



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년11월23일
(11) 등록번호 10-2469219
(24) 등록일자 2022년11월16일

(51) 국제특허분류(Int. Cl.)
G06F 11/22 (2017.01) G06F 11/30 (2006.01)
G06N 20/00 (2019.01)
(52) CPC특허분류
G06F 11/2263 (2013.01)
G06F 11/3068 (2013.01)
(21) 출원번호 10-2022-0065470
(22) 출원일자 2022년05월27일
심사청구일자 2022년05월27일
(56) 선행기술조사문헌
KR1020210094456 A
(뒷면에 계속)

(73) 특허권자
국방과학연구소
대전광역시 유성구 북유성대로488번길 160 (수남동)
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
고진원
서울특별시 성북구 안암로 73-15, 신공학관 405호
(74) 대리인
특허법인 광장리앤코

전체 청구항 수 : 총 10 항

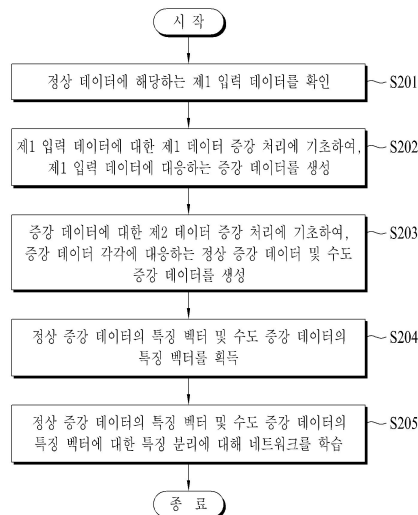
심사관 : 김계준

(54) 발명의 명칭 이상 데이터 검출 방법 및 이를 위한 전자 장치

(57) 요약

본 개시에 따르면, 전자 장치의 이상 데이터 검출 방법은, 정상 데이터에 해당하는 제1 입력 데이터를 확인하는 단계; 상기 제1 입력 데이터에 대한 제1 데이터 증강 처리에 기초하여, 상기 제1 입력 데이터에 대응하는 증강 데이터를 생성하는 단계; 상기 증강 데이터에 대한 제2 데이터 증강 처리에 기초하여, 상기 증강 데이터 각각에 대응하는 정상 증강 데이터 및 수도(pseudo) 증강 데이터를 생성하는 단계; 상기 정상 증강 데이터의 특징 벡터 및 상기 수도 증강 데이터의 특징 벡터를 획득하는 단계; 및 상기 정상 증강 데이터 및 상기 수도 증강 데이터를 각각 상기 제1 입력 데이터에 관한 정상 데이터 및 이상 데이터로 식별하기 위해, 상기 정상 증강 데이터의 특징 벡터 및 상기 수도 증강 데이터의 특징 벡터에 대한 특징 분리에 대해 네트워크를 학습시키는 단계를 포함할 수 있다.

대표도 - 도2



(52) CPC특허분류

G06F 11/3072 (2013.01)

G06N 20/00 (2021.08)

(56) 선행기술조사문헌

KR1020190143192 A

KR102372508 B1

KR1020210033774 A

KR102255833 B1

명세서

청구범위

청구항 1

전자 장치의 이상 데이터 검출 방법에 있어서,

정상 데이터에 해당하는 제1 입력 데이터를 확인하는 단계;

상기 제1 입력 데이터에 대한 제1 데이터 증강 처리에 기초하여, 상기 제1 입력 데이터에 대응하는 증강 데이터를 생성하는 단계;

상기 증강 데이터에 대한 제2 데이터 증강 처리에 기초하여, 상기 증강 데이터 각각에 대응하는 정상 증강 데이터 및 수도(pseudo) 증강 데이터를 생성하는 단계;

상기 정상 증강 데이터의 특징 벡터 및 상기 수도 증강 데이터의 특징 벡터를 획득하는 단계; 및

상기 정상 증강 데이터 및 상기 수도 증강 데이터를 각각 상기 제1 입력 데이터에 관한 정상 데이터 및 이상 데이터로 식별하기 위해, 상기 정상 증강 데이터의 특징 벡터 및 상기 수도 증강 데이터의 특징 벡터에 대한 특징 분리에 대해 네트워크를 학습시키는 단계를 포함하는, 이상 데이터 검출 방법.

