



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0043833

(43) 공개일자 2021년04월22일

(51) 국제특허분류(Int. Cl.)

G10L 17/26 (2013.01) G10L 15/20 (2006.01)

G10L 25/18 (2013.01) G10L 25/90 (2013.01)

(52) CPC특허분류

G10L 17/26 (2013.01)

G10L 15/20 (2013.01)

(21) 출원번호 10-2019-0126635

(22) 출원일자 2019년10월14일

심사청구일자 2019년10월14일

(71) 출원인

고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자

고한석

서울특별시 용산구 이촌로88길 30, 101동 1402호
(이촌동, 삼성리버스위트아파트)

이영로

서울특별시 성북구 안암로9가길 68, 인하우스 20
6호(안암동5가)

(뒷면에 계속)

(74) 대리인

윤귀상

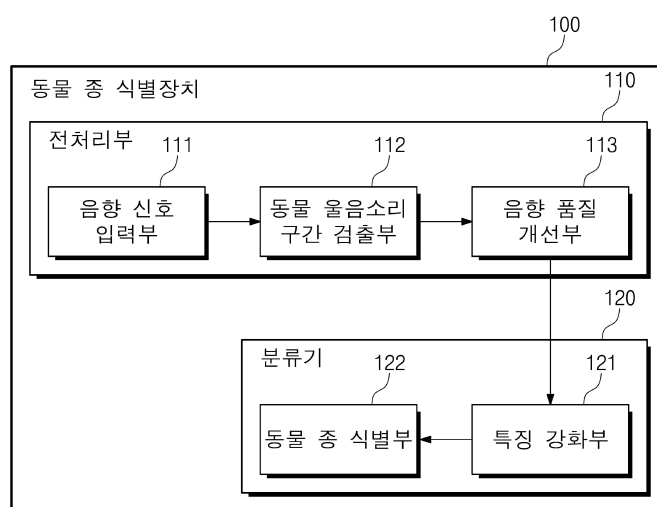
전체 청구항 수 : 총 18 항

(54) 발명의 명칭 잡음 환경에 강인한 동물 종 식별 장치 및 방법

(57) 요약

잡음 환경에 강인한 동물 종 식별 장치 및 방법은 동물 종들의 울음소리에 기초하여 동물 울음소리 구간을 검출하고, 동물 울음소리가 존재하지 않는 구간을 바탕으로 잡음을 추정하여 수신된 소리에서 잡음을 제거한 후, 다양한 잡음에도 강화된 음향 특징을 학습하여 동물 종을 식별할 수 있다.

대표도 - 도1



(52) CPC특허분류

G10L 25/18 (2013.01)

G10L 25/90 (2013.01)

(72) 발명자

김동현

경상북도 경산시 압량면 청운로 75 오페라 101호

박충호

대구광역시 중구 대봉로 254 (대봉동 청구맨션아파트) A동 407호

김정민

서울특별시 노원구 한글비석로 480 (상계동 보람아파트2단지) 202동 208호

고경득

경기도 하남시 대청로59번길 20 (신장동 더우개마을 통일아파트) 102동 403호

이 발명을 지원한 국가연구개발사업

과제고유번호

ARQ201702091003

부처명

환경부

과제관리(전문)기관명

한국환경산업기술원

연구사업명

환경정책기반공공기술개발사업(R&D)

연구과제명

인공지능 및 기계학습 기반 동물 울음소리를 통한 동물 종 구별 원천 기술 개발

기 여 율

1/1

과제수행기관명

고려대학교 산학협력단

연구기간

2019.01.01 ~ 2019.12.31

명세서

청구범위

청구항 1

동물 울음소리의 음향 신호인 1차원 음향 신호를 입력받고, 상기 1차원 음향 신호를 주파수 축과 시간 축을 가지는 2차원 음향 특징인 log-mel 스펙트로그램으로 변환하며, 상기 변환된 2차원 음향 특징을 소리 구간 검출을 통해 상기 동물 울음소리가 존재하는 구간을 검출하는 동물 울음소리 구간 검출부;

상기 검출한 동물 울음소리가 존재하는 구간을 입력 데이터로 수신하고, 상기 수신한 입력 데이터를 신경망 기반 특징 추출 방법을 이용하여 잡음 신호를 추정하며, 상기 입력 데이터에서 상기 추정된 잡음 신호를 뺄셈 연산으로 불필요한 신호 성분을 제거하여 음향 신호의 특징을 강화하는 특징 강화부; 및

상기 강화된 음향 신호의 특징을 컨볼루션 레이어(Convolutional Layer)와 완전 연결 레이어(Fully Connected Layer, FCL)로 이루어진 분류 알고리즘을 이용하여 출력 레이어에서 레이블 수만큼의 결과를 산출하고, 상기 산출한 결과에서 가장 높은 점수를 가진 레이블을 최종 결과로 도출하는 동물 종 식별부를 포함하는 것을 특징으로 하는 동물 종 식별 장치.

청구항 2

제1항에 있어서,

상기 동물 울음소리 구간 검출부는 상기 1차원 음향 신호를 STFT(Short Time Fourier Transform) 변환을 통해 특정 주파수로 샘플링된 음향 신호의 특정 구간을 계산하고, 상기 계산한 음향 신호의 특정 구간을 멜 필터 뱅크(Mel Filter Bank)를 통과시킨 후, 로그 연산을 수행하여 상기 2차원 음향 특징인 log-mel 스펙트로그램으로 변환하며, 이 중 이웃한 음향 특징 벡터들을 선택하는 음향 특징 추출부를 더 포함하는 것을 특징으로 하는 동물 종 식별 장치.

청구항 3

제2항에 있어서,

상기 동물 울음소리 구간 검출부는 상기 음향 특징 추출부로부터 상기 음향 특징 벡터들을 입력받아 주파수 축 방향으로 32개의 1×1 크기의 필터인 컨볼루션 레이어(Convolutional Layer)로 컨볼루션(Convolution) 연산을 수행하고, 64개의 5×1 크기의 필터인 컨볼루션 레이어로 dilated 컨볼루션 연산을 수행하며, residual connection을 위해 다시 32개의 1×1 크기의 필터인 컨볼루션 레이어로 컨볼루션 연산을 수행하여 특징을 추출하는 컨볼루션 연산부를 더 포함하는 것을 특징으로 하는 동물 종 식별 장치.

청구항 4

제3항에 있어서,

상기 동물 울음소리 구간 검출부는 상기 컨볼루션 연산부로부터 수신한 컨볼루션 연산 결과에 0에서 1 사이의 값을 갖는 시그모이드(Sigmoid) 함수를 적용하여 동물 울음소리 주파수 대역이 개선된 음향 특징이 추출하는 음향 특징 개선부를 더 포함하는 것을 특징으로 하는 동물 종 식별 장치.

청구항 5

제4항에 있어서,

상기 동물 울음소리 구간 검출부는 RNN(Recurrent Neural Network) 모델의 LSTM(Long Short-Term Memory)과 어텐션(Attention) 레이어로 구성되고, 상기 음향 특징 개선부로부터 추출된 음향 특징이 입력되면, 상기 추출한 음향 특징 중 어느 음향 특징에 동물 울음소리가 존재하는지 하기의 수학적 식 1에 따라 어텐션(Attention) 정보를 계산하는 어텐션 모듈부를 더 포함하는 것을 특징으로 하는 동물 종 식별 장치.

[수학식 1]

$$attention_i = \frac{\exp(a^T \tanh(W_x' x'_i + W_h h))}{\sum_{j=1}^7 \exp(a^T \tanh(W_x' x'_j + W_h h))}$$

여기서, a , W_x' 와 W_h 는 Attention Layer의 학습 파라미터이며, h 는 LSTM Layer의 내부 state, x'_i 는 개선된 i 번째 음향 특징임.

청구항 6

제5항에 있어서,

상기 동물 울음소리 구간 검출부는 상기 어텐션 모듈부에서 계산된 어텐션 정보를 하나의 출력 노드를 갖는 2개의 완전 연결 계층부(Fully Connected Layer, FCL)와 Sigmoid 활성화 함수에 입력하여 최종적으로 해당 구간에 동물 울음소리가 존재할 확률을 도출하는 최종 확률 계산부를 더 포함하는 것을 특징으로 하는 동물 종 식별 장치.

