



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0093826
(43) 공개일자 2023년06월27일

- | | |
|--|---|
| <p>(51) 국제특허분류(Int. Cl.)
 <i>G06V 40/10</i> (2022.01) <i>G06N 3/04</i> (2023.01)
 <i>G06T 7/11</i> (2017.01) <i>G06T 7/70</i> (2017.01)
 <i>G06V 10/764</i> (2022.01) <i>G06V 10/82</i> (2022.01)
 <i>G06V 20/70</i> (2022.01) <i>G10L 25/18</i> (2013.01)</p> <p>(52) CPC특허분류
 <i>G06V 40/10</i> (2022.01)
 <i>G06N 3/045</i> (2023.01)</p> <p>(21) 출원번호 10-2021-0182748
 (22) 출원일자 2021년12월20일
 심사청구일자 2021년12월20일</p> | <p>(71) 출원인
 세종대학교산학협력단
 서울특별시 광진구 능동로 209 (군자동, 세종대학교)</p> <p>(72) 발명자
 백성욱
 서울특별시 강남구 영동대로 22, 813동 1303호(일원동, 디에이치 자이 개포)</p> <p>이미영
 서울특별시 강남구 학동로82길 6(삼성동, 삼성리치빌)
 (뒷면에 계속)</p> <p>(74) 대리인
 특허법인위더피플</p> |
|--|---|

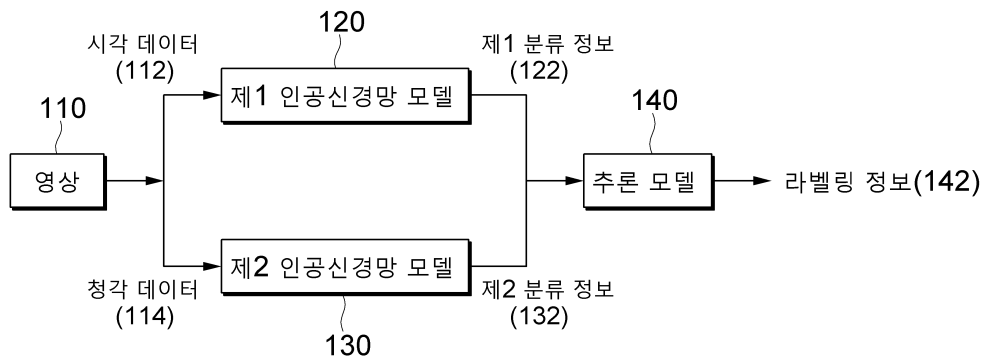
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법 및 장치

(57) 요약

본 발명은 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법에 관한 것이다. 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법은, 영상 데이터를 수신하는 단계, 수신된 영상 데이터로부터 동물과 연관된 시각 데이터 및 청각 데이터를 추출하는 단계, 추출된 시각 데이터를 학습된 제1 인공신경망 모델에 제공하여 동물의 이미지에 대응하는 제1 분류 정보를 생성하는 단계, 추출된 청각 데이터를 학습된 제2 인공신경망 모델에 제공하여 동물의 음성에 대응하는 제2 분류 정보를 생성하는 단계 및 생성된 제1 분류 정보 및 제2 분류 정보를 이용하여 영상 데이터에 포함된 동물에 대한 라벨링을 수행하는 단계를 포함한다.

대표도 - 도1



(52) CPC특허분류

G06T 7/11 (2017.01)
G06T 7/70 (2017.01)
G06V 10/764 (2023.01)
G06V 10/82 (2022.01)
G06V 20/70 (2023.01)
G10L 25/18 (2013.01)

(72) 발명자

이자즈 올 하크

서울특별시 광진구 능동로 209 세종대학교 대양AI
 센터 614호

히크마트 야르

서울특별시 광진구 능동로 209 세종대학교 대양AI
 센터 411호

우미르 하룬

서울특별시 광진구 능동로 209 세종대학교 대양AI
 센터 411호

아바스 칸

서울특별시 광진구 능동로 209 세종대학교 대양AI
 센터 614호

김민제

서울특별시 광진구 천호대로110길 83-5, 101호(능
 동, 예코빌)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711126258
과제번호	2020-0-00062-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	SW컴퓨팅산업원천기술개발(R&D, 정보화)
연구과제명	이중 정보 활용 및 데이터 융합을 통한 데이터 증식 기술 개발
기 여 율	1/1
과제수행기관명	(주)와이즈넷
연구기간	2021.01.01 ~ 2021.12.31

명세서

청구범위

청구항 1

적어도 하나의 프로세서에 의해 수행되는 동물 탐지 및 분류를 위한 영상 데이터 라벨링(labeling) 방법으로서, 영상 데이터(video data)를 수신하는 단계;

상기 수신된 영상 데이터로부터 동물과 연관된 시각 데이터(visual data) 및 청각 데이터(aural data)를 추출하는 단계;

상기 추출된 시각 데이터를 학습된 제1 인공신경망 모델에 제공하여 동물의 이미지(image)에 대응하는 제1 분류 정보를 생성하는 단계;

상기 추출된 청각 데이터를 학습된 제2 인공신경망 모델에 제공하여 동물의 음성(sound)에 대응하는 제2 분류 정보를 생성하는 단계; 및

상기 생성된 제1 분류 정보 및 제2 분류 정보를 이용하여 상기 영상 데이터에 포함된 동물에 대한 라벨링을 수행하는 단계;

를 포함하는 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법.

청구항 2

제1항에 있어서,

상기 추출된 시각 데이터를 학습된 제1 인공신경망 모델에 제공하여 동물의 이미지에 대응하는 제1 분류 정보를 생성하는 단계는,

상기 시각 데이터에 포함된 복수의 객체(object)를 검출하는 단계;

상기 검출된 복수의 객체 중 동물에 대응하는 적어도 일부의 객체를 결정하는 단계; 및

상기 결정된 적어도 일부의 객체와 연관된 동물의 이미지에 대응하는 제1 분류 정보를 생성하는 단계;

를 포함하는 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법.

청구항 3

제2항에 있어서,

상기 제1 인공신경망 모델은 상기 시각 데이터에서 상기 동물에 대응하는 적어도 일부의 객체와 연관된 위치 및 영역을 검출하기 위한 YOLO(You Only Look Once) 모델을 포함하는 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법.

청구항 4

제1항에 있어서,

상기 추출된 청각 데이터를 학습된 제2 인공신경망 모델에 제공하여 동물의 음성에 대응하는 제2 분류 정보를 생성하는 단계는,

상기 청각 데이터에 포함된 복수의 유형의 음성 정보를 검출하는 단계;

상기 검출된 복수의 유형의 음성 정보 중 동물에 대응하는 적어도 일부의 음성 정보를 결정하는 단계; 및

상기 결정된 적어도 일부의 음성 정보와 연관된 동물의 음성에 대응하는 제2 분류 정보를 생성하는 단계;
를 포함하는 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법.

청구항 5

제4항에 있어서,

상기 제2 인공신경망 모델은 CNN(Convolution Neural Network) 모델을 포함하고,

상기 추출된 청각 데이터를 학습된 제2 인공신경망 모델에 제공하여 동물의 음성에 대응하는 제2 분류 정보를 생성하는 단계는,

상기 동물에 대응하는 적어도 일부의 음성 정보에 대한 전처리를 수행하여 MFCC(Mel Frequency Cepstral Coefficient) 특징 정보를 추출하는 단계; 및

상기 추출된 MFCC 특징 정보를 상기 CNN 모델에 제공하여 동물의 음성에 대응하는 상기 제2 분류 정보를 생성하는 단계;

를 더 포함하는 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법.

청구항 6

제5항에 있어서,

상기 동물에 대응하는 적어도 일부의 음성 정보에 대한 전처리를 수행하여 MFCC 특징 정보를 추출하는 단계는,

상기 동물에 대응하는 적어도 일부의 음성 정보를 사전 결정된 윈도우(windowing) 알고리즘을 이용하여 복수의 프레임(frame)으로 분할하는 단계;

를 포함하는 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법.

청구항 7

제6항에 있어서,

상기 동물에 대응하는 적어도 일부의 음성 정보에 대한 전처리를 수행하여 MFCC 특징 정보를 추출하는 단계는,

이산 푸리에 변환(Discrete Fourier Transformation)을 이용하여 상기 복수의 프레임을 규모 스펙트럼(magnitude spectrum)으로 변환하는 단계;

를 더 포함하는 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법.

청구항 8

제7항에 있어서,

상기 동물에 대응하는 적어도 일부의 음성 정보에 대한 전처리를 수행하여 MFCC 특징 정보를 추출하는 단계는,

사전 결정된 멜 스케일(Mel scale) 알고리즘을 이용하여 상기 규모 스펙트럼을 로그 멜 스펙트럼(log Mel spectrum)으로 변환하는 단계;

를 더 포함하는 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법.

청구항 9

제8항에 있어서,

상기 동물에 대응하는 적어도 일부의 음성 정보에 대한 전처리를 수행하여 MFCC 특징 정보를 추출하는 단계는,
사전 결정된 이산 코사인 변환(Discrete Cosine Transform) 알고리즘을 이용하여 상기 로그 멜 스펙트럼을 기초로 상기 MFCC 특징 정보를 생성하는 단계;

를 더 포함하는 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법.

청구항 10

제5항에 있어서,

상기 CNN 모델은 3개의 컨볼루션 블록(convolution block)을 포함하는 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법.

청구항 11

제1항 내지 제10항 중 어느 한 항에 따른 방법을 컴퓨터에서 실행하기 위해 컴퓨터 판독 가능한 기록 매체에 저장된 컴퓨터 프로그램.