청구항 2

제1항에 있어서,

상기 제1 데이터 증강 처리는 복수의 유형을 포함하는 2차원 변환을 포함하고,

상기 제1 입력 데이터에 대응하는 증강 데이터를 생성하는 단계는,

상기 제1 입력 데이터에 대해, 상기 복수의 유형 각각의 2차원 변환이 적용된 증강 데이터를 생성하는 단계를 포함하는, 이상 데이터 검출 방법.

청구항 3

제2항에 있어서,

상기 제2 데이터 증강 처리는 데이터의 크롭(crop) 및 색 변환에 기초하고,

상기 증강 데이터 각각에 대응하는 정상 증강 데이터는, 상기 2차원 변환의 복수의 유형 각각에 대해 생성된 증강 데이터를 제1 크기로 크롭하고 색 변환하여 생성되고,

상기 증강 데이터 각각에 대응하는 수도 증강 데이터는, 상기 2차원 변환의 복수의 유형 각각에 대해 생성된 증강 데이터를 제2 크기로 크롭하고 색 변환하여 생성되는, 이상 데이터 검출 방법.

청구항 4

제2항에 있어서,

상기 복수의 유형의 2차원 변환 각각의 중심점을 획득하는 단계를 더 포함하고,

상기 복수의 유형의 2차원 변환 각각의 중심점은 상기 2차원 변환의 유형 별로 계산되는 상기 정상 증강 데이터의 특징 벡터의 평균 벡터를 나타내는, 이상 데이터 검출 방법.

청구항 5

제4항에 있어서,

상기 특징 분리는 상기 정상 증강 데이터의 특징 벡터에 대한 제1 특징 분리 및 상기 수도 증강 데이터의 특징 벡터에 대한 제2 특징 분리를 포함하고,

상기 제1 특징 분리는,

상기 복수의 유형 중 일 유형에 따른 2차원 변환에 기초한 정상 증강 데이터의 특징 벡터가 동일한 유형의 2차원 변환의 중심점과 큰 유사도를 가지고, 상이한 유형의 2차원 변환의 중심점과 낮은 유사도를 가지도록 하는, 이상 데이터 검출 방법.

청구항 6

제5항에 있어서,

상기 제2 특징 분리는,

상기 수도 증강 데이터의 특징 벡터가 상기 복수의 유형의 2차원 변환 각각의 중심점과 낮은 유사도를 가지도록 하는, 이상 데이터 검출 방법.

청구항 7

제2항에 있어서,

상기 학습된 네트워크를 이용하여 새로운 제2 입력 데이터에 대한 이상 데이터 검출을 수행하는 단계를 더 포함하고,

상기 이상 데이터 검출을 수행하는 단계는,

상기 복수의 유형 각각에 따른 2차원 변환에 기초하여, 상기 제2 입력 데이터에 대응하는 증강 데이터를 생성하는 단계;

상기 제2 입력 데이터에 대응하는 증강 데이터의 특징 벡터를 획득하는 단계; 및

상기 제2 입력 데이터에 대응하는 증강 데이터의 특징 벡터 및 상기 복수의 유형의 2차원 변환 각각의 중심점에 기초하여, 상기 제2 입력 데이터가 정상 데이터인지 또는 이상 데이터인지 여부를 판단하는 단계를 포함하는, 이상 데이터 검출 방법.

청구항 8

제7항에 있어서,

상기 제2 입력 데이터가 정상 데이터인지 또는 이상 데이터인지 여부는, 상기 제2 입력 데이터에 대응하는 증강 데이터의 특징 벡터 및 상기 복수의 유형의 2차원 변환 각각의 중심점 간의 유사도로부터 획득되는 정상치 점수에 기초하여 판단되는, 이상 데이터 검출 방법.