청구항 7

제6항에 있어서,

상기 최종 확률 계산부로부터 동물 울음소리 구간이 검출된 음향 신호를 수신하고, 상기 수신한 음향 신호에서 하기의 수학식 2와 수학식 3에 의해 사후 신호 대 잡음비(Posterior Signal to Noise Ratio, SNR)와 사전 신호 대 잡음비(Prior SNR)와 사전 SNR을 계산하고, 계산된 사전 SNR을 하기의 수학식 4의 이득 함수에 대입하여 스펙트럼 이득을 계산하고, 상기 수신한 음향 신호의 각 주파수 성분에 상기 스펙트럼 이득을 곱셈 연산을 수행하면, 잡음 신호를 제거하는 음향 품질 개선부를 더 포함하는 것을 특징으로 하는 동물 종 식별 장치.

[수학식 2]

$$\gamma_k = \frac{Y_k^2}{\lambda_d(k)}$$

[수학식 3]

$$\xi_k = 0.98 \times \xi_{k,old} + 0.02 \times \max[\gamma_k - 1, 0]$$

여기서, γ_k 는 k 번째 주파수 성분에서 계산된 사후 SNR, Y_k^2 는 음향 신호의 k 번째 주파수 파워 스펙트럼, $\lambda_d(k)$ 는 잡음 신호의 k 번째 주파수 파워 스펙트럼, ξ_k 는 k 번째 주파수 성분에서 추정된 사전 SNR, $\xi_{k,old}$ 는 이전 시간에 추정된 사전 SNR임.

[수학식 4]

$$gain_k = \frac{\xi_k}{1 + \xi_k}$$

여기서, $gain_k$ 는 스펙트럼 이득, ξ_k 는 k 번째 주파수 성분에서 추정된 사전 SNR임.

청구항 8

제7항에 있어서,

상기 특징 강화부는 상기 음향 품질 개선부로부터 동물 울음소리가 존재하는 입력 스펙트로그램을 수신하고, 상기 수신한 입력 스펙트로그램을 RNN(Recurrent Neural Network) 모델의 LSTM(Long Short-Term Memory)을 이용하여 상기 입력 스펙트로그램의 각 축 별로 특징을 추출하며, 상기 추출한 특징을 잡음 성분과 잡음이 아닌 성분을 구분한 스펙트로그램으로 계산하는 것을 특징으로 하는 동물 종 식별 장치.

청구항 9

제8항에 있어서,

상기 특징 강화부는 상기 계산한 스펙트로그램에 CNN(Convolutional Neural Network) 기반의 하나 이상의 인코더 레이어와 하나 이상의 디코더 레이어를 구성하여 고차원의 입력 데이터를 저차원으로 압축한 후, 다시 원래 데이터로 복원하는 오토 엔코더(Auto Encoder) 알고리즘을 적용하여 도메인 적응을 수행하는 것을 특징으로 하는 동물 종 식별 장치.

청구항 10

제9항에 있어서,

상기 특징 강화부는 상기 수신한 입력 스펙트로그램에서 상기 도메인 적응을 수행한 결과인 잡음 특징을 뺄셈 연산으로 불필요한 성분을 제거한 후, 비선형 활성화 함수인 ReLU(Recified Linear Unit)을 적용하고, 0에서 1 범위의 min-max 정규화(Normalization)를 통해 특징의 크기를 조절하여 음향 신호의 특징을 강화하는 것을 특징으로 하는 동물 종 식별 장치.

청구항 11

제10항에 있어서,

상기 동물 종 식별부는 상기 특징 강화부로부터 강화된 음향 신호의 특징이 입력되면, 5개의 컨볼루션 레이어(Convolutional Layer)와 2개의 완전 연결 레이어(Fully Connected Layer, FCL)로 이루어진 CNN 기반 분류 알고리즘을 이용하여 출력 레이어에서 레이블 수만큼의 결과를 산출하고, 상기 산출된 결과를 하기의 수학적 5의 소프트맥스(Softmax) 함수를 통하여 점수(Score)로 나타내며, 가장 높은 점수를 가진 레이블(Label)을 상기 최종 결과($S(y_i)$)로 분류하는 것을 특징으로 하는 동물 종 식별 장치.

[수학적 5]

$$S(y_i) = \frac{\exp(y_i)}{\sum_{j=1}^N \exp(y_j)}$$

청구항 12

동물 울음소리의 음향 신호인 1차원 음향 신호를 입력받고, 상기 1차원 음향 신호를 주파수 축과 시간 축을 가지는 2차원 음향 특징인 log-mel 스펙트로그램으로 변환하며, 상기 변환된 2차원 음향 특징을 소리 구간 검출을 통해 상기 동물 울음소리가 존재하는 구간을 검출하는 단계;

상기 검출한 동물 울음소리가 존재하는 구간을 입력 데이터로 수신하고, 상기 수신한 입력 데이터를 신경망 기반 특징 추출 방법을 이용하여 잡음 신호를 추정하며, 상기 입력 데이터에서 상기 추정한 잡음 신호를 뺄셈 연산으로 불필요한 신호 성분을 제거하여 음향 신호의 특징을 강화하는 단계; 및

상기 강화된 음향 신호의 특징을 컨볼루션 레이어(Convolutional Layer)와 완전 연결 레이어(Fully Connected Layer, FCL)로 이루어진 분류 알고리즘을 이용하여 출력 레이어에서 레이블 수만큼의 결과를 산출하고, 상기 산출한 결과에서 가장 높은 점수를 가진 레이블을 최종 결과로 도출하는 단계를 포함하는 것을 특징으로 하는 동물 종 식별 방법.

청구항 13

제12항에 있어서,

상기 동물 울음소리가 존재하는 구간을 검출하는 단계는,

상기 1차원 음향 신호를 STFT(Short Time Fourier Transform) 변환을 통해 특정 주파수로 샘플링된 음향 신호의 특정 구간을 계산하고, 상기 계산한 음향 신호의 특정 구간을 멜 필터 뱅크(Mel Filter Bank)를 통과시킨 후, 로그 연산을 수행하여 상기 2차원 음향 특징인 log-mel 스펙트로그램으로 변환하며, 이 중 이웃한 음향 특징 벡터들을 선택하는 단계를 더 포함하는 것을 특징으로 하는 동물 종 식별 방법.

청구항 14

제13항에 있어서,

상기 동물 울음소리가 존재하는 구간을 검출하는 단계는,

RNN(Recurrent Neural Network) 모델의 LSTM(Long Short-Term Memory)과 어텐션(Attention) 레이어로 구성되고, 상기 음향 특징 개선부로부터 추출된 음향 특징이 입력되면, 상기 추출한 음향 특징 중 어느 음향 특징에 동물 울음소리가 존재하는지 하기의 수학적 식 1에 따라 어텐션(Attention) 정보를 계산하는 단계를 더 포함하는 것을 특징으로 하는 동물 종 식별 방법.

[수학적 식 1]

$$attention_i = \frac{\exp(a^T \tanh(W_x' x'_i + W_h h))}{\sum_{j=1}^7 \exp(a^T \tanh(W_x' x'_j + W_h h))}$$

여기서, a , W_x' 와 W_h 는 Attention Layer의 학습 파라미터이며, h 는 LSTM Layer의 내부 state, x'_i 는 개선된 i 번째 음향 특징임.

청구항 15

제14항에 있어서,

상기 계산된 어텐션 정보를 하나의 출력 노드를 갖는 2개의 완전 연결 계층부(Fully Connected Layer, FCL)와 Sigmoid 활성화 함수에 입력하여 최종적으로 해당 구간에 동물 울음소리가 존재할 확률을 도출하는 단계를 더 포함하는 것을 특징으로 하는 동물 종 식별 방법.

청구항 16

제15항에 있어서,

상기 음향 신호의 특징을 강화하는 단계는,

상기 동물 울음소리가 존재하는 입력 스펙트로그램을 수신하고, 상기 수신한 입력 스펙트로그램을 RNN(Recurrent Neural Network) 모델의 LSTM(Long Short-Term Memory)을 이용하여 상기 입력 스펙트로그램의 각 축 별로 특징을 추출하며, 상기 추출한 특징을 잡음 성분과 잡음이 아닌 성분을 구분한 스펙트로그램으로 계산하는 단계를 더 포함하는 것을 특징으로 하는 동물 종 식별 방법.