청구항 12

컴퓨팅 장치로서,

통신 모듈;

메모리; 및

상기 메모리와 연결되고, 상기 메모리에 포함된 컴퓨터 판독 가능한 적어도 하나의 프로그램을 실행하도록 구성된 적어도 하나의 프로세서;

를 포함하고,

상기 적어도 하나의 프로그램은,

영상 데이터를 수신하고,

상기 수신된 영상 데이터로부터 동물과 연관된 시각 데이터 및 청각 데이터를 추출하고,

상기 추출된 시각 데이터를 학습된 제1 인공신경망 모델에 제공하여 동물의 이미지에 대응하는 제1 분류 정보를 생성하고,

상기 추출된 청각 데이터를 학습된 제2 인공신경망 모델에 제공하여 동물의 음성에 대응하는 제2 분류 정보를 생성하고,

상기 생성된 제1 분류 정보 및 제2 분류 정보를 이용하여 상기 영상 데이터에 포함된 동물에 대한 라벨링을 수행하기 위한 명령어들을 포함하는 컴퓨팅 장치.

발명의 설명

기술 분야

본 발명은 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법 및 장치에 관한 것으로, 구체적으로, TV 프로그램 등의 동물과 연관된 영상을 이용하여 동물 탐지 및 분류를 수행하는 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법 및 장치에 관한 것이다.

[0001]

배경 기술

- [0003] 기술이 발전함에 따라 생태학적 연구를 위한 영상 자료 수집이 비교적 간단해지고 있으며, 이에 따라 수많은 영상 자료가 수집되고 있다. 이와 같이 수집된 영상 자료는 야생동물 모니터링, 동물 행동 분석, 교육 영역 등에 다양하게 활용되고 있다.
- [0004] 한편, 영상 자료를 효과적으로 활용하기 위해 AI(Artificial Intelligence)를 이용한 영상 자료 분석이 요구될 수 있다. AI를 이용하기 위해, 각 영상 자료에 포함된 동물(animal)에 대한 라벨링(labeling)을 수행하는 것이 필요할 수 있다. 그러나, 사람이 수많은 영상 자료에 포함된 동물에 대한 라벨링을 직접 수행하기 어려운 문제가 있다. 또한, 종래의 임의의 장치를 이용하여 라벨링을 수행하는 경우, 유사한 동물의 생김새 등으로 인해 라벨링에 대한 정확성이 떨어지는 문제가 있다.

발명의 내용

해결하려는 과제

- [0006] 본 발명은 상기와 같은 문제점을 해결하기 위한 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법, 컴퓨터 판독 가능 매체에 저장된 컴퓨터 프로그램, 컴퓨터 프로그램이 저장된 컴퓨터 판독 가능 매체 및 장치(시스템)를 제공한다.

과제의 해결 수단

- [0008] 본 발명은 방법, 시스템(장치), 컴퓨터 판독 가능 매체에 저장된 컴퓨터 프로그램 또는 컴퓨터 프로그램이 저장된 컴퓨터 판독 가능 매체를 포함한 다양한 방식으로 구현될 수 있다.
- [0009] 본 발명의 일 실시예에 따르면, 적어도 하나의 프로세서에 의해 수행되는 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법은, 영상 데이터를 수신하는 단계, 수신된 영상 데이터로부터 동물과 연관된 시각 데이터 및 청각 데이터를 추출하는 단계, 추출된 시각 데이터를 학습된 제1 인공신경망 모델에 제공하여 동물의 이미지에 대응하는 제1 분류 정보를 생성하는 단계, 추출된 청각 데이터를 학습된 제2 인공신경망 모델에 제공하여 동물의 음성에 대응하는 제2 분류 정보를 생성하는 단계 및 생성된 제1 분류 정보 및 제2 분류 정보를 이용하여 영상 데이터에 포함된 동물에 대한 라벨링을 수행하는 단계를 포함한다.
- [0010] 본 발명의 일 실시예에 따르면, 추출된 시각 데이터를 학습된 제1 인공신경망 모델에 제공하여 동물의 이미지에 대응하는 제1 분류 정보를 생성하는 단계는, 시각 데이터에 포함된 복수의 객체를 검출하는 단계, 검출된 복수의 객체 중 동물에 대응하는 적어도 일부의 객체를 결정하는 단계 및 결정된 적어도 일부의 객체와 연관된 동물의 이미지에 대응하는 제1 분류 정보를 생성하는 단계를 포함한다.
- [0011] 본 발명의 일 실시예에 따르면, 제1 인공신경망 모델은 시각 데이터에서 동물에 대응하는 적어도 일부의 객체와 연관된 위치 및 영역을 검출하기 위한 YOLO 모델을 포함한다.
- [0012] 본 발명의 일 실시예에 따르면, 추출된 청각 데이터를 학습된 제2 인공신경망 모델에 제공하여 동물의 음성에 대응하는 제2 분류 정보를 생성하는 단계는, 청각 데이터에 포함된 복수의 유형의 음성 정보를 검출하는 단계, 검출된 복수의 유형의 음성 정보 중 동물에 대응하는 적어도 일부의 음성 정보를 결정하는 단계 및 결정된 적어도 일부의 음성 정보와 연관된 동물의 음성에 대응하는 제2 분류 정보를 생성하는 단계를 포함한다.
- [0013] 본 발명의 일 실시예에 따르면, 제2 인공신경망 모델은 CNN 모델을 포함한다. 추출된 청각 데이터를 학습된 제2 인공신경망 모델에 제공하여 동물의 음성에 대응하는 제2 분류 정보를 생성하는 단계는, 동물에 대응하는 적어도 일부의 음성 정보에 대한 전처리를 수행하여 MFCC 특징 정보를 추출하는 단계 및 추출된 MFCC 특징 정보를 CNN 모델에 제공하여 동물의 음성에 대응하는 제2 분류 정보를 생성하는 단계를 더 포함한다.
- [0014] 본 발명의 일 실시예에 따르면, 동물에 대응하는 적어도 일부의 음성 정보에 대한 전처리를 수행하여 MFCC 특징 정보를 추출하는 단계는, 동물에 대응하는 적어도 일부의 음성 정보를 사전 결정된 윈도우 알고리즘을 이용하여 복수의 프레임으로 분할하는 단계를 포함한다.

- [0015] 본 발명의 일 실시예에 따르면, 동물에 대응하는 적어도 일부의 음성 정보에 대한 전처리를 수행하여 MFCC 특징 정보를 추출하는 단계는, 이산 푸리에 변환을 이용하여 복수의 프레임을 규모 스펙트럼으로 변환하는 단계를 더 포함한다.
- [0016] 본 발명의 일 실시예에 따르면, 동물에 대응하는 적어도 일부의 음성 정보에 대한 전처리를 수행하여 MFCC 특징 정보를 추출하는 단계는, 사전 결정된 멜 스케일 알고리즘을 이용하여 규모 스펙트럼을 로그 멜 스펙트럼으로 변환하는 단계를 더 포함한다.
- [0017] 본 발명의 일 실시예에 따르면, 동물에 대응하는 적어도 일부의 음성 정보에 대한 전처리를 수행하여 MFCC 특징 정보를 추출하는 단계는, 사전 결정된 이산 코사인 변환 알고리즘을 이용하여 로그 멜 스펙트럼을 기초로 MFCC 특징 정보를 생성하는 단계를 더 포함한다.
- [0018] 본 발명의 일 실시예에 따르면, CNN 모델은 3개의 컨볼루션 블록을 포함한다.
- [0019] 본 발명의 일 실시예에 따른 상술된 방법을 컴퓨터에서 실행하기 위해 컴퓨터 판독 가능한 기록 매체에 저장된 컴퓨터 프로그램이 제공된다.
- [0020] 본 발명의 일 실시예에 따른 컴퓨팅 장치는, 통신 모듈, 메모리 및 메모리와 연결되고, 메모리에 포함된 컴퓨터 판독 가능한 적어도 하나의 프로그램을 실행하도록 구성된 적어도 하나의 프로세서를 포함한다. 적어도 하나의 프로그램은, 영상 데이터를 수신하고, 수신된 영상 데이터로부터 동물과 연관된 시각 데이터 및 청각 데이터를 추출하고, 추출된 시각 데이터를 학습된 제1 인공신경망 모델에 제공하여 동물의 이미지에 대응하는 제1 분류 정보를 생성하고, 추출된 청각 데이터를 학습된 제2 인공신경망 모델에 제공하여 동물의 음성에 대응하는 제2 분류 정보를 생성하고, 생성된 제1 분류 정보 및 제2 분류 정보를 이용하여 영상 데이터에 포함된 동물에 대한 라벨링을 수행하기 위한 명령어들을 포함한다.