청구항 9

이상 데이터 검출을 위한 전자 장치에 있어서,

트랜시버;

적어도 하나의 프로그램이 저장된 메모리; 및

상기 트랜시버 및 상기 메모리와 연결되는 프로세서를 포함하고, 상기 프로세서는 상기 적어도 하나의 프로그램을 실행함으로써,

정상 데이터에 해당하는 입력 데이터를 확인하고,

상기 입력 데이터에 대한 제1 데이터 증강 처리에 기초하여, 상기 입력 데이터에 대응하는 증강 데이터를 생성하고,

상기 증강 데이터에 대한 제2 데이터 증강 처리에 기초하여, 상기 증강 데이터 각각에 대응하는 정상 증강 데이터 및 수도 증강 데이터를 생성하고,

상기 정상 증강 데이터의 특징 벡터 및 상기 수도 증강 데이터의 특징 벡터를 획득하고,

상기 정상 증강 데이터 및 상기 수도 증강 데이터를 각각 상기 입력 데이터에 관한 정상 데이터 및 이상 데이터로 식별하기 위해, 상기 정상 증강 데이터의 특징 벡터 및 상기 수도 증강 데이터의 특징 벡터에 대한 특징 분리에 대해 네트워크를 학습시키도록 구성되는, 전자 장치.

청구항 10

비일시적 컴퓨터 판독 가능 저장 매체로서,

상기 비일시적 컴퓨터 판독 가능 저장 매체는, 전자 장치의 프로세서에 의해 실행되는 경우, 상기 전자 장치로 하여금 이상 데이터 검출 방법을 수행하도록 하는 적어도 하나의 프로그램을 저장하고, 상기 방법은,

정상 데이터에 해당하는 입력 데이터를 확인하는 단계;

상기 입력 데이터에 대한 제1 데이터 증강 처리에 기초하여, 상기 입력 데이터에 대응하는 증강 데이터를 생성하는 단계;

상기 증강 데이터에 대한 제2 데이터 증강 처리에 기초하여, 상기 증강 데이터 각각에 대응하는 정상 증강 데이터 및 수도 증강 데이터를 생성하는 단계;

상기 정상 증강 데이터의 특징 벡터 및 상기 수도 증강 데이터의 특징 벡터를 획득하는 단계; 및

상기 정상 증강 데이터 및 상기 수도 증강 데이터를 각각 상기 입력 데이터에 관한 정상 데이터 및 이상 데이터로 식별하기 위해, 상기 정상 증강 데이터의 특징 벡터 및 상기 수도 증강 데이터의 특징 벡터에 대한 특징 분리에 대해 네트워크를 학습시키는 단계를 포함하는, 비일시적 컴퓨터 판독 가능 저장 매체.

발명의 설명

기술 분야

[0001] 본 개시는 이상 데이터 검출 방법 및 이를 위한 전자 장치에 관한 것이다.

배경 기술

[0002] 4차 산업혁명으로 양질의 데이터를 요구하는 분야가 증가하고 있으며, 이에 따라 이상 데이터를 검출하기 위한 기술이 활발하게 사용되고 있다. 이상 데이터 검출은 기존의 데이터가 가진 특징이나 패턴과는 다른 데이터를 찾아내는 것으로서, 제조업 분야에서 장비 이상 탐지나 제품 불량 탐지, 보안 기술 분야에서 위조 기술 탐지나 CCTV 영상 보안, 의학 분야에서 의료 영상을 통한 진단 등과 같이 다양한 기술 분야에 접목되어 기술 발전에 크게 이바지하고 있다.

[0003] 이와 같은 이상 데이터 검출 기술에 있어서, 일반적으로 이상 데이터는 정상 데이터에 비해 발생 빈도가 매우 낮아 데이터 확보에 큰 비용과 시간이 소모되기 때문에, 정상 데이터에만 기초하여 이상 데이터 검출을 수행하는 다양한 방안이 연구되고 있다. 이에 따라, 정상 데이터를 이용하여 이상 데이터 검출을 효과적으로 수행할 수 있는 기술이 요구된다.

발명의 내용

해결하려는 과제

[0004] 본 개시의 실시예는 상술한 문제점을 해결하기 위하여 제안된 것으로, 정상 데이터에 해당하는 입력 데이터를 증강시킨 정상 증강 데이터와 수도 증강 데이터에 기초하여 정상 데이터 및 이상 데이터 분류를 학습하고, 이를 통해 이상 데이터의 검출을 수행하는 기술에 관한 것이다.

[0005] 본 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제들로 한정되지 않으며, 이하의 실시예들로부터 또 다른 기술적 과제들이 유추될 수 있다.

