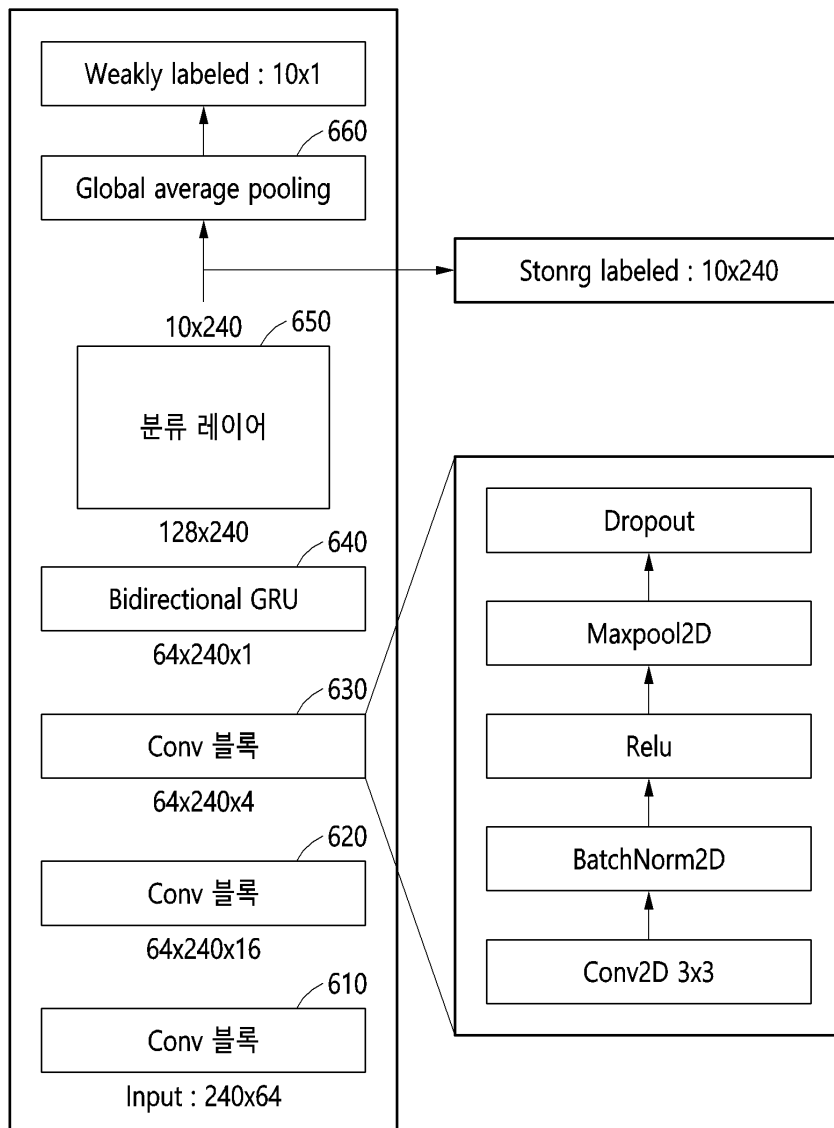
도면4



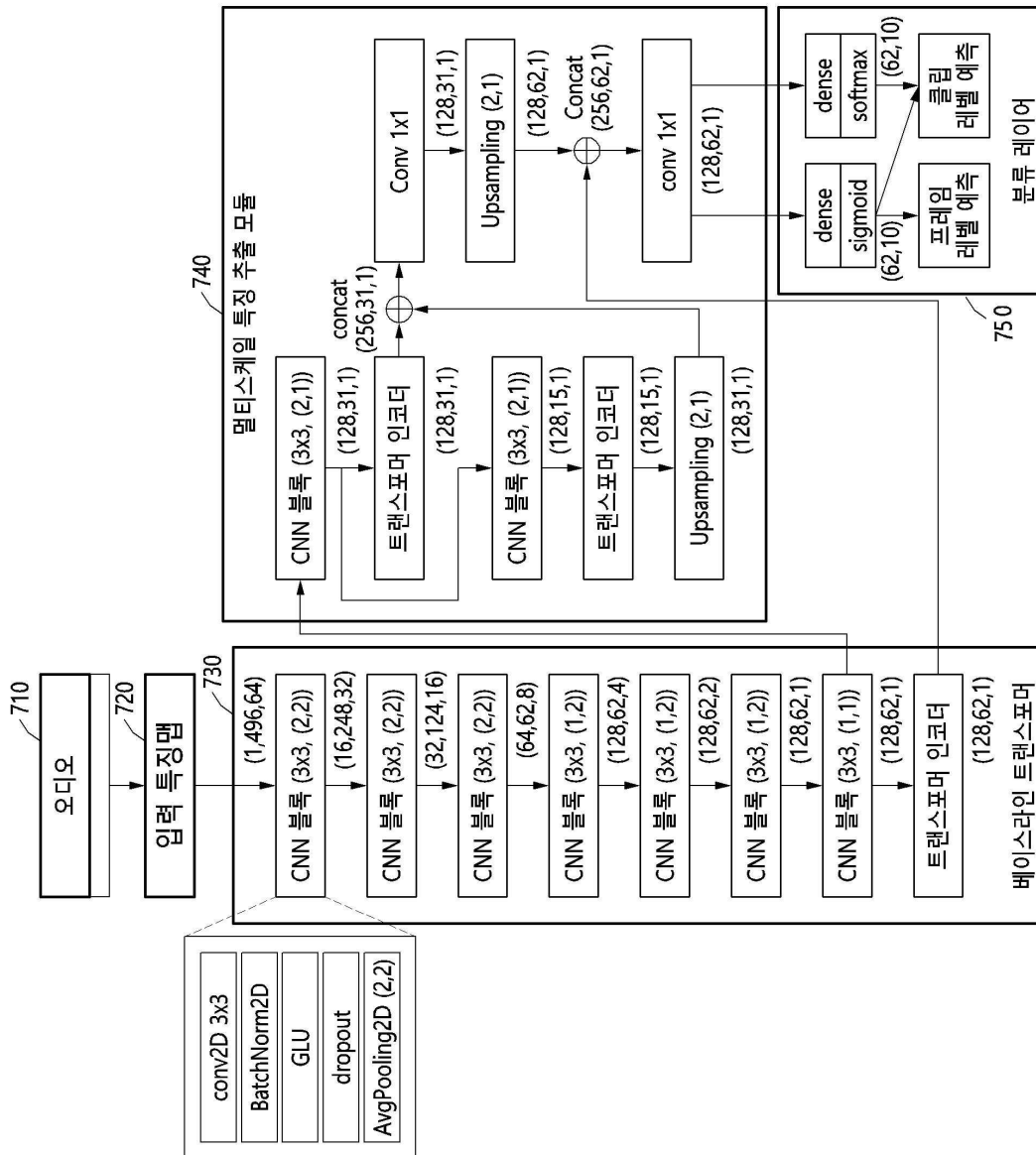
도면5



도면6



도면7



도면8