발명의 효과

- [0022] 본 발명의 다양한 실시예에서 사람이 직접 영상 데이터에 대한 라벨링을 수행하지 않는 경우에도, 영상 데이터의 시각적 특징 및 청각적 특징을 모두 이용하여 높은 정확도로 동물에 대한 라벨링이 수행될 수 있다.
- [0023] 본 발명의 다양한 실시예에서 컴퓨팅 장치는 이미지 처리를 수행하여 라벨링을 수행하는 일반적인 방식에 비해 청각 데이터를 추가적으로 이용함으로써, 임의의 영상에 포함된 동물에 대한 높은 정확성을 갖는 라벨링을 수행할 수 있다.
- [0024] 본 발명의 다양한 실시예에서 컴퓨팅 장치는 온라인 상에서 검색 가능한 임의의 영상 데이터로부터 간단히 동물에 대한 라벨링을 수행할 수 있으며, 이에 따라 수많은 소스로부터 라벨링 데이터가 효율적으로 확보될 수 있다.
- [0025] 본 발명의 다양한 실시예에서 영상으로부터 추출된 시각 데이터를 YOLO V4 모델에 제공하는 경우, 높은 신뢰성을 갖는 분류 정보가 생성될 수 있다.
- [0026] 본 발명의 다양한 실시예에서 3개의 컨볼루션 블록을 포함하는 CNN 모델을 이용함으로써, 다른 임의의 개수의 컨볼루션 블록을 이용하는 경우보다 높은 성능으로 동물 클래스 분류가 수행될 수 있다.
- [0027] 본 발명의 효과는 이상에서 언급한 효과로 제한되지 않으며, 언급되지 않은 다른 효과들은 청구범위의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자("통상의 기술자"라 함)에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [0029] 본 발명의 실시예들은, 이하 설명하는 첨부 도면들을 참조하여 설명될 것이며, 여기서 유사한 참조 번호는 유사한 요소들을 나타내지만, 이에 한정되지는 않는다.

도 1은 본 발명의 일 실시예에 따른 동물과 연관된 라벨링 정보가 생성되는 예시를 나타내는 도면이다.

도 2는 본 발명의 일 실시예에 따른 컴퓨팅 장치의 내부 구성을 나타내는 기능적인 블록도이다.

도 3은 본 발명의 일 실시예에 따른 YOLO V4 모델의 세부적인 구성을 나타내는 예시적인 도면이다.

도 4는 본 발명의 일 실시예에 따른 청각 데이터와 연관된 분류 정보가 생성되는 과정을 나타내는 예시적인 흐름도이다.

도 5는 본 발명의 일 실시예에 따른 CNN 모델의 구조를 나타내는 예시적인 도면이다.

도 6은 본 발명의 일 실시예에 따른 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법의 예시를 나타내는 도면이다.

도 7은 본 발명의 일 실시예에 따른 컴퓨팅 장치의 내부 구성을 나타내는 블록도이다.

발명을 실시하기 위한 구체적인 내용

- [0030] 이하, 본 발명의 실시를 위한 구체적인 내용을 첨부된 도면을 참조하여 상세히 설명한다. 다만, 이하의 설명에서는 본 발명의 요지를 불필요하게 흐릴 우려가 있는 경우, 널리 알려진 기능이나 구성에 관한 구체적 설명은 생략하기로 한다.
- [0031] 첨부된 도면에서, 동일하거나 대응하는 구성요소에는 동일한 참조부호가 부여되어 있다. 또한, 이하의 실시예들의 설명에 있어서, 동일하거나 대응되는 구성요소를 중복하여 기술하는 것이 생략될 수 있다. 그러나, 구성요소에 관한 기술이 생략되어도, 그러한 구성요소가 어떤 실시예에 포함되지 않는 것으로 의도되지는 않는다.
- [0032] 개시된 실시예의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나, 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명이 완전하도록 하고, 본 발명이 통상의 기술자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것일 뿐이다.
- [0033] 본 명세서에서 사용되는 용어에 대해 간략히 설명하고, 개시된 실시예에 대해 구체적으로 설명하기로 한다. 본 명세서에서 사용되는 용어는 본 발명에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 관련 분야에 종사하는 기술자의 의도 또는 판례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 발명의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서, 본 발명에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 발명의 전반에 걸친 내용을 토대로 정의되어야 한다.
- [0034] 본 명세서에서의 단수의 표현은 문맥상 명백하게 단수인 것으로 특정하지 않는 한, 복수의 표현을 포함한다. 또한, 복수의 표현은 문맥상 명백하게 복수인 것으로 특정하지 않는 한, 단수의 표현을 포함한다. 명세서 전체에서 어떤 부분이 어떤 구성요소를 포함한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있음을 의미한다.
- [0035] 본 발명에서, "포함하다", "포함하는" 등의 용어는 특징들, 단계들, 동작들, 요소들 및/또는 구성 요소들이 존재하는 것을 나타낼 수 있으나, 이러한 용어가 하나 이상의 다른 기능들, 단계들, 동작들, 요소들, 구성 요소들 및/또는 이들의 조합이 추가되는 것을 배제하지는 않는다.
- [0036] 본 발명에서, 특정 구성 요소가 임의의 다른 구성 요소에 "결합", "조합", "연결" 되거나, "반응" 하는 것으로 언급된 경우, 특정 구성 요소는 다른 구성 요소에 직접 결합, 조합 및/또는 연결되거나, 반응할 수 있으나, 이에 한정되지 않는다. 예를 들어, 특정 구성 요소와 다른 구성 요소 사이에 하나 이상의 중간 구성 요소가 존재할 수 있다. 또한, 본 발명에서 "및/또는"은 열거된 하나 이상의 항목의 각각 또는 하나 이상의 항목의 적어도 일부의 조합을 포함할 수 있다.
- [0037] 본 발명에서, "제1", "제2" 등의 용어는 특정 구성 요소를 다른 구성 요소와 구별하기 위해 사용되는 것으로, 이러한 용어에 의해 상술된 구성 요소가 제한되진 않는다. 예를 들어, "제1" 구성 요소는 "제2" 구성 요소와 동일하거나 유사한 형태의 요소일 수 있다.
- [0038] 본 발명에서, "인공신경망 모델"은, 기계학습 모델의 일 예로서, 주어진 입력에 대한 답을 추론하는데 사용하는 임의의 모델을 포함할 수 있다. 일 실시예에 따르면, 인공신경망 모델은 입력 레이어(층), 복수 개의 은닉 레이어 및 출력 레이어를 포함한 뉴럴 네트워크를 포함할 수 있다. 여기서, 각 레이어는 하나 이상의 노드를 포함할 수 있다. 또한, 인공신경망 모델은 인공신경망 모델에 포함된 복수의 노드와 연관된 가중치를 포함할 수 있다. 여기서, 가중치는 인공신경망 모델과 연관된 임의의 파라미터를 포함할 수 있다.

- [0039] 본 발명에서, "윈도잉(windowing)"은, 연속적이고 시간 의존적(time dependent)인 청각 데이터(예: 오디오 데이터)를 시간의 영향을 받지 않는 짧은 구간의 신호로 분할하는 기법 및/또는 작업을 지칭할 수 있다. 청각 데이터를 짧은 구간의 신호로 분할하는 경우 각 구간의 경계가 불연속해지는 문제를 방지하기 위해, 각 구간의 끝값을 0으로 수렴시키기 위한 임의의 윈도우 함수(window function)가 적용될 수 있다.
- [0040] 본 발명에서, "푸리에 변환(Fourier Transform)"은, 청각 데이터와 연관된 오디오 신호를 시간이 아닌 서로 다른 주파수(frequency)로 표현하는 변환 기법 및/또는 작업을 지칭할 수 있다. 또한, 이산 푸리에 변환(Discrete Fourier Transformation)은 이산적인 데이터 집합에 대한 푸리에 변환을 지칭할 수 있다. 이산 푸리에 변환은 윈도잉이 수행된 각 프레임마다 수행될 수 있으며, 이산 푸리에 변환된 각 프레임을 결합하는 경우 규모 스펙트럼(magnitude spectrum)이 생성될 수 있다.
- [0041] 본 발명에서, "멜 스케일(Mel scale)"은, 임의의 오디오 데이터를 사람이 인식 가능한 스케일로 변환하는 스케일 기법 및/또는 작업일 수 있다. 예를 들어, 멜 스케일은 사람이 인식 가능한 사전 결정된 기준 범위의 진동수를 갖는 주파수에 대해서는 선형 스케일(linear scale) 변환을 수행하고, 그 범위 이상의 진동수를 갖는 주파수에 대해서는 로그 스케일(log scale) 변환을 수행하는 기법일 수 있다.
- [0042] 본 발명에서, "이산 코사인 변환(Discrete Cosine Transform, DCT)"은, 임의의 함수를 코사인 함수의 합으로 변환하는 기법 및/또는 작업일 수 있다. 이산 코사인 변환을 수행하는 경우, 코사인 함수의 계수가 변환 전 데이터의 대부분의 정보를 가지고 있을 수 있다.
- [0043] 본 발명에서, "MFCC(Mel Frequency Cepstral Coefficient) 특징 정보"는, 계수(cepstrum coefficient) 중 많은 정보 및 에너지를 포함하는 12개의 계수 및 12개의 계수에 해당하는 각 프레임의 에너지에 대한 정보를 포함하는 특징 정보일 수 있다. 즉, MFCC 특징 정보는 청각 데이터와 연관된 특징(feature)을 포함할 수 있다.
- [0044] 본 발명에서, "분류 정보"는, 인공신경망 모델에 의해 추출되거나 생성된 동물의 클래스(class) 정보를 포함할 수 있다.
- [0045] 도 1은 본 발명의 일 실시예에 따른 동물과 연관된 라벨링 정보(142)가 생성되는 예시를 나타내는 도면이다. 도시된 것과 같이, 영상(110) 및/또는 영상 데이터(video data)가 제공되는 경우, 제공된 영상(110)에 대한 처리(processing)가 수행되어 해당 영상(110)과 연관된 라벨링 정보(142)가 생성될 수 있다. 여기서, 영상(110) 및/또는 영상 데이터는 온라인(online) 상에서 이용 가능한 임의의 영상 및/또는 해당 영상과 연관된 데이터일 수 있으며, 예를 들어, TV 프로그램, 미디어 플랫폼 등에 게재된 영상 등을 포함할 수 있으나, 이에 한정되지 않는다. 또한, 라벨링 정보(142)는 영상(110)에 포함된 동물의 좌표, 영역, 클래스(class) 등을 특정하기 위한 정보일 수 있다.
- [0046] 일 실시예에 따르면, 영상(110)으로부터 동물에 대응하는 라벨링 정보(142)를 추출하기 위해, 해당 영상(110)과 연관된 시각 데이터(112) 및/또는 청각 데이터(114)가 추출되거나 생성될 수 있다. 이 경우, 영상(110)으로부터 시각 데이터(112) 및/또는 청각 데이터(114)를 구분하여 추출하기 위한 임의의 알고리즘이 이용될 수 있다. 즉, 영상(110) 및/또는 영상 데이터를 임의의 알고리즘에 제공하는 경우, 시각 데이터(112) 및/또는 청각 데이터(114)가 추출되거나 생성될 수 있다. 그리고 나서, 이와 같이 추출된 시각 데이터(112)와 청각 데이터(114)를 이용하여 임의의 동물을 특정하기 위한 정보가 생성되거나 추출될 수 있다.
- [0047] 일 실시예에 따르면, 영상(110)과 연관된 시각 데이터(112)는 학습된 제1 인공신경망 모델(120)에 제공될 수 있다. 또한, 제1 인공신경망 모델(120)은 제공된 시각 데이터(112)를 이용하여 동물의 이미지(image)에 대응하는 제1 분류 정보(122)를 추출하거나 생성할 수 있다. 여기서, 제1 인공신경망 모델(120)은 적어도 하나의 분류기(classifier)와 연관된 모델일 수 있다. 예를 들어, 제1 분류 정보(122)를 생성하기 위해 사전 결정된 임의의 알고리즘을 통해 시각 데이터(112)에 포함된 복수의 객체(object)가 검출될 수 있다. 이와 같이 검출된 복수의 객체는 제1 분류기를 통해 차량, 동물, 과일 등의 카테고리(category)로 분류될 수 있다. 그리고 나서, 동물 카테고리로 분류된 적어도 일부의 객체는 제2 분류기에 제공될 수 있으며, 이에 따라 제2 분류기는 적어도 일부의 객체에 대응하는 동물의 종류 및/또는 클래스(예: 개, 고양이, 호랑이 등)에 대한 분류를 수행할 수 있다.
- [0048] 추가적으로, 영상(110)과 연관된 청각 데이터(114)는 학습된 제2 인공신경망 모델(130)에 제공될 수 있다. 또한, 제2 인공신경망 모델(130)은 제공된 청각 데이터(114)를 이용하여 동물의 음성(sound)에 대응하는 제2 분류 정보를 추출하거나 생성할 수 있다. 여기서, 인공신경망 모델(130)은 적어도 하나의 세분화부(segmentation unit) 및/또는 분류기와 연관될 수 있다. 예를 들어, 제2 분류 정보(132)를 생성하기 위해 임의의 세분화부로부터 청각 데이터(114)는 배경음악(background music), 대화(dialogues), 동물 울음소리(animal's sound) 등으로