과제의 해결 수단

[0006] 일 실시예에 따르면, 전자 장치의 이상 데이터 검출 방법은, 정상 데이터에 해당하는 제1 입력 데이터를 확인하는 단계; 상기 제1 입력 데이터에 대한 제1 데이터 증강 처리에 기초하여, 상기 제1 입력 데이터에 대응하는 증강 데이터를 생성하는 단계; 상기 증강 데이터에 대한 제2 데이터 증강 처리에 기초하여, 상기 증강 데이터 각각에 대응하는 정상 증강 데이터 및 수도(pseudo) 증강 데이터를 생성하는 단계; 상기 정상 증강 데이터의 특징 벡터 및 상기 수도 증강 데이터의 특징 벡터를 획득하는 단계; 및 상기 정상 증강 데이터 및 상기 수도 증강 데이터를 각각 상기 제1 입력 데이터에 관한 정상 데이터 및 이상 데이터로 식별하기 위해, 상기 정상 증강 데이

터의 특징 벡터 및 상기 수도 증강 데이터의 특징 벡터에 대한 특징 분리에 대해 네트워크를 학습시키는 단계를 포함할 수 있다.

- [0007] 일 실시예에 따르면, 상기 제1 데이터 증강 처리는 복수의 유형을 포함하는 2차원 변환을 포함할 수 있고, 상기 제1 입력 데이터에 대응하는 증강 데이터를 생성하는 단계는, 상기 제1 입력 데이터에 대해, 상기 복수의 유형 각각의 2차원 변환이 적용된 증강 데이터를 생성하는 단계를 포함할 수 있다.
- [0008] 일 실시예에 따르면, 상기 제2 데이터 증강 처리는 데이터의 크롭(crop) 및 색 변환에 기초할 수 있고, 상기 증강 데이터 각각에 대응하는 정상 증강 데이터는 상기 2차원 변환의 복수의 유형 각각에 대해 생성된 증강 데이터를 제1 크기로 크롭하고 색 변환하여 생성될 수 있고, 상기 증강 데이터 각각에 대응하는 수도 증강 데이터는 상기 2차원 변환의 복수의 유형 각각에 대해 생성된 증강 데이터를 제2 크기로 크롭하고 색 변환하여 생성될 수 있다.
- [0009] 일 실시예에 따르면, 전자 장치의 이상 데이터 검출 방법은 상기 복수의 유형의 2차원 변환 각각의 중심점을 획득하는 단계를 더 포함할 수 있고, 상기 복수의 유형의 2차원 변환 각각의 중심점은 상기 2차원 변환의 유형 별로 계산되는 상기 정상 증강 데이터의 특징 벡터의 평균 벡터를 나타낼 수 있다.
- [0010] 일 실시예에 따르면, 상기 특징 분리는 상기 정상 증강 데이터의 특징 벡터에 대한 제1 특징 분리 및 상기 수도 증강 데이터의 특징 벡터에 대한 제2 특징 분리를 포함할 수 있고, 상기 제1 특징 분리는, 상기 복수의 유형 중 일 유형에 따른 2차원 변환에 기초한 정상 증강 데이터의 특징 벡터가 동일한 유형의 2차원 변환의 중심점과 큰 유사도를 가지고, 상이한 유형의 2차원 변환의 중심점과 낮은 유사도를 가지도록 할 수 있다.
- [0011] 일 실시예에 따르면, 상기 제2 특징 분리는, 상기 수도 증강 데이터의 특징 벡터가 상기 복수의 유형의 2차원 변환 각각의 중심점과 낮은 유사도를 가지도록 할 수 있다.