청구항 17

제16항에 있어서,

상기 계산한 스펙트로그램에 CNN(Convolutional Neural Network) 기반의 하나 이상의 인코더 레이어와 하나 이상의 디코더 레이어를 구성하여 고차원의 입력 데이터를 저차원으로 압축한 후, 다시 원래 데이터로 복원하는 오토 엔코더(Auto Encoder) 알고리즘을 적용하여 도메인 적응을 수행하는 단계; 및

상기 수신한 입력 스펙트로그램에서 상기 도메인 적응을 수행한 결과인 잡음 특징을 뺄셈 연산으로 불필요한 성분을 제거한 후, 비선형 활성화 함수인 ReLU(Rectified Linear Unit)를 적용하고, 0에서 1 범위의 min-max 정규화(Normalization)를 통해 특징의 크기를 조절하여 음향 신호의 특징을 강화하는 단계를 더 포함하는 것을 특징으로 하는 동물 종 식별 방법.

청구항 18

제17항에 있어서,

상기 최종 결과로 도출하는 단계는,

상기 강화된 음향 신호의 특징이 입력되면, 5개의 컨볼루션 레이어(Convolutional Layer)와 2개의 완전 연결 레이어(Fully Connected Layer, FCL)로 이루어진 CNN 기반 분류 알고리즘을 이용하여 출력 레이어에서 레이블 수 만큼의 결과를 산출하고, 상기 산출된 결과를 하기의 수학적 2의 소프트맥스(Softmax) 함수를 통하여 점수(Score)로 나타내며, 가장 높은 점수를 가진 레이블(Label)을 상기 최종 결과(S(yi))로 분류하는 단계를 더 포함하는 것을 특징으로 하는 동물 종 식별 방법.

[수학적식 2]

$$S(y_i) = \frac{\exp(y_i)}{\sum_{j=1}^N \exp(y_j)}$$

발명의 설명

기술 분야

[0001] 본 발명은 동물 종 식별 장치에 관한 것으로서, 더욱 상세하게는 다양한 잡음에도 동물 종들의 울음소리에 기초하여 동물 울음소리 구간을 검출하고, 강화된 음향 특징을 학습하여 동물 종을 식별할 수 있는 잡음 환경에 강한 동물 종 식별 장치 및 방법에 관한 것이다.

배경 기술

[0002] 현재까지 환경 음원의 자동 분류에 관한 연구는 여러 해 동안 꾸준히 진행되어 왔다. 음원에 대한 자동 분류는 직접적 또는 간접적으로 음성 인식, 패턴 인식, 상황 감지 또는 상황 인식 등의 다양한 분야에 적용될 수 있기 때문에 그 중요성이 더 커져가고 있다.

[0003] 공공 사업과 같은 철도, 전기, 가스, 수도 사업 등은 대규모 사업으로 자연 환경에 미치는 영향이 크기 때문에 환경 영향 평가를 시행하여 지역 환경에 미칠 수 있는 부정적인 요소를 미리 파악 및 분석하여 최소한의 피해를 줄 수 있는 수단을 취해야 한다.

[0004] 환경 영향 평가에서 생물 종을 조사하는 활동은 해당 지역에 서식하는 생물들의 다양성을 보전하고 생태계를 유지하는데 필수적으로 필요한 활동이다.

[0005] 생물 종을 식별하기 위해서는 실제로 생물을 채집하거나 생물의 울음소리를 듣고 분별할 수 있지만, 사람이 특정 지역 내의 모든 생물 종을 직접 분별하는 것이 불가능하다.

[0006] 실외 환경은 실내 환경에서 생물의 울음소리 정보를 수집하는 것과 다르게 잡음이 많은 환경이기 때문에 생물의 울음소리만을 습득하는 것이 어려운 일이다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 한국 공개특허번호 제10-2014-0122881호

발명의 내용

해결하려는 과제

[0008] 이와 같은 문제점을 해결하기 위하여, 본 발명은 동물 종들의 울음소리에 기초하여 동물 울음소리 구간을 검출하고, 동물 울음소리가 존재하지 않는 구간을 바탕으로 잡음을 추정하여 수신된 소리에서 잡음을 제거한 후, 다양한 잡음에도 강화된 음향 특징을 학습하여 동물 종을 식별할 수 있는 잡음 환경에 강인한 동물 종 식별 장치 및 방법을 제공하는 것을 목적으로 한다.

[0009] 또한, 본 발명은 다양한 잡음이 많이 발생하는 환경에서 획득된 동물 종 식별 관련 음향 신호이더라도 좋은 분류 성능을 낼 수 있는 잡음에 강인한 인공지능 기반 알고리즘과 시스템에 개발을 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0010] 상기 목적을 달성하기 위한 본 발명의 특징에 따른 잡음 환경에 강인한 동물 종 식별 장치는,

[0011] 동물 울음소리의 음향 신호인 1차원 음향 신호를 입력받고, 상기 1차원 음향 신호를 주파수 축과 시간 축을 가지는 2차원 음향 특징인 log-mel 스펙트로그램으로 변환하며, 상기 변환된 2차원 음향 특징을 소리 구간 검출을 통해 상기 동물 울음소리가 존재하는 구간을 검출하는 동물 울음소리 구간 검출부;

[0012] 상기 검출한 동물 울음소리가 존재하는 구간을 입력 데이터로 수신하고, 상기 수신한 입력 데이터를 신경망 기반 특징 추출 방법을 이용하여 잡음 신호를 추정하며, 상기 입력 데이터에서 상기 추정한 잡음 신호를 뺄셈 연산으로 불필요한 신호 성분을 제거하여 음향 신호의 특징을 강화하는 특징 강화부; 및

[0013] 상기 강화된 음향 신호의 특징을 컨볼루션 레이어(Convolutional Layer)와 완전 연결 레이어(Fully Connected Layer, FCL)로 이루어진 분류 알고리즘을 이용하여 출력 레이어에서 레이블 수만큼의 결과를 산출하고, 상기 산출한 결과에서 가장 높은 점수를 가진 레이블을 최종 결과로 도출하는 동물 종 식별부를 포함하는 것을 특징으로 한다.

[0015] 상기 동물 울음소리 구간 검출부는 상기 1차원 음향 신호를 STFT(Short Time Fourier Transform) 변환을 통해 특정 주파수로 샘플링된 음향 신호의 특정 구간을 계산하고, 상기 계산한 음향 신호의 특정 구간을 멜 필터 뱅크(Mel Filter Bank)를 통과시킨 후, 로그 연산을 수행하여 상기 2차원 음향 특징인 log-mel 스펙트로그램으로 변환하며, 이 중 이웃한 음향 특징 벡터들을 선택하는 음향 특징 추출부를 더 포함하는 것을 특징으로 한다.

[0017] 상기 동물 울음소리 구간 검출부는 상기 음향 특징 추출부로부터 상기 음향 특징 벡터들을 입력받아 주파수 축 방향으로 32개의 1×1 크기의 필터인 컨볼루션 레이어(Convolutional Layer)로 컨볼루션(Convolution) 연산을 수행하고, 64개의 5×1 크기의 필터인 컨볼루션 레이어로 diated 컨볼루션 연산을 수행하며, residual connection을 위해 다시 32개의 1×1 크기의 필터인 컨볼루션 레이어로 컨볼루션 연산을 수행하여 특징을 추출하는 컨볼루션 연산부를 더 포함하는 것을 특징으로 한다.

[0019] 상기 동물 울음소리 구간 검출부는 상기 컨볼루션 연산부로부터 수신한 컨볼루션 연산 결과에 0에서 1 사이의 값을 갖는 시그모이드(Sigmoid) 함수를 적용하여 동물 울음소리 주파수 대역이 개선된 음향 특징이 추출하는 음향 특징 개선부를 더 포함하는 것을 특징으로 한다.