세분화될 수 있다. 그리고 나서, 동물 울음소리로 세분화된 데이터는 임의의 분류기에 제공될 수 있으며, 이에 따라 임의의 분류기는 동물 울음소리에 대응하는 동물의 종류 및/또는 클래스(예: 개, 고양이, 호랑이 등)에 대한 분류를 수행할 수 있다.

[0049] 일 실시예에 따르면, 추출되거나 생성된 제1 분류 정보(122) 및 제2 분류 정보(132)를 이용하여 영상(110)에 포함된 동물에 대한 라벨링이 수행되어 라벨링 정보(142)가 생성될 수 있다. 여기서, 제1 분류 정보(122) 및 제2 분류 정보(132)는 상술한 바와 같이 시각 및 청각을 기초로 분류된 동물에 대한 클래스 정보를 포함할 수 있다. 도시된 것과 같이, 제1 분류 정보(122) 및 제2 분류 정보(132)는 추론 모델(140)에 제공될 수 있으며, 추론 모델(140)은 제1 분류 정보(122) 및 제2 분류 정보(132)를 기초로 최종적인 라벨링 정보(142)를 생성할 수 있다. 여기서, 추론 모델(140)은 임의의 추론 알고리즘(reasoning algorithm)과 연관될 수 있으며, 예를 들어, 퍼지 논리(fuzzy logic) 기반의 결정 모듈을 포함할 수 있다. 이와 같은 구성에 의해, 사람이 직접 영상 데이터에 대한 라벨링을 수행하지 않는 경우에도, 영상 데이터의 시각적 특징 및 청각적 특징을 모두 이용하여 높은 정확도로 동물에 대한 라벨링이 수행될 수 있다.

[0050] 도 2는 본 발명의 일 실시예에 따른 컴퓨팅 장치(200)의 내부 구성을 나타내는 기능적인 블록도이다. 도시된 바와 같이, 컴퓨팅 장치(200)는 시각 데이터 처리부(210), 청각 데이터 처리부(220), 라벨링 수행부(230) 등을 포함할 수 있으나, 이에 한정되지 않는다. 컴퓨팅 장치(200)는 임의의 외부 장치와 통신하며 동물에 대한 라벨링에 필요한 데이터 및/또는 정보를 주고받을 수 있다.

[0051] 일 실시예에 따르면, 컴퓨팅 장치(200)는 임의의 영상 데이터를 수신할 수 있다. 이 경우, 시각 데이터 처리부(210)는 수신된 영상 데이터로부터 동물과 연관된 시각 데이터(visual data)를 추출할 수 있다. 여기서, 시각 데이터는 영상 데이터를 프레임(frame) 단위로 분할할 복수의 이미지 등을 포함할 수 있다. 시각 데이터 처리부(210)는 추출된 시각 데이터를 학습된 제1 인공신경망 모델에 제공하여 동물의 이미지에 대응하는 제1 분류 정보를 생성할 수 있다. 예를 들어, 시각 데이터 처리부(210)는 시각 데이터에 포함된 복수의 객체를 검출하고, 검출된 복수의 객체 중 동물에 대응하는 적어도 일부의 객체를 결정할 수 있다. 그리고 나서, 시각 데이터 처리부(210)는 결정된 적어도 일부의 객체와 연관된 동물의 이미지에 대응하는 제1 분류 정보를 생성할 수 있다.

[0052] 일 실시예에 따르면, 제1 인공신경망 모델은 시각 데이터에서 동물에 대응하는 적어도 일부의 객체와 연관된 위치 및 영역을 검출하기 위한 YOLO(You Only Look Once) 모델(예: YOLO-V4 모델)을 포함할 수 있다. 여기서, YOLO 모델은 이미지에서 관심 객체를 배경과 구분해 식별하기 위한 객체 탐지 모델의 일종일 수 있다. 즉, 시각 데이터 처리부(210)는 이러한 YOLO 모델을 이용하여 시각 데이터에서 동물에 대응하는 적어도 일부의 객체와 연관된 위치, 영역, 클래스 정보 등을 높은 정확도로 탐지할 수 있다.

[0053] 일 실시예에 따르면, 청각 데이터 처리부(220)는 수신된 영상 데이터로부터 동물과 연관된 청각 데이터(aural data)를 추출할 수 있다. 여기서, 청각 데이터는 영상 데이터에 삽입된 임의의 오디오 데이터를 포함할 수 있다. 청각 데이터 처리부(220)는 추출된 청각 데이터를 학습된 제2 인공신경망 모델에 제공하여 동물의 음성(sound)에 대응하는 제2 분류 정보를 생성할 수 있다. 예를 들어, 청각 데이터 처리부(220)는 청각 데이터에 포함된 복수의 유형의 음성 정보를 검출하고, 검출된 복수의 유형의 음성 정보 중 동물에 대응하는 적어도 일부의 음성 정보를 결정할 수 있다. 그리고 나서, 청각 데이터 처리부(220)는 결정된 적어도 일부의 음성 정보와 연관된 동물의 음성에 대응하는 제2 분류 정보를 생성할 수 있다.

[0054] 일 실시예에 따르면, 제2 인공신경망 모델은 CNN(Convolution Neural Network) 모델을 포함할 수 있다. 청각 데이터 처리부(220)는 동물에 대응하는 적어도 일부의 음성 정보에 대한 전처리를 수행하여 MFCC(Mel Frequency Cepstral Coefficient) 특징 정보를 추출하고, 추출된 MFCC 특징 정보를 CNN 모델에 제공하여 동물의 음성에 대응하는 제2 분류 정보를 생성할 수 있다. 여기서, MFCC 특징 정보는 청각 데이터와 연관된 소리의 고유한 특징을 나타내는 정보일 수 있다. 즉, 청각 데이터 처리부(220)는 청각 데이터의 특징을 나타내는 MFCC 특징 정보를 추출한 후, 추출된 MFCC 특징 정보를 학습된 CNN 모델에 제공함으로써 동물의 음성에 대응하는 제2 분류 정보를 효과적으로 생성할 수 있다.

[0055] 일 실시예에 따르면, 라벨링 수행부(230)는 제1 분류 정보 및 제2 분류 정보를 이용하여 영상 데이터에 포함된 동물에 대한 라벨링을 수행할 수 있다. 즉, 라벨링 수행부(230)는 이미지 처리를 수행하여 라벨링을 수행하는 일반적인 방식에 비해 청각 데이터를 추가적으로 이용함으로써, 임의의 영상에 포함된 동물에 대한 높은 정확성을 갖는 라벨링을 수행할 수 있다. 상술한 바와 같이, 라벨링 수행부(230)는 퍼지 논리 기반의 결정 모듈을 통해 라벨링을 수행할 수 있으나, 이에 한정되지 않는다.