- [0012] 일 실시예에 따르면, 전자 장치의 이상 데이터 검출 방법은 상기 학습된 네트워크를 이용하여 새로운 제2 입력 데이터에 대한 이상 데이터 검출을 수행하는 단계를 더 포함할 수 있고, 상기 이상 데이터 검출을 수행하는 단계는, 상기 복수의 유형 각각에 따른 2차원 변환에 기초하여, 상기 제2 입력 데이터에 대응하는 증강 데이터를 생성하는 단계; 상기 제2 입력 데이터에 대응하는 증강 데이터의 특징 벡터를 획득하는 단계; 및 상기 제2 입력 데이터에 대응하는 증강 데이터의 특징 벡터 및 상기 복수의 유형의 2차원 변환 각각의 중심점에 기초하여, 상기 제2 입력 데이터가 정상 데이터인지 또는 이상 데이터인지 여부를 판단하는 단계를 포함할 수 있다.
- [0013] 일 실시예에 따르면, 상기 제2 입력 데이터가 정상 데이터인지 또는 이상 데이터인지 여부는, 상기 제2 입력 데이터에 대응하는 증강 데이터의 특징 벡터 및 상기 복수의 유형의 2차원 변환 각각의 중심점 간의 유사도로부터 획득되는 정상치 점수(Normality Score)에 기초하여 판단될 수 있다.
- [0014] 일 실시예에 따르면, 이상 데이터 검출을 위한 전자 장치는, 트랜시버; 적어도 하나의 프로그램이 저장된 메모리; 및 상기 트랜시버 및 상기 메모리와 연결되는 프로세서를 포함할 수 있고, 상기 프로세서는 상기 적어도 하나의 프로그램을 실행함으로써, 정상 데이터에 해당하는 입력 데이터를 확인하고, 상기 입력 데이터에 대한 제1 데이터 증강 처리에 기초하여, 상기 입력 데이터에 대응하는 증강 데이터를 생성하고, 상기 증강 데이터에 대한 제2 데이터 증강 처리에 기초하여, 상기 증강 데이터 각각에 대응하는 정상 증강 데이터 및 수도 증강 데이터를 생성하고, 상기 정상 증강 데이터의 특징 벡터 및 상기 수도 증강 데이터의 특징 벡터를 획득하고, 상기 정상 증강 데이터 및 상기 수도 증강 데이터를 각각 상기 제1 입력 데이터에 관한 정상 데이터 및 이상 데이터로 식별하기 위해, 상기 정상 증강 데이터의 특징 벡터 및 상기 수도 증강 데이터의 특징 벡터에 대한 특징 분리에 대해 네트워크를 학습시키도록 구성될 수 있다.
- [0015] 일 실시예에 따르면, 비밀식 컴퓨터 판독 가능 저장 매체는, 전자 장치의 프로세서에 의해 실행되는 경우, 상기 전자 장치로 하여금 이상 데이터 검출 방법을 수행하도록 하는 적어도 하나의 프로그램을 저장할 수 있고, 상기 방법은, 정상 데이터에 해당하는 입력 데이터를 확인하는 단계; 상기 입력 데이터에 대한 제1 데이터 증강 처리에 기초하여, 상기 입력 데이터에 대응하는 증강 데이터를 생성하는 단계; 상기 증강 데이터에 대한 제2 데이터 증강 처리에 기초하여, 상기 증강 데이터 각각에 대응하는 정상 증강 데이터 및 수도 증강 데이터를 생성하는 단계; 상기 정상 증강 데이터의 특징 벡터 및 상기 수도 증강 데이터의 특징 벡터를 획득하는 단계; 및 상기 정상 증강 데이터 및 상기 수도 증강 데이터를 각각 상기 제1 입력 데이터에 관한 정상 데이터 및 이상 데이터로 식별하기 위해, 상기 정상 증강 데이터의 특징 벡터 및 상기 수도 증강 데이터의 특징 벡터에 대한 특징 분리에 대해 네트워크를 학습시키는 단계를 포함할 수 있다.