[0021] 본 발명의 특징에 따른 잡음 환경에 강인한 동물 종 식별 방법은,

[0022] 동물 울음소리의 음향 신호인 1차원 음향 신호를 입력받고, 상기 1차원 음향 신호를 주파수 축과 시간 축을 가지는 2차원 음향 특징인 log-mel 스펙트로그램으로 변환하며, 상기 변환된 2차원 음향 특징을 소리 구간 검출을

통해 상기 동물 울음소리가 존재하는 구간을 검출하는 단계;

[0023] 상기 검출한 동물 울음소리가 존재하는 구간을 입력 데이터로 수신하고, 상기 수신한 입력 데이터를 신경망 기반 특징 추출 방법을 이용하여 잡음 신호를 추정하며, 상기 입력 데이터에서 상기 추정한 잡음 신호를 뺄셈 연산으로 불필요한 신호 성분을 제거하여 음향 신호의 특징을 강화하는 단계; 및

[0024] 상기 강화된 음향 신호의 특징을 컨볼루션 레이어(Convolutional Layer)와 완전 연결 레이어(Fully Connected Layer, FCL)로 이루어진 분류 알고리즘을 이용하여 출력 레이어에서 레이블 수만큼의 결과를 산출하고, 상기 산출한 결과에서 가장 높은 점수를 가진 레이블을 최종 결과로 도출하는 단계를 포함하는 것을 특징으로 한다.

발명의 효과

[0025] 전술한 구성에 의하여, 본 발명은 잡음 환경에 강인한 종 식별 알고리즘 및 시스템을 적용하면 잡음이 많은 환경에서 얻은 신호이더라도 생물 종 식별 성능을 통해 정확한 생태계 환경을 파악할 수 있는 효과가 있다.

[0026] 본 발명은 생물 종 식별 뿐만 아니라 음성 데이터를 활용한 알고리즘 및 시스템에 적용하여 어떠한 잡음 환경에서도 개선된 성능을 낼 수 있는 효과가 있다.

[0027] 본 발명은 수집된 데이터들로부터 정확한 생태계 환경 정보를 바탕으로 해당 지역의 생물들의 고유 특성을 파악할 수 있으며, 환경 개선 및 유지를 통한 생물들을 보전할 수 있는 효과가 있다.

도면의 간단한 설명

[0028] 도 1은 본 발명의 실시예에 따른 잡음 환경에 강인한 동물 종 식별 장치의 구성을 나타낸 블록도이다.

도 2는 본 발명의 실시예에 따른 동물 울음소리 구간 검출부의 내부 구성을 간략하게 나타낸 블록도이다.

도 3은 본 발명의 실시예에 따른 특징 강화부에서 이루어지는 특징 강화 과정을 나타낸 도면이다.

도 4는 본 발명의 실시예에 따른 잡음 성분 검출 과정을 나타낸 도면이다.

도 5는 본 발명의 실시예에 따른 도메인 적응 과정을 나타낸 도면이다.

도 6은 본 발명의 실시예에 따른 강화된 음향 신호 특징을 인식하기 위한 CNN 구조도를 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

[0029] 이하 도면을 참조하여 본 발명의 바람직한 실시예를 설명함으로써 본 발명을 상술한다.

[0030] 본 발명의 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

[0031] 도 1은 본 발명의 실시예에 따른 잡음 환경에 강인한 동물 종 식별 장치의 구성을 나타낸 블록도이고, 도 2는 본 발명의 실시예에 따른 동물 울음소리 구간 검출부의 내부 구성을 간략하게 나타낸 블록도이고, 도 3은 본 발명의 실시예에 따른 특징 강화부에서 이루어지는 특징 강화 과정을 나타낸 도면이고, 도 4는 본 발명의 실시예에 따른 잡음 성분 검출 과정을 나타낸 도면이고, 도 5는 본 발명의 실시예에 따른 도메인 적응 과정을 나타낸 도면이고, 도 6은 본 발명의 실시예에 따른 강화된 음향 신호 특징을 인식하기 위한 CNN 구조도를 나타낸 도면이다.

[0032] 본 발명의 실시예에 따른 잡음 환경에 강인한 동물 종 식별 장치(100)는 동물 울음소리 구간을 검출하는 전처리부(110) 및 잡음 성분 검출과 도메인 적응을 수행하는 분류기(120)를 포함한다.

[0033] 동물 종 식별 장치(100)는 동물 종들의 울음소리에 기초하여 수신된 소리에서 동물 울음소리 구간을 검출하고, 동물 울음소리가 존재하지 않는 구간을 바탕으로 잡음을 추정하여 수신된 소리에서 잡음을 제거한 후, 다양한 잡음에도 강화된 음향 특징을 학습하여 동물 종을 식별할 수 있다.

[0034] 전처리부(110)는 소리를 분류하기 전에 수행되는 작업으로 음향 특징 추출, 소리 구간 검출과 잡음 제거로 구성된다.

[0035] 전처리부(110)는 동물 울음소리의 구간 검출 기능을 수행하는 것으로 음향 신호 입력부(111), 동물 울음소리 구간 검출부(112) 및 음향 품질 개선부(113)를 포함한다.

- [0036] 동물 울음소리 구간 검출부(112)는 입력된 음향 신호에서 동물 울음소리가 존재하는 구간을 검출하기 위해 NAVAD(Neural Attentive Voice Activity Detection) 알고리즘을 적용한다.
- [0037] 동물 울음소리 구간 검출부(112)는 음향 특징 추출부(112a), 컨볼루션 연산부(112b), 음향 특징 개선부(112c), 어텐션 모듈부(112d) 및 최종 확률 계산부(112e)를 포함하며, 이러한 구성 모듈은 NAVAD 알고리즘의 구조도라고 볼 수 있다.
- [0038] 음향 신호 입력부(111)는 탐지 대상이 되는 동물 울음소리의 음향 신호를 입력받는다.
- [0039] 음향 특징 추출부(112a)는 1차원 음향 신호를 주파수 축과 시간 축을 가지는 2차원 음향 특징인 log-mel 스펙트로그램으로 변환하는 과정이다.
- [0040] 음향 특징 추출부(112a)는 음향 신호 입력부(111)로부터 1차원 음향 신호를 입력받고, 1차원 음향 신호를 STFT(Short Time Fourier Transform) 변환과, 멜 필터 뱅크부(Mel Filter Bank)를 통과시킨 후, 로그 연산을 수행하여 2차원 음향 특징인 log-mel 스펙트로그램으로 변환한다.
- [0041] 음향 특징 추출부(112a)는 STFT 변환 모듈 및 멜 필터 뱅크부를 포함한다.
- [0042] STFT 변환 모듈은 음향 신호를 일정한 길이의 윈도우(Window)를 이용하여 세분화한 후, 음향 신호의 주파수를 분석할 수 있다.
- [0043] 멜 필터 뱅크부는 STFT 변환 모듈에서 분석한 주파수를 이용하여 MFCC(Mel Frequency Cepstral Coefficient) 방식에 의하여 입력된 음향 신호의 특징값을 추출할 수 있다.
- [0044] 음향 특징 추출부(112a)는 2차원 음향 특징인 log-mel 스펙트로그램으로 변환한 후, 이 중 이웃한 7개의 음향 특징 벡터들을 선택하여 컨볼루션 연산부(112b)로 전송한다.
- [0045] 다시 말해, 음향 특징 추출부(112a)는 STFT 변환을 통해 특정 주파수로 샘플링된 음향 신호의 특정 구간을 계산하고, 계산한 음향 신호의 특정 구간을 멜 필터 뱅크(Mel Filter Bank)를 통과시켜 특정 주파수 밴드의 멜 필터 뱅크값(Mel Filterbank)을 구하며, 구해진 멜 필터 뱅크값에 로그 연산을 수행하여 2차원 음향 특징인 log-mel 스펙트로그램으로 변환된다.
- [0046] 컨볼루션 연산부(112b)는 입력 데이터와 가중치들의 집합체인 필터(Filter)와의 컨볼루션 연산을 통해 입력 데이터의 특징(Feature)을 추출하는 기능을 수행한다.
- [0047] 컨볼루션 연산부(112b)는 7개의 음향 특징 벡터들을 입력받는데, 즉 하기의 수학적 식 1과 같이 입력된 음향 특징에 주파수 축 방향으로 32개의 1×1 크기의 필터인 컨볼루션 레이어(Convolutional Layer)로 컨볼루션(Convolution) 연산을 수행한 후, 비선형 활성화 함수인 ReLU(Rectified Linear Unit)를 적용한다.