- [0056] 도 2에서는 컴퓨팅 장치(200)에 포함된 각각의 기능적인 구성이 구분되어 상술되었으나, 이는 발명의 이해를 돕기 위한 것일 뿐이며, 하나의 연산 장치에서 둘 이상의 기능을 수행할 수도 있다. 이와 같은 구성에 의해, 컴퓨팅 장치(200)는 온라인 상에서 검색 가능한 임의의 영상 데이터로부터 간단히 동물에 대한 라벨링을 수행할 수 있으며, 이에 따라 수많은 소스(source)로부터 라벨링 데이터가 효율적으로 확보될 수 있다.
- [0057] 도 3은 본 발명의 일 실시예에 따른 YOLO V4 모델(300)의 세부적인 구성을 나타내는 예시적인 도면이다. 상술된 바와 같이, 제1 인공신경망 모델(도 1의 120)은 시각 데이터에서 동물에 대응하는 적어도 일부의 객체와 연관된 위치 및 영역을 검출하기 위한 YOLO 모델을 포함할 수 있다. 예를 들어, YOLO 모델은 YOLO V4 모델(300)일 수 있다. 도시된 바와 같이, YOLO V4 모델(300)은 입력 모듈(input module), 백본 모듈(backbone module), 넥 모듈(neck module), 예측 모듈(dense prediction module) 등을 포함할 수 있다.
- [0058] 일 실시예에 따르면, YOLO V4 모델(300)은 지도 학습(supervised learning)에 의해 학습될 수 있는 단일 단계 검출기(single stage detector)일 수 있으며, 이에 따라 입력 모듈을 통해 수신되는 입력은 객체 검출에 따라 바운딩 박스(bounding box) 등으로 지정된 복수의 객체를 포함하는 이미지 등일 수 있다. 또한, YOLO V4 모델(300)의 백본 모듈은 이미지에 포함된 특징을 추출하기 위한 모듈로서, YOLO V3 모델의 Darknet53에 CSP(Cross-Stage-Partial)가 연결되어 개선된 모듈일 수 있다.
- [0059] 또한, YOLO V4 모델(300)의 넥 모듈은 백본 모듈로부터 추출된 특징에 대한 혼합 및/또는 결합을 수행하기 위한 모듈일 수 있다. YOLO V4 모델(300)에서 넥 모듈은 최신 버전의 PANet(Path Aggregation Network)을 포함할 수 있다. 이와 같이 혼합 및/또는 결합된 특징을 기초로 최종적인 객체 검출을 수행하기 위해 혼합 및/또는 결합된 특징은 완전 연결 레이어(fully connected layer)에 제공되어 압축될 수 있다.
- [0060] 이와 관련하여, YOLO V4 모델(300)의 예측 모듈은 추출된 특징을 이용하여 동물의 클래스에 관한 최종적인 예측을 수행하고, 해당 동물이 위치한 영역을 나타내는 바운딩 박스 좌표, 신뢰도 점수 등을 생성할 수 있다. 이와 같은 구성에 의해, 영상으로부터 추출된 시각 데이터를 YOLO V4 모델(300)에 제공하는 경우, 높은 신뢰성을 갖는 분류 정보(예: 동물의 클래스 정보 등)가 생성될 수 있다.
- [0061] 도 4는 본 발명의 일 실시예에 따른 청각 데이터와 연관된 분류 정보가 생성되는 과정을 나타내는 예시적인 흐름도이다. 상술된 바와 같이, 영상 데이터가 수신될 수 있으며, 영상 데이터로부터 동물과 연관된 청각 데이터가 획득될 수 있다(410). 이 경우, 컴퓨팅 장치(도 2의 200)는 획득된 청각 데이터를 이용하여 동물의 음성에 대응하는 제2 분류 정보를 생성할 수 있다. 여기서, 청각 데이터는 사인파(sine wave) 형태의 파형을 가질 수 있으나, 이에 한정되지 않는다.
- [0062] 일 실시예에 따르면, 컴퓨팅 장치는 청각 데이터에 포함된 복수의 유형의 음성 정보를 검출하고, 검출된 복수의 유형의 음성 정보 중 동물에 대응하는 적어도 일부의 음성 정보를 결정할 수 있다. 그리고 나서, 컴퓨팅 장치는 결정된 적어도 일부의 음성 정보와 연관된 동물의 음성에 대응하는 제2 분류 정보를 생성할 수 있다. 예를 들어, 컴퓨팅 장치는 동물에 대응하는 적어도 일부의 음성 정보에 대한 전처리를 수행할 수 있다(420).
- [0063] 일 실시예에 따르면, 컴퓨팅 장치는 동물에 대응하는 적어도 일부의 음성 정보를 사전 결정된 윈도우(windowing) 알고리즘을 이용하여 복수의 프레임(frame)으로 분할할 수 있다. 예를 들어, 연속적인 1차원 밴드 패스(band-pass) 필터는 작은 프레임으로 분할될 수 있으며, 이 때 윈도우 알고리즘이 이용될 수 있다. 여기서, 윈도우 알고리즘은 연속적인 데이터에서 특정 개수의 구간을 결정하기 위한 알고리즘일 수 있다. 이러한 윈도우 알고리즘을 이용하는 경우, 스펙트럼 왜곡이 감소하고 신호의 연속성이 유지될 수 있다.
- [0064] 일 실시예에 따르면, 윈도우 알고리즘은 사각 윈도우(rectangular window), 해밍 윈도우(Hamming window), 해닝 윈도우(Hanning window), 삼각 윈도우(triangular window) 등을 포함할 수 있으나, 이에 한정되지 않는다. 여기서, 해밍 윈도우는 각 프레임의 시작 부분과 끝 부분에서 샘플을 0으로 테이퍼(taper)하기 위해 사용될 수 있다. 예를 들어, 해밍 윈도우는 다음의 수학적 식 1과 같이 구성될 수 있다.

수학적 식 1

- [0066] $w[n] = 0.54 - 0.46\cos(2\pi nN - 1)$

[0068] 여기서, N 은 각 프레임 사이즈는 나타낼 수 있으며, n 은 샘플을 나타낼 수 있다. 이 경우, n 은 $0 \leq n \leq N-1$ 의 값을 가질 수 있다. 추가적으로, 원도영이 수행된 각 프레임에 대한 이산 푸리에 변환(Discrete Fourier Transformation)이 수행될 수 있다. 일 실시예에 따르면, 컴퓨팅 장치는 이산 푸리에 변환을 이용하여 복수의 프레임을 규모 스펙트럼으로 변환할 수 있다. 여기서, 이산 푸리에 변환은 이산적인 데이터 집합의 푸리에 변환으로서, 신호를 진동수의 성분으로 분해하는 수학적 기법일 수 있다. 예를 들어, 이산 푸리에 변환은 다음의 수식 2를 이용하여 수행될 수 있다.

수학식 2

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N}, 0 \leq k \leq N-1$$

[0072] 여기서, N 은 이산 푸리에 변환을 연산하는데 사용되는 포인트의 수일 수 있다. 추가적으로, 이산 푸리에 변환에 의해 생성된 규모 스펙트럼에 대한 멜 스케일이 수행될 수 있다. 즉, 컴퓨팅 장치는 사전 결정된 멜 스케일(Mel scale) 알고리즘을 이용하여 규모 스펙트럼을 로그 멜 스펙트럼(log Mel spectrum)으로 변환할 수 있다. 이 경우, 멜 스케일과 연관된 멜 필터(Mel filter)는 두 개의 인접한 필터의 중심 주파수에 의해 결정되는 컷오프(cut-off) 주파수를 갖는 중첩된 삼각형 필터로 구성될 수 있다. 이러한 필터는 선형적 간격을 갖는 중심 주파수와 멜 스케일에서 고정 대역폭을 가질 수 있다. 예를 들어, 로그 멜 스펙트럼은 다음의 수식 3에 의해 생성될 수 있다.

수학식 3

$$F(Mel) = [2595 * \log_{10}\{1 + F\}700]$$

[0076] 일 실시예에 따르면, 로그 멜 스펙트럼에 대한 이산 코사인 변환(Discrete Cosine Transform)이 수행될 수 있다. 이산 코사인 변환이 수행되는 경우 로그 멜 스펙트럼이 시간 영역으로 변환될 수 있으며, 변환에 따라 MFCC 특징 정보가 생성될 수 있다(430).

[0077] 일 실시예에 따르면, 이와 같이 생성된 MFCC 특징 정보를 이용하여 분류 정보가 생성될 수 있다. 다시 말해, 컴퓨팅 장치는 MFCC 특징 정보를 이용하여 청각 데이터와 연관된 분류 정보를 생성할 수 있다(440). 예를 들어, 컴퓨팅 장치는 MFCC 특징 정보를 학습된 CNN 모델에 제공하여 동물의 클래스 정보를 포함하는 분류 정보를 생성할 수 있다.

[0078] 도 5는 본 발명의 일 실시예에 따른 CNN 모델(500)의 구조를 나타내는 예시적인 도면이다. 상술된 바와 같이, 제2 인공신경망 모델은 CNN 모델(500)을 포함할 수 있으며, 컴퓨팅 장치(도 2의 200)는 MFCC 특징 정보를 CNN 모델(500)에 제공하여 동물의 음성에 대응하는 분류 정보를 생성할 수 있다. 여기서, CNN 모델(500)은 MFCC 특징 정보 중 가장 두드러진 특징만을 선택하여 분류를 수행하기 위한 모델일 수 있다.