[0016] 기타 실시예들의 구체적인 사항은 상세한 설명 및 도면들에 포함된다.

발명의 효과

[0017] 본 개시에 따르면, 정상 데이터로부터 생성되는 수도 증강 데이터를 이상 데이터로 간주하여 네트워크를 학습시킴으로써, 정상 데이터만을 이용한 경우에도 정상 데이터 및 이상 데이터 판별이 효과적으로 학습될 수 있다.

[0018] 본 개시에 따르면, 정상 데이터로부터 생성되는 수도 증강 데이터를 이상 데이터로 간주하여 네트워크를 학습시킴으로써, 데이터의 양이 부족한 상황에서도 효과적인 학습이 가능하며, 학습을 위한 데이터의 수집 비용이 절감될 수 있다.

[0019] 발명의 효과는 이상에서 언급한 효과로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 청구범위의 기재로부터 당해 기술 분야의 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0020] 도 1은 일 실시예에 따른 전자 장치의 블록도를 나타낸다.

도 2는 일 실시예에 따른 전자 장치에 의해 수행되는 이상 데이터 검출을 위한 네트워크 학습의 흐름도를 나타낸다.

도 3은 일 실시예에 따른 제1 데이터 증강에 관한 예시도를 나타낸다.

도 4는 일 실시예에 따른 제2 데이터 증강에 관한 예시도를 나타낸다.

도 5는 일 실시예에 따른 전자 장치에 의해 수행되는 이상 데이터 검출의 흐름도를 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0021] 실시 예들에서 사용되는 용어는 본 개시에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 당 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 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서 본 개시에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 개시의 전반에 걸친 내용을 토대로 정의되어야 한다.

[0022] 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 있음을 의미한다. 또한, 명세서에 기재된 "...부", "...모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.

[0023] 명세서 전체에서 기재된 "a, b, 및 c 중 적어도 하나"의 표현은, 'a 단독', 'b 단독', 'c 단독', 'a 및 b', 'a 및 c', 'b 및 c', 또는 'a,b,c 모두'를 포괄할 수 있다.

[0024] 이하에서 언급되는 "단말", "장치" 및 "전자 장치"는 네트워크를 통해 서버나 타 단말에 접속할 수 있는 컴퓨터나 휴대용 단말로 구현될 수 있다. 여기서, 컴퓨터는 예를 들어, 웹 브라우저(WEB Browser)가 탑재된 노트북, 데스크톱(desktop), 랩톱(laptop) 등을 포함하고, 휴대용 단말은 예를 들어, 휴대성과 이동성이 보장되는 무선 통신 장치로서, IMT(International Mobile Telecommunication), CDMA(Code Division Multiple Access), W-CDMA(W-Code Division Multiple Access), LTE(Long Term Evolution) 등의 통신 기반 단말, 스마트폰, 태블릿 PC 등과 같은 모든 종류의 핸드헬드(Handheld) 기반의 무선 통신 장치를 포함할 수 있다.

[0025] 위에서 언급된 바와 같이, 비정상 데이터, 즉 이상 데이터는 정상 데이터에 비해 발생 빈도가 매우 적기 때문에 데이터 확보에 큰 비용과 시간이 요구된다. 예를 들어, CCTV 등과 같은 감시 영상에서 이상 데이터는 범죄 현상이나 범죄 가능성이 있는 영상을 정의하게 되며 이는 획득하기 매우 어려울 수 있다. 이에 따라, 일반적으로 정상 데이터만을 이용하여 정상 데이터 또는 이상 데이터 여부의 분류를 수행하는 방법이 이용된다. 이러한 방법들은 정상 데이터만 이용하기 때문에 이상 데이터에 해당하는 특징 정보를 학습하지 못하므로, 데이터의 정상 또는 이상의 판별이 어렵다는 문제점을 가진다. 본 개시는 정상 데이터만을 이용하여 이상 데이터 검출을 효과적으로 수행하는 방법을 제안한다. 우선, 본 개시의 전반에서 사용되는 용어들이 아래에서 상세히 설명된다.

[0026] 코사인 유사도

[0027] 코사인(cosine) 유사도는 두 벡터가 이루는 각도의 코사인 값으로서, 두 벡터의 유사 정도를 나타낼 수 있다. 이 값은 벡터의 크기가 아닌 방향의 유사도를 판단하는 목적으로 사용될 수 있다. 구체적으로, 벡터의 크기와는 관계 없이, 두 벡터의 방향이 완전히 동일할 경우(즉, 두 벡터가 이루는 각의 크기가 0도인 경우), 코사인 유사도는 1의 값을 가지며, 두 벡터의 방향이 완전히 반대일 경우(즉, 두 벡터가 이루는 각의 크기가 180도인 경우), 코사인 유사도는 -1의 값을 가질 수 있다. 코사인 유사도의 계산에 있어서는 일반적으로 정규화된 두 벡터가 이용될 수 있다.

[0028] 이상 데이터

[0029] 이상 데이터는 이상치로 지칭될 수도 있으며, 기존 데이터의 패턴이나 특징과는 다른 새로운 데이터로서, 일반적으로 관측된 데이터의 범위에서 벗어난 데이터를 의미할 수 있다. 기존에는 관측된 데이터의 특징 벡터로부터 관측 데이터를 대표하는 평균 특징 벡터를 획득하고, 새로운 데이터와 평균 특징 벡터 간의 유사도를 코사인 유사도 또는 거리로서 측정하여, 측정된 유사도가 특정 수준 이상일 경우 해당 데이터를 정상 데이터로 판별하고, 측정된 유사도가 특정 수준 이하일 경우 해당 데이터를 이상 데이터로 판별한다. 이러한 이상 데이터의 개념은 영상 처리 분야 뿐만 아니라, 의료 분야, 사이버 보안, 통신 분야의 신호 처리 등 다양한 경우에 적용되고 있으며, 이를 검출하고자 하는 기술이 활발히 연구되고 있다.