수학적 식 1

$$y_{i,j} = \max(x_i * w_{j,pre}, 0)$$

- [0048]
- [0049] 여기서, $\mathbf{x}_i$ 는 i번째 음향 특징 벡터, $w_{j,pre}$ 는 1×1 크기를 갖는 j번째 필터이며, $y_{i,j}$ 는 i번째 입력 $\mathbf{x}_i$ 과 j번째 필터 $w_{j,pre}$ 의 컨볼루션(Convolution) 연산을 통해 나온 결과값이다.

- [0050] 그런 다음, 컨볼루션 연산부(112b)는 64개의 5×1 크기의 필터인 컨볼루션 레이어로 diated 컨볼루션 연산을 수행하며, 앞서 추출한 특징과의 residual connection을 위해 다시 32개의 1×1 크기의 필터인 컨볼루션 레이어로 컨볼루션 연산을 수행한다. 각각의 컨볼루션 연산 뒤에는 비선형 활성화 함수 ReLU를 적용한다.
- [0051] 해당 컨볼루션 연산은 하기의 수학적 식 2와 같다.

수학식 2

$$z_{i,l} = \max(\max(y_{i,j} * f_{j,k}, 0) * w_{k,l,post}, 0) + y_{i,j}$$

여기서, $f_{j,k}$ 는 5×1 크기를 갖는 j 번째 입력 필터에 상응하는 k 번째 필터이며, $w_{k,l,post}$ 는 5×1 크기를 갖는 k 번째 입력 필터에 상응하는 l 번째 필터이다.

네트워크에 사용되는 입력과 출력의 차원을 맞추기 위해 1×1 크기의 필터로 컨볼루션 연산이 하기의 수학식 3과 같이 최종적으로 수행된다.

수학식 3

$$v_i = z_{i,l} * w_{l,match}$$

여기서, $w_{l,match}$ 는 1×1 크기를 갖는 l 번째 필터이고, v_i 는 입력 x_i 와 크기와 차원이 같다.

음향 특징 개선부(112c)는 컨볼루션 연산부(112b)로부터 수신한 컨볼루션 연산 결과에 0에서 1 사이의 값을 갖는 시그모이드(Sigmoid) 함수를 적용하여 동물 울음소리 주파수 대역이 개선된 음향 특징 x'_i 이 추출된다(수학식 4). 여기서, 시그모이드 함수는 입력을 0과 1 사이의 출력값으로 정규화시키는 활성화 함수이다.

수학식 4

$$x'_i = x_i \cdot \frac{1}{1 + \exp(-v_i)}$$

음향 특징 개선부(112c)는 추출된 개선된 음향 특징 x'_i 을 어텐션 모듈부(112d)에 입력한다.

어텐션 모듈부(112d)는 음향 신호를 패턴 인식 기법을 이용하여 다양한 음향 이벤트를 찾아내는 기법으로 재발 신경망(Recurrent Neural Network, RNN) 모델의 LSTM(Long Short-Term Memory)과 Attention Layer로 구성된다.

패턴 인식 기법은 인공 신경망을 이용한 예측 방법으로 입력층으로부터 출력층의 결과값을 예측한 경우, 학습 과정에서 결과값들로부터 입력값을 예측할 수 있다. 인공 신경망은 입력값과 출력값이 일대일 대응 관계에 있지 아니하므로, 출력층으로서 입력층을 그대로 복구하는 것은 불가능하나, 예측 알고리즘을 고려하여 역전파(Backpropagation, Backpropa) 알고리즘에 의해 결과값으로부터 산출된 출력 데이터가 최초의 입력 데이터와 상이하다면, 인공 신경망의 예측이 부정확하다고 볼 수 있으므로, 제약 조건 하에서 산출된 출력 데이터가 최초의 입력 데이터와 유사해지도록 예측 계수를 변경하여 학습을 훈련하게 된다.

심층 신경망이란 신경망 알고리즘 중에서 여러 개의 층으로 이루어진 신경망을 의미한다. 한 층은 여러 개의 노드로 이루어져 있고, 노드에서 실제 연산이 이루어지는데, 이러한 연산 과정은 인간의 신경망을 구성하는 뉴런에서 일어나는 과정을 모사하도록 설계되어 있다. 통상적인 인공 신경망은 입력층(Input Layer), 은닉층(Hidden Layer), 출력층(Output Layer)로 나뉘며, 입력 데이터는 입력층의 입력이 되며, 입력층의 출력은 은닉층의 입력이 되고, 은닉층의 출력은 출력층의 입력이 되고, 출력층의 출력이 최종 출력이 된다.

어텐션 모듈부(112d)는 입력층으로부터 입력 데이터를 입력받아 예측값을 출력층의 버퍼에 출력하는 예측 심층 신경망을 사용하며, 예측 심층 신경망의 구조나 형태는 제한되지 않고, 대표적인 방법으로 DNN(Deep Neural Network), CNN(Convolutional Neural Network), RNN(Recurrent Neural Network) 등이 있으며, 각각의 신경망의

조합으로 예측 심층 신경망을 구성하여 다양한 구조의 심층 신경망을 구성할 수 있다. 본 발명의 어텐션 모듈부(112d)는 예측 심층 신경망의 구조로 RNN을 사용한다.

- [0064] RNN은 입력으로 T 프레임의 40차의 ($T \times 40$) 로그 멜 필터 뱅크값이 사용되고, 그 전체적인 구조가 256개의 유닛을 가진 3개의 GRU(Gated Recurrent Unit)층과 4개층의 FNN, 그리고 마지막에 Sigmoid 함수를 가진 출력층으로 구성되어 있다.
- [0065] 여기서, FNN(Feedforward Neural Network)은 40차의 로그 멜 필터 뱅크값의 특징은 연속적인 5개의 프레임들이 합쳐져서 200차원의 특징 벡터를 형성하고 이들은 FNN의 입력으로 사용된다. 2개의 은닉층은 각각 Relu 활성화 함수를 가지는 1600개의 유닛들로 구성된다.
- [0066] 출력층은 분류하고자 하는 음향 이벤트 클래스의 수만큼의 유닛을 가지며, 각 유닛은 Sigmoid 활성화 함수를 이용한다. 여기서, Sigmoid 활성화 함수의 출력값은 각 클래스에 대한 사후 확률로 간주되며, 0.5의 기준치와 비교하여 이진화된 후, Ground Truth Table과 비교하여 FNN의 정확도 산출을 위하여 사용된다.
- [0067] RNN 모델은 음성, 문장 등의 순차적인 데이터를 처리하는 신경망 구조이다. RNN의 메모리라 할 수 있는 state는 입력 데이터가 들어올 때마다 반복적으로 갱신이 되어 처음부터 현재 시간 t까지의 입력 데이터를 요약한 정보를 가지고 있다.
- [0068] 결과적으로 입력을 모두 처리한 후, RNN의 state는 입력 데이터의 전체를 요약하는 정보가 된다. 실제 모델 훈련 시 t에서의 state에서는 t에서의 입력 데이터와 t-1의 state 값으로 t의 결과값과 state값을 계산한다.
- [0069] RNN은 Back-Propagation 기반의 Gradient Descent 방법을 이용하여 비용 함수를 최소화하며, 최적의 변수값을 찾는다.
- [0070] LSTM은 RNN의 가중치 대신 망각 게이트를 이용하여 결과값을 예측하는 RNN 방법의 일종을 의미한다. 시계열적인 입력 데이터에 대한 예측에 있어서 순차적으로 데이터를 처리할 때, RNN 방식으로 지난 데이터를 처리하는 경우, 오래된 데이터의 경우는 가중치에 따라 감소되어 일정 단계를 넘게 되면 그 값이 0이 되어 가중치와 관계없이 더 이상 반영하지 않는 문제가 있다.
- [0071] LSTM의 경우 곱셈 대신 덧셈을 사용하므로, Recurrent 입력값이 0이 되지 않는 장점이 있게 된다.
- [0072] 어텐션 모듈부(112d)는 음향 특징 개선부(112c)로부터 추출된 음향 특징이 입력되면, 서로 이웃한 7개의 음향 특징 중 어느 음향 특징에 동물 울음소리가 존재하는지 하기의 수학적 식 5에 따라 어텐션(Attention) 정보를 확률로 계산한다.