[0079] 도시된 것과 같이, CNN 모델(500)은 제1 컨볼루션 블록(convolution block)(510), 제2 컨볼루션 블록(520) 및 제3 컨볼루션 블록(530)을 포함하는 3개의 컨볼루션 블록(510, 520, 530)을 포함할 수 있다. 이 경우, 각각의 컨볼루션 블록(510, 520, 530)은 컨볼루션 레이어(convolution layer), 배치 정규화 레이어(batch normalization layer), 맥스 풀링 레이어(max pooling layer) 등을 포함할 수 있다. 또한, 각 컨볼루션 레이어

는 ReLu 활성화 함수와 연결될 수 있다.

- [0080] 일 실시예에 따르면, 제1 컨볼루션 블록(510)에 포함된 컨볼루션 레이어는 1D 데이터를 수신하고, 3x3 크기의 64개의 서로 다른 필터를 적용하여 MFCC 특징 정보 중 두드러진 특징을 출력할 수 있다. 이와 같은 출력된 정보는 128개의 필터 수를 갖는 제2 컨볼루션 블록(520)에 포함된 컨볼루션 레이어에 제공될 수 있다. 이 경우, 출력의 크기(size)는 변경되지 않을 수 있다.
- [0081] 일 실시예에 따르면, 제3 컨볼루션 블록(530)에 포함된 컨볼루션 레이어는 2x2 크기의 192개의 필터를 포함할 수 있다. 예를 들어, 제3 컨볼루션 블록(530)의 출력은 네트워크의 다른 뉴런과 직접 연결된 64개의 뉴런을 갖는 텐스 레이어(540)에 제공될 수 있다. 여기서, 텐스 레이어(540)는 출력에 대한 과적합을 줄이기 위한 드롭아웃 레이어(dropout layer)와 연결될 수 있다. 최종적으로, 드롭아웃 레이어의 출력은 출력 레이어(550)(예: 소프트맥스 레이어)에 제공되어 청각 데이터에 포함된 각 음성이 특정 동물 클래스로 분류될 수 있다. 이와 같은 구성에 의해, 3개의 컨볼루션 블록(510, 520, 530)을 포함하는 CNN 모델을 이용함으로써, 다른 임의의 개수의 컨볼루션 블록을 이용하는 경우보다 높은 성능으로 동물 클래스 분류가 수행될 수 있다.
- [0082] 도 6은 본 발명의 일 실시예에 따른 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법(600)의 예시를 나타내는 도면이다. 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법(600)은 적어도 하나의 프로세서(예: 컴퓨팅 장치의 적어도 하나의 프로세서)에 의해 수행될 수 있다. 동물 탐지 및 분류를 위한 영상 데이터 라벨링 방법(600)은 프로세서가 영상 데이터를 수신함으로써 개시될 수 있다(S610).
- [0083] 프로세서는 수신된 영상 데이터로부터 동물과 연관된 시각 데이터 및 청각 데이터를 추출할 수 있다(S620). 이 경우, 프로세서는 추출된 시각 데이터를 학습된 제1 인공신경망 모델에 제공하여 동물의 이미지에 대응하는 제1 분류 정보를 생성할 수 있다(S630). 또한, 프로세서는 추출된 청각 데이터를 학습된 제2 인공신경망 모델에 제공하여 동물의 음성에 대응하는 제2 분류 정보를 생성할 수 있다(S640). 여기서, 제1 인공신경망 모델은 시각 데이터에서 동물에 대응하는 적어도 일부의 객체와 연관된 위치 및 영역을 검출하기 위한 YOLO 모델을 포함할 수 있으며, 제2 인공신경망 모델은 CNN 모델을 포함할 수 있다. 그리고 나서, 프로세서는 생성된 제1 분류 정보 및 제2 분류 정보를 이용하여 영상 데이터에 포함된 동물에 대한 라벨링을 수행할 수 있다(S650).
- [0084] 일 실시예에 따르면, 프로세서는 시각 데이터에 포함된 복수의 객체를 검출하고, 검출된 복수의 객체 중 동물에 대응하는 적어도 일부의 객체를 결정할 수 있다. 이 경우, 프로세서는 결정된 적어도 일부의 객체와 연관된 동물의 이미지에 대응하는 제1 분류 정보를 생성할 수 있다. 추가적으로 또는 대안적으로, 프로세서는 청각 데이터에 포함된 복수의 유형의 음성 정보를 검출하고, 검출된 복수의 유형의 음성 정보 중 동물에 대응하는 적어도 일부의 음성 정보를 결정할 수 있다. 또한, 프로세서는 결정된 적어도 일부의 음성 정보와 연관된 동물의 음성에 대응하는 제2 분류 정보를 생성할 수 있다.
- [0085] 프로세서는 동물에 대응하는 적어도 일부의 음성 정보에 대한 전처리를 수행하여 MFCC 특징 정보를 추출하고, 추출된 MFCC 특징 정보를 CNN 모델에 제공하여 동물의 음성에 대응하는 제2 분류 정보를 생성할 수 있다. 예를 들어, 프로세서는 동물에 대응하는 적어도 일부의 음성 정보를 사전 결정된 윈도우 알고리즘을 이용하여 복수의 프레임으로 분할할 수 있다. 추가적으로, 프로세서는 이산 푸리에 변환을 이용하여 복수의 프레임을 규모 스펙트럼으로 변환하고, 사전 결정된 멜 스케일 알고리즘을 이용하여 규모 스펙트럼을 로그 멜 스펙트럼으로 변환할 수 있다. 프로세서는 사전 결정된 이산 코사인 변환 알고리즘을 이용하여 로그 멜 스펙트럼을 기초로 MFCC 특징 정보를 생성할 수 있다.
- [0086] 도 7은 본 발명의 일 실시예에 따른 컴퓨팅 장치(700)의 내부 구성을 나타내는 블록도이다. 컴퓨팅 장치(700)는 메모리(710), 프로세서(720), 통신 모듈(730) 및 입출력 인터페이스(740)를 포함할 수 있다. 도 7에 도시된 바와 같이, 컴퓨팅 장치(700)는 통신 모듈(730)을 이용하여 네트워크를 통해 정보 및/또는 데이터를 통신할 수 있도록 구성될 수 있다.
- [0087] 메모리(710)는 비-일시적인 임의의 컴퓨터 판독 가능한 기록매체를 포함할 수 있다. 일 실시예에 따르면, 메모리(710)는 RAM(random access memory), ROM(read only memory), 디스크 드라이브, SSD(solid state drive), 플래시 메모리(flash memory) 등과 같은 비소멸성 대용량 저장 장치(permanent mass storage device)를 포함할 수 있다. 다른 예로서, ROM, SSD, 플래시 메모리, 디스크 드라이브 등과 같은 비소멸성 대용량 저장 장치는 메모리와는 구분되는 별도의 영구 저장 장치로서 컴퓨팅 장치(700)에 포함될 수 있다. 또한, 메모리(710)에는 운영체제와 적어도 하나의 프로그램 코드가 저장될 수 있다.
- [0088] 이러한 소프트웨어 구성요소들은 메모리(710)와는 별도의 컴퓨터에서 판독 가능한 기록매체로부터 로딩될 수 있

다. 이러한 별도의 컴퓨터에서 판독 가능한 기록매체는 이러한 컴퓨팅 장치(700)에 직접 연결가능한 기록 매체를 포함할 수 있는데, 예를 들어, 플로피 드라이브, 디스크, 테이프, DVD/CD-ROM 드라이브, 메모리 카드 등의 컴퓨터에서 판독 가능한 기록매체를 포함할 수 있다. 다른 예로서, 소프트웨어 구성요소들은 컴퓨터에서 판독 가능한 기록매체가 아닌 통신 모듈(730)을 통해 메모리(710)에 로딩될 수도 있다. 예를 들어, 적어도 하나의 프로그램은 개발자들 또는 어플리케이션의 설치 파일을 배포하는 파일 배포 시스템이 통신 모듈(730)을 통해 제공하는 파일들에 의해 설치되는 컴퓨터 프로그램에 기반하여 메모리(710)에 로딩될 수 있다.

[0089] 프로세서(720)는 기본적인 산술, 로직 및 입출력 연산을 수행함으로써, 컴퓨터 프로그램의 명령을 처리하도록 구성될 수 있다. 명령은 메모리(710) 또는 통신 모듈(730)에 의해 사용자 단말(미도시) 또는 다른 외부 시스템으로 제공될 수 있다.

[0090] 통신 모듈(730)은 네트워크를 통해 사용자 단말(미도시)과 컴퓨팅 장치(700)가 서로 통신하기 위한 구성 또는 기능을 제공할 수 있으며, 컴퓨팅 장치(700)가 외부 시스템(일례로 별도의 클라우드 시스템 등)과 통신하기 위한 구성 또는 기능을 제공할 수 있다. 일례로, 컴퓨팅 장치(700)의 프로세서(720)의 제어에 따라 제공되는 제어 신호, 명령, 데이터 등이 통신 모듈(730)과 네트워크를 거쳐 사용자 단말 및/또는 외부 시스템의 통신 모듈을 통해 사용자 단말 및/또는 외부 시스템으로 전송될 수 있다.

[0091] 또한, 컴퓨팅 장치(700)의 입출력 인터페이스(740)는 컴퓨팅 장치(700)와 연결되거나 컴퓨팅 장치(700)가 포함할 수 있는 입력 또는 출력을 위한 장치(미도시)와의 인터페이스를 위한 수단일 수 있다. 도 7에서는 입출력 인터페이스(740)가 프로세서(720)와 별도로 구성된 요소로서 도시되었으나, 이에 한정되지 않으며, 입출력 인터페이스(740)가 프로세서(720)에 포함되도록 구성될 수 있다. 컴퓨팅 장치(700)는 도 7의 구성요소들보다 더 많은 구성요소들을 포함할 수 있다. 그러나, 대부분의 종래기술적 구성요소들을 명확하게 도시할 필요성은 없다.