[0030] 데이터 증강

[0031] 데이터 증강은 기존 데이터에 인위적인 변환을 가함으로써 새로운 데이터를 대량 확보하는 기법을 의미할 수 있다. 인위적인 변환의 예로서, 회전, 좌우 반전, 크롭(잘라내기), 밝기 조절, 색상 변경 등이 있을 수 있다. 데이터 증강은 기존 데이터의 정보는 그대로 포함하면서 변환된 데이터를 추가로 이용하기 때문에, 적은 양의 데이터만 보유한 경우이거나 또는 추가적인 데이터를 획득하는 데에 큰 비용이 드는 경우에 사용될 수 있다. 또한, 데이터 증강은 데이터 양을 증가시켜 네트워크의 학습 범위를 확장할 수 있으므로 네트워크의 훈련에 효과적으로 이용될 수 있다.

[0032] 인코딩

[0033] 인코딩은 보안, 저장 공간 절약(압축), 표준화 등 다양한 목적을 위해 정보의 형태나 형식을 변환하는 처리 또는 처리 방식을 의미할 수 있다. 영상 처리 분야에서, 인코더에 입력되는 영상에 대해 인코딩이 수행되고, 인코딩된 영상은 벡터로 표현될 수 있다. 여기서, 인코더를 통해, 관측된 데이터의 특징 벡터가 추출될 수 있다. 이상 데이터 검출은 정상 데이터의 특징 벡터와 이상 데이터의 특징 벡터가 효과적으로 분리되도록 하는 인코더를 얻는 기술일 수 있다. 따라서, 이상 데이터 검출 기술의 주 목적은 인코더가 정상 및 이상 데이터를 효과적으로 검출할 수 있도록 하는 좋은 특징 벡터를 추출하는 것이 될 수 있다.

[0034] 손실 함수

[0035] 손실 함수는 네트워크의 출력 값과 원하는 출력값 간의 차이, 즉 오차를 의미할 수 있다. 네트워크는 이러한 손실 함수를 감소시키도록 학습된다. 또한, 손실 함수는 네트워크의 학습 방향을 함축하고, 목적 함수로도 지칭될 수 있다. 일반적으로, 손실 함수로서 평균 제곱 오차 및 교차 엔트로피 오차가 이용될 수 있다.

[0036] 아래에서는 첨부한 도면을 참고하여 본 개시의 실시 예에 대하여 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 개시는 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시 예에 한정되지 않는다.

[0037] 이하, 도면을 참조하여 본 개시의 실시 예들이 상세히 설명된다.

[0038] 도 1은 일 실시예에 따른 전자 장치(100)의 블록도를 나타낸다. 도 1을 참조하면, 전자 장치(100)는 트랜시버(transceiver)(110), 프로세서(120) 및 메모리(130)를 포함할 수 있다. 도 1에는 본 개시의 실시예와 관련된 전자 장치(100)의 구성요소들만이 도시되어 있다. 따라서, 도 1에 도시된 구성요소들 외에 다른 범용적인 구성요소들이 전자 장치에 더 포함될 수 있음을 본 실시 예와 관련된 기술분야에서 통상의 지식을 가진 자라면 이해할 수 있다.

[0039] 트랜시버(110)는 유/무선 통신을 수행하여 관련 정보를 송수신할 수 있다. 트랜시버(110)가 이용하는 통신 기술에는 GSM(Global System for Mobile communication), CDMA(Code Division Multi Access), LTE(Long Term Evolution), 5G, WLAN(Wireless LAN), Wi-Fi(Wireless-Fidelity), 블루투스(Bluetooth™), RFID(Radio Frequency Identification), 적외선 통신(Infrared Data Association; IrDA), ZigBee, NFC(Near Field

Communication) 등이 있을 수 있다.