수학적 식 5

$$attention_i = \frac{\exp(a^T \tanh(W_x' x'_i + W_h h))}{\sum_{j=1}^7 \exp(a^T \tanh(W_x' x'_j + W_h h))}$$

- [0073]
- [0074] 여기서, a, W_x' 와 W_h 는 Attention Layer의 학습 파라미터이며, h 는 LSTM Layer의 내부 state, x'_i 는 개선된 i번째 음향 특징이다.
- [0075] 어텐션 모듈부(112d)는 수학적 식 5를 통해 계산된 어텐션 정보와 개선된 음향 특징 x'_i 을 가중 평균 계산을 수행하여 하나의 벡터를 계산한다(수학적 식 6).

수학식 6

$$c = \sum_{i=1}^7 \text{attention}_i \cdot \mathbf{x}'_i$$

[0076]

[0077] 어텐션 모듈부(112d)는 계산된 하나의 벡터 c 를 최종 확률 계산부(112e)로 전송한다.

[0078] RNN과 같이 순차적인 데이터를 처리하는 신경망 구조는 시간이 길어질수록 초기에 저장된 정보가 사라지는 단점이 있는데, 이를 보완하기 위해서 Attention 알고리즘을 사용한다. Attention 알고리즘은 순차적인 입력 데이터를 RNN으로 처리하여 모든 시간의 state 벡터 h 를 출력한다. 이때, 모든 시간의 벡터를 같은 비율로 참고하는 것이 아니라 특정 시간에 Attention을 주어 최종 결과를 도출한다.

[0079] 최종 확률 계산부(112e)는 어텐션 모듈부(112d)에서 계산된 어텐션 정보를 하나의 출력 노드를 갖는 2개의 완전 연결 계층부(Fully Connected Layer, FCL)와 Sigmoid 활성화 함수에 입력하여 최종적으로 해당 구간에 동물 울음소리가 존재할 확률을 도출한다(수학식 7).

수학식 7

$$p(\text{animal sound} | \mathbf{x}_1, \dots, \mathbf{x}_7) = \frac{1}{1 + \exp(-W_{fc2} \max(W_{fc1} \mathbf{u}, 0))}$$

[0080]

[0081] 입력된 음향 신호의 모든 구간에 대하여 수학식 1 내지 수학식 7와 같은 연산을 수행한다.

[0082] 음향 신호에서 동물 울음소리가 존재하지 않는 부분은 잡음 제거 단계를 적용하며, 동물 울음소리가 존재하는 부분은 특징 강화 단계를 적용한다.

[0083] 음향 품질 개선부(113)는 동물 울음소리 구간 검출부(112)로부터 동물 울음소리 구간이 검출된 음향 신호를 수신하고, 수신한 음향 신호에서 잡음 신호를 제거하여 음향 신호의 품질을 개선한다.

[0084] 잡음 신호는 시간에 따라 변하기 때문에 동물 울음소리가 존재하지 않을 때마다 잡음의 분산을 갱신하여 사후 신호 대 잡음비(Posterior Signal to Noise Ratio, SNR)와 사전 신호 대 잡음비(Prior SNR)를 추정한다.

[0085] 잡음 제거 기술은 신호 대 잡음비에 따른 스펙트럼 이득으로 표현할 수 있다. 이에 따라 음향 품질 개선부(113)는 동물 울음소리 구간 검출부(112)로부터 동물 울음소리 구간이 검출된 음향 신호를 수신하고, 수신한 음향 신호에서 하기의 수학식 8과 수학식 9에 의해 사후 SNR과 사전 SNR을 계산한다.

[0086] 사후 SNR γ 은 잡음이 섞인 신호 대비 잡음비이고, 사전 SNR ξ 은 잡음이 섞이지 않은 신호 대비 잡음비이다.

[0087] 주어진 환경에서는 잡음이 섞이지 않은 신호를 알 수 없기 때문에 사전 SNR을 추정하는 과정이 필요하며, 해당 사후 SNR과 사전 SNR의 추정은 하기의 수학식 8과 수학식 9와 같이 산출된다.

수학식 8

$$\gamma_k = \frac{Y_k^2}{\lambda_d(k)}$$

[0088]

수학식 9

$$\xi_k = 0.98 \times \xi_{k,old} + 0.02 \times \max[\gamma_k - 1, 0]$$

여기서, γ_k 는 k번째 주파수 성분에서 계산된 사후 SNR, Y_k^2 는 음향 신호의 k번째 주파수 파워 스펙트럼, $\lambda_d(k)$ 는 잡음 신호의 k번째 주파수 파워 스펙트럼, ξ_k 는 k번째 주파수 성분에서 추정된 사전 SNR, $\xi_{k,old}$ 는 이전 시간에 추정된 사전 SNR을 나타낸다. 최초의 사전 SNR 값은 0으로 초기화한다.

음향 품질 개선부(113)는 수학식 8과 수학식 9를 이용하여 추정된 사전 SNR을 하기의 수학식 10의 이득 함수에 대입하여 스펙트럼 이득을 계산하고, 0과 1 사이의 값을 가지며, 입력 음향 신호의 각 주파수 성분에 직접 곱해진다.

수학식 10

$$gain_k = \frac{\xi_k}{1 + \xi_k}$$

여기서, $gain_k$ 는 스펙트럼 이득, ξ_k 는 k번째 주파수 성분에서 추정된 사전 SNR이다.

음향 품질 개선부(113)는 이득 함수 $Gain_k$ 를 잡음이 있는 음향 신호의 k번째 주파수 파워 스펙트럼 Y_k^2 에 직접 곱하여 잡음이 제거된 음향 신호의 k번째 주파수 파워 스펙트럼 X_k^2 를 계산하게 된다.

수학식 11

$$X_k^2 = gain_k \cdot Y_k^2$$

음향 품질 개선부(113)는 동물 울음소리 구간 검출부(112)로부터 동물 울음소리 구간이 검출된 음향 신호를 수신하고, 수신한 음향 신호의 각 주파수 성분에 스펙트럼 이득을 곱셈 연산을 수행하면, 잡음 신호를 제거할 수 있다.

음향 품질 개선부(113)는 잡음 신호를 제거한 동물 울음소리가 존재하는 구간을 잘라서 특징 강화부(121)로 전송한다.

분류기(120)는 특징 강화부(121) 및 동물 종 식별부(122)를 포함한다.

특징 강화부(121)는 음향 품질 개선부(113)로부터 동물 울음소리가 존재하는 입력 스펙트로그램을 수신하고, 상기 수신한 입력 스펙트로그램을 신경망 기반 특징 추출 방법을 통해 잡음 신호를 추정하고, 입력 데이터에서 추정된 잡음 신호를 뺄셈 연산으로 불필요한 성분을 제거하여 음향 신호의 특징을 강화한다. 스펙트로그램은 소리나 파동을 시각화하여 파악하기 위한 것으로 파형과 스펙트럼의 특징이 포함되어 있다. 파형에서는 시간 축의 변화에 따른 진폭 축의 변화를 볼 수 있고, 스펙트럼에서는 주파수 축의 변화에 따른 진폭 축의 변화를 볼 수 있다. 스펙트로그램은 시간 축과 주파수 축의 변화에 따라 진폭의 차이를 인쇄 농도와 표시 색상의 차이로 나타낸다.

- [0101] 특징 강화부(121)는 뿔샘 연산 이후에 비선형 활성화 함수 ReLU를 통해 음수 부분을 0으로 처리해주고, 정규화(Normalization) 과정을 통해 특징의 범위를 0 내지 1로 조절한다. 이는 Positive Definition을 유지하고, 뿔샘 연산으로 인한 특징의 손실을 막기 위함이다.
- [0102] 특징 강화부(121)는 RNN-LSTM과 Auto Encoder 알고리즘을 이용하여 신경망이 잡음 및 녹음 장소 등과 같은 도메인 변화에 강인하도록 학습하는 과정이다.
- [0103] 다시 말해, 특징 강화 네트워크의 학습(잡음 성분 검출 및 도메인 적응(Adaptation))은 분류기(120)와 동시에 End to End 방식으로 학습된다.
- [0104] 특징 강화부(121)는 잡음 성분 검출(도 4)을 에너지 기반으로 이루어진다.
- [0105] 특징 강화부(121)는 $R^{M \times N}$ 차원을 가지는 입력 스펙트로그램에서 특징 요소 간 에너지 크기를 학습하고, 에너지가 상대적으로 작으면서 학습에 불필요한 성분을 제거한다.
- [0106] 특징 강화부(121)는 차원 축소 및 확장을 통해 이루어지며, RNN-LSTM을 이용하여 입력 스펙트로그램의 각 축 별로 저차원 특징을 추출한다.
- [0107] 시간 축의 경우, $R^{K \times N}$ 차원을 가지며 $N > K$ 의 조건을 만족한다. 주파수 축은 $R^{M \times K}$ 차원을 가지며 $M > K$ 의 조건을 만족한다.
- [0108] 특징 강화부(121)는 하이퍼볼릭 탄젠트(Hypoblic Tangent) 함수(Tanh)인 수학식 12를 이용하여 각각의 특징 요소들을 -1에서 1의 범위로 표현하고, 비선형 활성화 함수(Rectified Linear Unit, RELU)인 수학식 13을 통해 음의 값들을 0으로 만든다. 이러한 과정을 통해 각 축 별 LSTM 출력 벡터는 항상 0에서 1 범위의 확률값을 가지게 된다.