[0092] 컴퓨팅 장치(700)의 프로세서(720)는 복수의 사용자 단말 및/또는 복수의 외부 시스템으로부터 수신된 정보 및/또는 데이터를 관리, 처리 및/또는 저장하도록 구성될 수 있다.

[0093] 상술된 방법 및/또는 다양한 실시예들은, 디지털 전자 회로, 컴퓨터 하드웨어, 펌웨어, 소프트웨어 및/또는 이들의 조합으로 실현될 수 있다. 본 발명의 다양한 실시예들은 데이터 처리 장치, 예를 들어, 프로그래밍 가능한 하나 이상의 프로세서 및/또는 하나 이상의 컴퓨팅 장치에 의해 실행되거나, 컴퓨터 판독 가능한 기록 매체 및/또는 컴퓨터 판독 가능한 기록 매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다. 상술된 컴퓨터 프로그램은 컴파일된 언어 또는 해석된 언어를 포함하여 임의의 형태의 프로그래밍 언어로 작성될 수 있으며, 독립 실행형 프로그램, 모듈, 서브 루틴 등의 임의의 형태로 배포될 수 있다. 컴퓨터 프로그램은 하나의 컴퓨팅 장치, 동일한 네트워크를 통해 연결된 복수의 컴퓨팅 장치 및/또는 복수의 상이한 네트워크를 통해 연결되도록 분산된 복수의 컴퓨팅 장치를 통해 배포될 수 있다.

[0094] 상술된 방법 및/또는 다양한 실시예들은, 입력 데이터를 기초로 동작하거나 출력 데이터를 생성함으로써, 임의의 기능, 함수 등을 처리, 저장 및/또는 관리하는 하나 이상의 컴퓨터 프로그램을 실행하도록 구성된 하나 이상의 프로세서에 의해 수행될 수 있다. 예를 들어, 본 발명의 방법 및/또는 다양한 실시예는 FPGA(Field Programmable Gate Array) 또는 ASIC(Application Specific Integrated Circuit)과 같은 특수 목적 논리 회로에 의해 수행될 수 있으며, 본 발명의 방법 및/또는 실시예들을 수행하기 위한 장치 및/또는 시스템은 FPGA 또는 ASIC와 같은 특수 목적 논리 회로로서 구현될 수 있다.

[0095] 컴퓨터 프로그램을 실행하는 하나 이상의 프로세서는, 범용 목적 또는 특수 목적의 마이크로 프로세서 및/또는 임의의 종류의 디지털 컴퓨팅 장치의 하나 이상의 프로세서를 포함할 수 있다. 프로세서는 읽기 전용 메모리, 랜덤 액세스 메모리의 각각으로부터 명령 및/또는 데이터를 수신하거나, 읽기 전용 메모리와 랜덤 액세스 메모리로부터 명령 및/또는 데이터를 수신할 수 있다. 본 발명에서, 방법 및/또는 실시예들을 수행하는 컴퓨팅 장치의 구성 요소들은 명령어들을 실행하기 위한 하나 이상의 프로세서, 명령어들 및/또는 데이터를 저장하기 위한 하나 이상의 메모리 디바이스를 포함할 수 있다.

[0096] 일 실시예에 따르면, 컴퓨팅 장치는 데이터를 저장하기 위한 하나 이상의 대용량 저장 장치와 데이터를 주고받을 수 있다. 예를 들어, 컴퓨팅 장치는 자기 디스크(magnetic disc) 또는 광 디스크(optical disc)로부터 데이터를 수신하거나/수신하고, 자기 디스크 또는 광 디스크로 데이터를 전송할 수 있다. 컴퓨터 프로그램과 연관된 명령어들 및/또는 데이터를 저장하기에 적합한 컴퓨터 판독 가능한 저장 매체는, EPROM(Erasable Programmable Read-Only Memory), EEPROM(Electrically Erasable PROM), 플래시 메모리 장치 등의 반도체 메모리 장치를 포함하는 임의의 형태의 비 휘발성 메모리를 포함할 수 있으나, 이에 한정되지 않는다. 예를 들어, 컴퓨터 판독

가능한 저장 매체는 내부 하드 디스크 또는 이동식 디스크와 같은 자기 디스크, 광 자기 디스크, CD-ROM 및 DVD-ROM 디스크를 포함할 수 있다.

[0097] 사용자와의 상호 작용을 제공하기 위해, 컴퓨팅 장치는 정보를 사용자에게 제공하거나 디스플레이하기 위한 디스플레이 장치(예를 들어, CRT (Cathode Ray Tube), LCD(Liquid Crystal Display) 등) 및 사용자가 컴퓨팅 장치 상에 입력 및/또는 명령 등을 제공할 수 있는 포인팅 장치(예를 들어, 키보드, 마우스, 트랙볼 등)를 포함할 수 있으나, 이에 한정되지 않는다. 즉, 컴퓨팅 장치는 사용자와의 상호 작용을 제공하기 위한 임의의 다른 종류의 장치들을 더 포함할 수 있다. 예를 들어, 컴퓨팅 장치는 사용자와의 상호 작용을 위해, 시각적 피드백, 청각 피드백 및/또는 촉각 피드백 등을 포함하는 임의의 형태의 감각 피드백을 사용자에게 제공할 수 있다. 이에 대해, 사용자는 시각, 음성, 동작 등의 다양한 제스처를 통해 컴퓨팅 장치로 입력을 제공할 수 있다.

[0098] 본 발명에서, 다양한 실시예들은 백엔드 구성 요소(예: 데이터 서버), 미들웨어 구성 요소(예: 애플리케이션 서버) 및/또는 프론트 엔드 구성 요소를 포함하는 컴퓨팅 시스템에서 구현될 수 있다. 이 경우, 구성 요소들은 통신 네트워크와 같은 디지털 데이터 통신의 임의의 형태 또는 매체에 의해 상호 연결될 수 있다. 예를 들어, 통신 네트워크는 LAN(Local Area Network), WAN(Wide Area Network) 등을 포함할 수 있다.

[0099] 본 명세서에서 기술된 예시적인 실시예들에 기반한 컴퓨팅 장치는, 사용자 디바이스, 사용자 인터페이스(UI) 디바이스, 사용자 단말 또는 클라이언트 디바이스를 포함하여 사용자와 상호 작용하도록 구성된 하드웨어 및/또는 소프트웨어를 사용하여 구현될 수 있다. 예를 들어, 컴퓨팅 장치는 랩톱(laptop) 컴퓨터와 같은 휴대용 컴퓨팅 장치를 포함할 수 있다. 추가적으로 또는 대안적으로, 컴퓨팅 장치는, PDA(Personal Digital Assistants), 태블릿 PC, 게임 콘솔(game console), 웨어러블 디바이스(wearable device), IoT(internet of things) 디바이스, VR(virtual reality) 디바이스, AR(augmented reality) 디바이스 등을 포함할 수 있으나, 이에 한정되지 않는다. 컴퓨팅 장치는 사용자와 상호 작용하도록 구성된 다른 유형의 장치를 더 포함할 수 있다. 또한, 컴퓨팅 장치는 이동 통신 네트워크 등의 네트워크를 통한 무선 통신에 적합한 휴대용 통신 디바이스(예를 들어, 이동 전화, 스마트 전화, 무선 셀룰러 전화 등) 등을 포함할 수 있다. 컴퓨팅 장치는, 무선 주파수(RF; Radio Frequency), 마이크로파 주파수(MWF; Microwave Frequency) 및/또는 적외선 주파수(IRF; Infrared Ray Frequency)와 같은 무선 통신 기술 및/또는 프로토콜들을 사용하여 네트워크 서버와 무선으로 통신하도록 구성될 수 있다.

[0100] 본 발명에서 특정 구조적 및 기능적 세부 사항을 포함하는 다양한 실시예들은 예시적인 것이다. 따라서, 본 발명의 실시예들은 상술된 것으로 한정되지 않으며, 여러 가지 다른 형태로 구현될 수 있다. 또한, 본 발명에서 사용된 용어는 일부 실시예를 설명하기 위한 것이며 실시예를 제한하는 것으로 해석되지 않는다. 예를 들어, 단수형 단어 및 상기는 문맥상 달리 명확하게 나타내지 않는 한 복수형도 포함하는 것으로 해석될 수 있다.

[0101] 본 발명에서, 달리 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함하여 본 명세서에서 사용되는 모든 용어는 이러한 개념이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 갖는다. 또한, 사전에 정의된 용어와 같이 일반적으로 사용되는 용어들은 관련 기술의 맥락에서의 의미와 일치하는 의미를 갖는 것으로 해석되어야 한다.

[0102] 본 명세서에서는 본 발명이 일부 실시예들과 관련하여 설명되었지만, 본 발명의 발명이 속하는 기술분야의 통상의 기술자가 이해할 수 있는 본 발명의 범위를 벗어나지 않는 범위에서 다양한 변형 및 변경이 이루어질 수 있다. 또한, 그러한 변형 및 변경은 본 명세서에 첨부된 특허청구의 범위 내에 속하는 것으로 생각되어야 한다.