수학식 12

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

수학식 13

$$ReLU(x) = \max(0, x)$$

- [0111] 도 4에 도시된 바와 같이, 특징 강화부(121)는 RNN-LSTM을 통해 얻은 2개의 특징을 하기의 수학식 14를 이용하여 잡음 성분과 잡음이 아닌 성분을 구분한 강화된 스펙트로그램을 계산한다.

수학식 14

$$Y(V_1, V_2, X) = (f(V_1) \otimes f(V_2)) \odot X$$

- [0113] V_1, V_2, X 는 각각 주파수축 LSTM 출력 벡터, 시간 축 LSTM 출력 벡터 및 입력 데이터를 나타낸다.

- [0114] $\otimes, \odot$ 는 각각 벡터 외적과 성분별 곱셈을 나타낸다. $f(\cdot)$ 는 수학식 12, 수학식 13을 연속적으로 적용한 함수이다.

- [0115] 각각의 LSTM 출력 벡터의 외적을 통해 얻은 $R^{M \times N}$ 차원의 행렬은 입력 스펙트로그램의 구성 요소에 0 내지

1 범위를 가지는 가중치를 주기 때문에 학습에 불필요한 성분을 검출하는 역할을 수행한다.

- [0116] 특징 강화부(121)는 잡음 성분 검출을 수행한 후, 도메인 적응(Adaptation)을 수행하게 되는데, 수학식 14를 이용하여 계산된 스펙트로그램에 도 5와 같은 CNN(Convolutional Neural Network) 기반 Bottleneck 구조의 오토 인코더(Auto Encoder) 알고리즘을 적용하여 도메인 적응을 수행한다.
- [0117] 일반적인 CNN 기반 오토 인코더 알고리즘의 경우, 입력 데이터를 빼기 연산에 적합한 도메인으로 변환하는 역할을 수행한다. 여기서, 오토 인코더(Auto Encoder) 알고리즘은 고차원의 입력 데이터를 저차원으로 압축한 후, 다시 원래 데이터로 복원하는 알고리즘이다. 인코딩(Encoding) 단계에서는 입력 데이터를 압축하는데, 압축하는 과정에서 신경망을 통해 입력 데이터의 중요한 특징을 추출한다.
- [0118] 디코딩(Decoding) 단계는 인코딩 단계에서 압축된 특징을 입력받아 초기의 입력 데이터로 복원하고, 이러한 과정에서 잡음을 제거해 원하는 입력 데이터의 정보만 볼 수 있다.
- [0119] CNN은 입력 데이터에 컨볼루션 연산과 풀링 연산을 반복적으로 수행하고, 입력 데이터의 특징을 보존하는 동시에 데이터의 크기를 줄이면서 학습해야 할 변수값의 개수를 줄인다. 각각의 연산을 수행하는 레이어를 컨볼루션 레이어(Convolutional Layer)와 풀링 레이어(Pooling Layer)로 나타낸다.
- [0120] 컨볼루션 레이어는 2차원 입력에 고정된 크기의 2차원 필터로 컨볼루션 연산을 수행한다. 풀링 레이어는 컨볼루션 레이어의 결과를 입력으로 사용하여 이웃한 영역 내부에서 최대값을 선정함으로써 차원을 축소한다.
- [0121] 특징 강화부(121)는 도메인 적응을 수행하기 위하여 4개의 인코더 레이어와 2개의 디코더 레이어를 구성한다.
- [0122] 인코더 레이어에서는 일정한 크기의 컨볼루션 레이어(Convolutional Layer)를 통해 출력 크기와 차원을 줄이면서 특징을 추출하고, 디코더 레이어에서는 일정한 크기의 컨볼루션 레이어를 통해 출력 크기의 차원을 늘리면서 출력 크기를 입력 스펙트로그램과 동일하게 한다.
- [0123] 특징 강화부(121)는 컨볼루션 연산을 수행한 인코더 레이어가 끝날 때마다 최대 풀링 레이어(Max Pooling Layer), 정규화 연산인 배치 정규화(Batch Normalization), 활성화 함수인 RELU 함수를 수행한다. 최대 풀링 레이어는 컨볼루션 레이어에서 컨볼루션 연산을 수행한 후, 사이즈를 리사이즈하여 만드는 과정을 의미한다.
- [0124] 배치 정규화는 활성화 함수의 앞쪽에 배치가 되며, 이는 역전파(Back Propagation)를 통해서 학습이 가능하다.
- [0125] 특징 강화부(121)는 컨볼루션 연산을 수행한 디코더 레이어가 끝날 때마다 정규화 연산인 배치 정규화(Batch Normalization), 활성화 함수인 RELU 함수를 수행한다.
- [0126] 특징 강화부(121)는 입력 스펙트로그램에서 도메인 적응을 수행한 결과인 잡음 특징을 뺄셈 연산으로 불필요한 성분을 제거한 후, 비선형 활성화 함수인 ReLU(Rectified Linear Unit)을 적용하고, 0에서 1 범위의 min-max 정규화(Normalization)를 통해 특징의 크기를 조절하여 음향 신호의 특징을 강화한다.
- [0127] 도 6에 도시된 바와 같이, 동물 중 식별부(122)는 CNN 기반 분류 알고리즘과 완전 연결 네트워크(Fully Connected Network, FCN)으로 구성되어 있으며, 마지막에 Softmax 함수를 통해 최종 결과를 도출한다. 다시 말해, 동물 중 식별부(122)는 5개의 컨볼루션 레이어(Convolutional Layer)와 2개의 완전 연결 레이어(Fully Connected Layer, FCL)로 구성된다.
- [0128] 동물 중 식별부(122)는 첫 번째, 두 번째, 마지막의 컨볼루션 레이어의 컨볼루션 연산을 수행한 후, 최대 풀링(Max Pooling)을 수행한다.
- [0129] 동물 중 식별부(122)는 특징 강화부(121)로부터 강화된 음향 신호의 특징이 입력되면, 컨볼루션 레이어에서 컨볼루션 연산을 수행하고, 마지막 컨볼루션 레이어에서 컨볼루션 연산을 수행한 후, FCN과 연결되어 출력 레이어에서 레이블 수만큼의 결과를 산출한다.
- [0130] 동물 중 식별부(122)는 산출된 결과를 하기의 수학식 15의 소프트맥스(Softmax) 함수를 통하여 점수(Score)로 나타나게 되고, 최종적으로 가장 높은 점수를 가진 레이블(Label)을 최종 결과(S(yi))로 분류한다.

수학식 15

$$S(y_i) = \frac{\exp(y_i)}{\sum_{j=1}^N \exp(y_j)}$$

[0131]

[0132]

이상에서 본 발명의 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속하는 것이다.

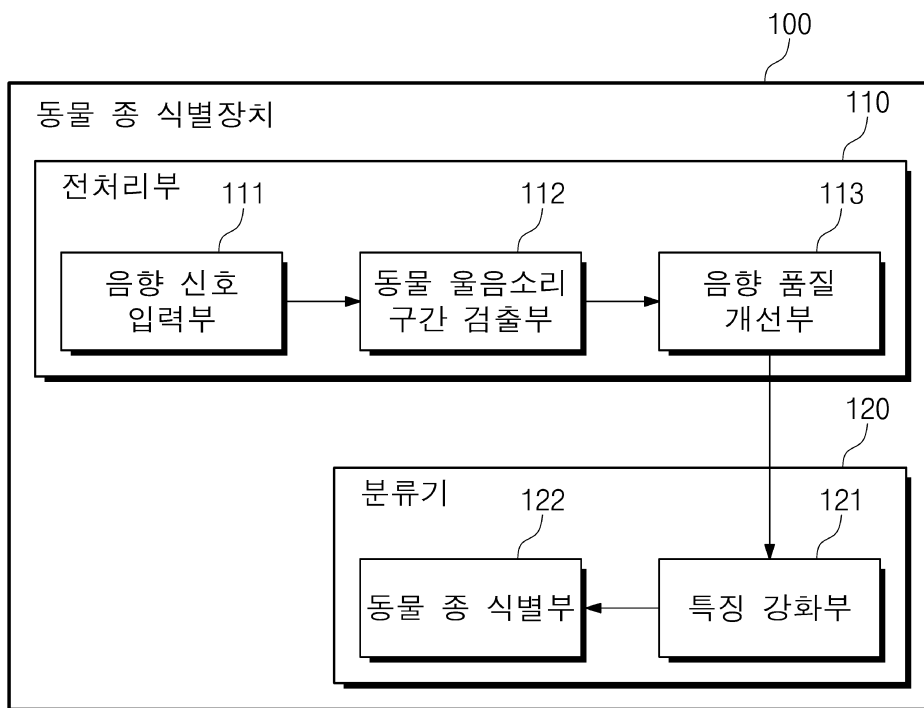
부호의 설명

[0133]

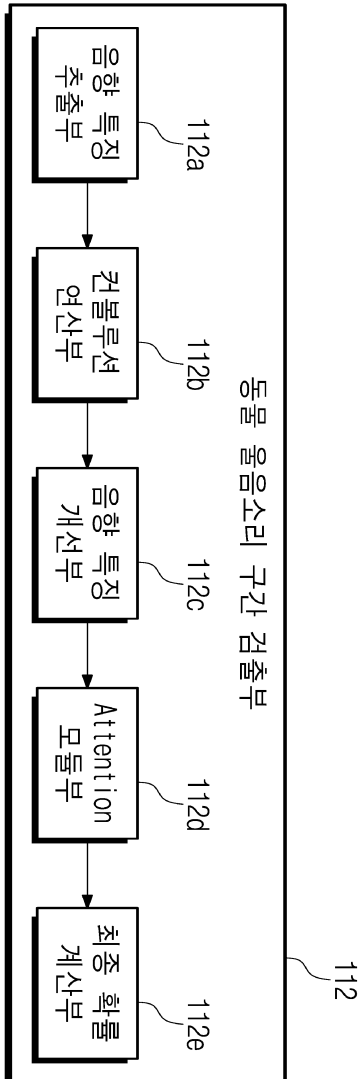
100: 동물 종 식별 장치	110: 전처리부
111: 음향 신호 입력부	112: 동물 울음소리 구간 검출부
112a: 음향 특징 추출부	112b: 컨볼루션 연산부
112c: 음향 특징 개선부	112d: 어텐션 모듈부
112e: 최종 확률 계산부	113: 음향 품질 개선부
120: 분류기	121: 특징 강화부
122: 동물 종 식별부	

도면

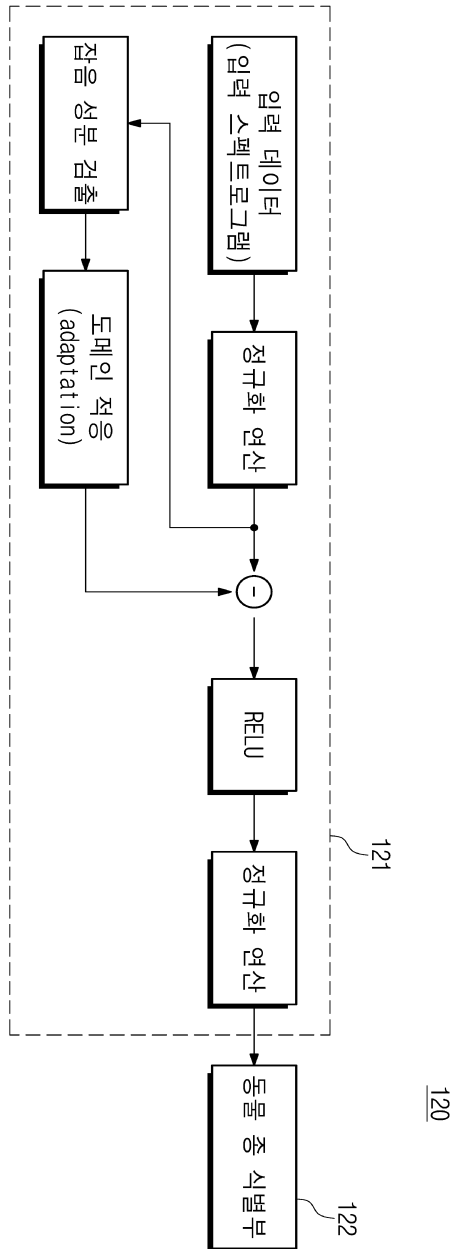
도면1



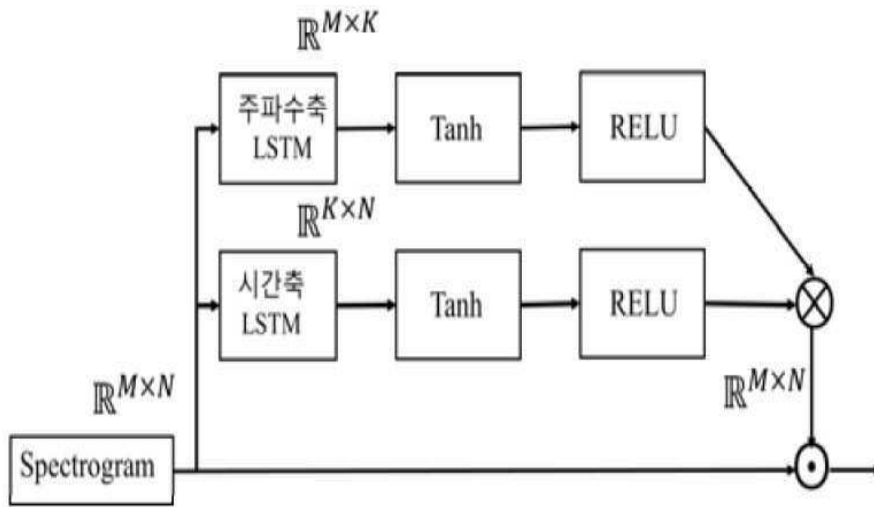
도면2



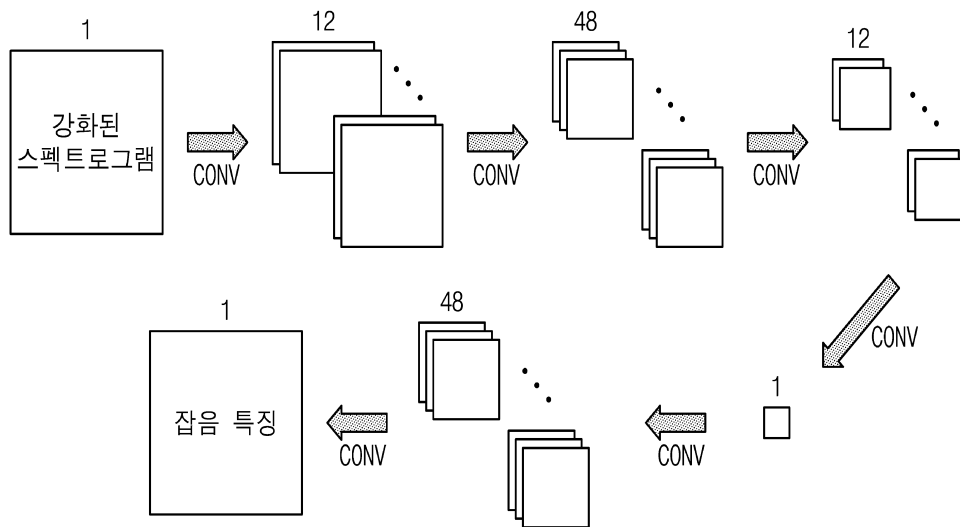
도면3



도면4



도면5



도면6