부호의 설명

- [0104] 110: 영상
- 112: 시각 데이터
- 114: 청각 데이터
- 120: 제1 인공신경망 모델
- 122: 제1 분류 정보
- 130: 제2 인공신경망 모델

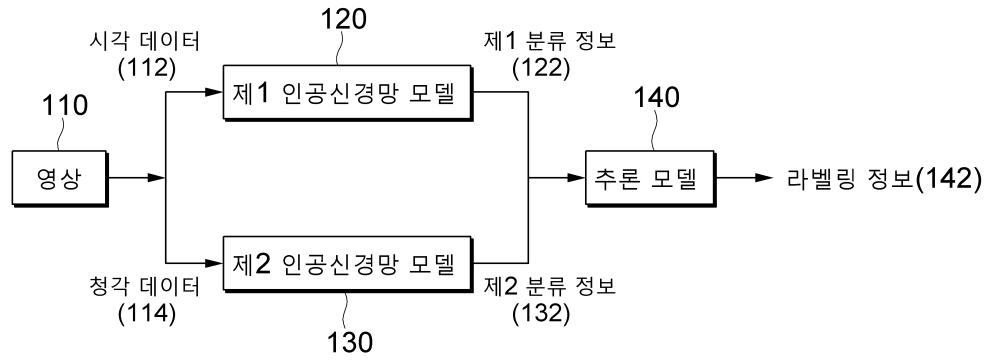
132: 제2 분류 정보

140: 추론 모델

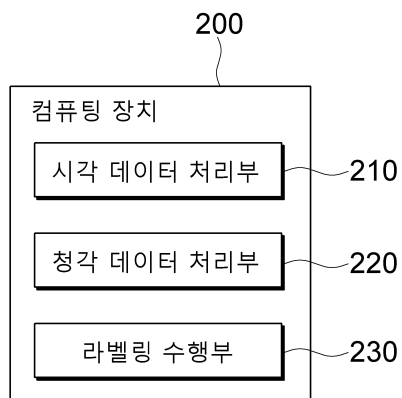
142: 라벨링 정보

도면

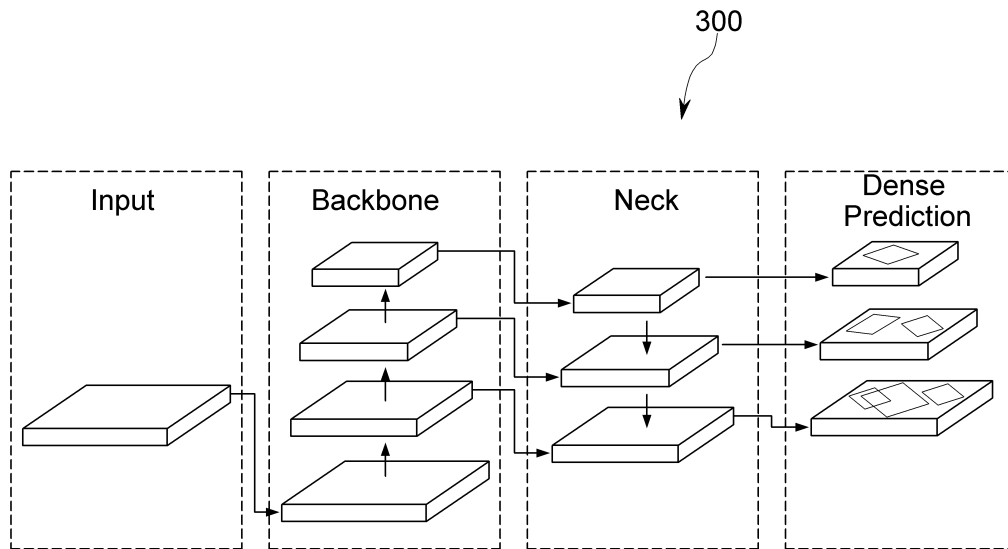
도면1



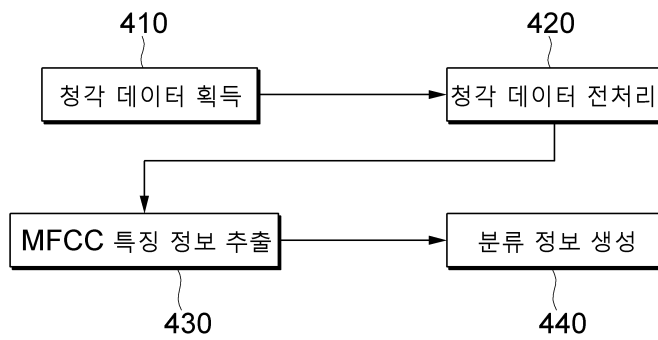
도면2



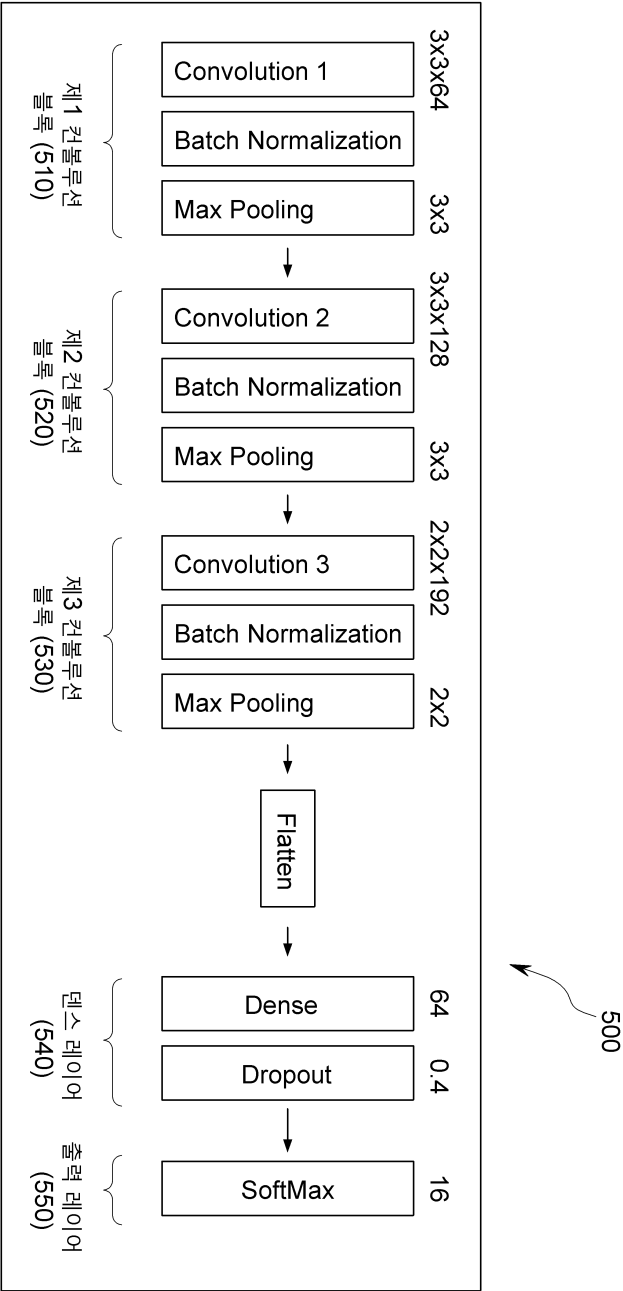
도면3



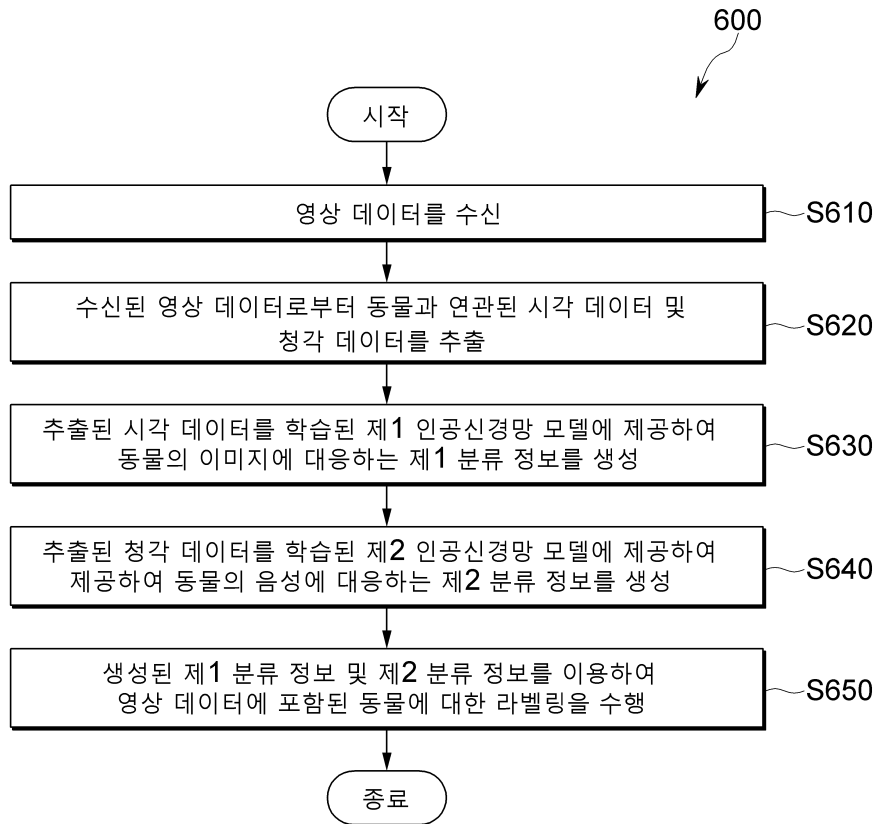
도면4



도면5



도면6



도면7