- [0040] 프로세서(120)는 전자 장치(100)의 전반적인 동작을 제어하고 데이터 및 신호를 처리할 수 있다. 프로세서(120)는 적어도 하나의 하드웨어 유닛으로 구성될 수 있다. 또한, 프로세서(120)는 메모리(130)에 저장된 프로그램 코드를 실행하여 생성되는 하나 이상의 소프트웨어 모듈에 의해 동작할 수 있다. 프로세서(120)는 메모리(130)에 저장된 프로그램 코드를 실행하여 전자 장치(100)의 전반적인 동작을 제어하고 데이터 및 신호를 처리할 수 있다. 프로세서(120)는 본 개시 전반에서 설명되는 이상 데이터 검출에 관한 방법에 관한 내용을 수행하도록 구성될 수 있다.
- [0041] 메모리(130)는 본 개시 전반에서 설명되는 이상 데이터 검출에 관한 방법을 수행하기 위한 프로그램을 실행하는데 필요한 프로그램 코드 및 정보를 저장할 수 있다. 메모리(130)는 휘발성 메모리 또는 비휘발성 메모리일 수 있다.
- [0042] 도 2는 일 실시예에 따른 전자 장치에 의해 수행되는 이상 데이터 검출을 위한 네트워크 학습의 흐름도를 나타낸다. 도 2는 도 1에 도시된 전자 장치(100)의 동작 방법을 예시할 수 있다.
- [0043] 도 2를 참조하면, 단계 S201에서, 전자 장치는 정상 데이터에 해당하는 제1 입력 데이터를 확인할 수 있다. 정상 데이터인 제1 입력 데이터는 이상 데이터 검출을 위한 학습에 이용되는 학습 데이터로서 전자 장치에 입력될 수 있다.
- [0044] 단계 S202에서, 전자 장치는 제1 입력 데이터에 대한 제1 데이터 증강 처리에 기초하여, 제1 입력 데이터에 대응하는 증강 데이터를 생성할 수 있다. 일 실시예에서, 제1 데이터 증강 처리는 복수의 유형을 포함하는 2차원 변환을 포함할 수 있다. 따라서, 전자 장치는 제1 입력 데이터에 대응하는 증강 데이터를 생성하기 위해, 제1 입력 데이터에 대해, 복수의 유형 각각의 2차원 변환이 적용된 증강 데이터를 생성할 수 있다.
- [0045] 보다 구체적으로, 전자 장치는 하나의 데이터로부터 보다 많은 특징 정보를 획득하기 위하여, 제1 입력 데이터에 대해 2차원 변환에 기초한 제1 데이터 증강 처리를 수행할 수 있다. 여기서, 2차원 변환은 영상의 좌우 반전 및 회전의 조합에 기초하여 수행될 수 있고, 입력 데이터에 포함된 정보는 2차원 변환이 수행되더라도 손실되지 않을 수 있다. 제1 데이터 증강 처리를 통해, 하나의 제1 입력 데이터가 복수의 증강 데이터로 증강될 수 있다. 예를 들어, 도 3을 참조하면, 하나의 제1 입력 데이터(300)는, 제1 데이터 증강 처리를 통해, 상이한 좌우 반전 및 회전의 조합으로 2차원 변환된 8개의 증강 데이터(310 내지 380)로 증강될 수 있다. 즉, 도 3에서, 전자 장치는 제1 입력 데이터(300)에 8개 유형 각각의 2차원 변환이 적용된 증강 데이터(310 내지 380)를 생성할 수 있다.
- [0046] 단계 S203에서, 전자 장치는 증강 데이터에 대한 제2 데이터 증강 처리에 기초하여, 증강 데이터 각각에 대응하는 정상 증강 데이터 및 수도 증강 데이터를 생성할 수 있다. 일 실시예에서, 제2 데이터 증강 처리는 데이터의 크롭(crop) 및 색 변환에 기초할 수 있다. 증강 데이터 각각에 대응하는 정상 증강 데이터는, 2차원 변환의 복수의 유형 각각에 대해 생성된 증강 데이터를 제1 크기로 크롭하고 색 변환하여 생성될 수 있다. 또한, 증강 데이터 각각에 대응하는 수도 증강 데이터는, 2차원 변환의 복수의 유형 각각에 대해 생성된 증강 데이터를 제2 크기로 크롭하고 색 변환하여 생성될 수 있다. 여기서, 제1 크기 및 제2 크기는 증강 데이터의 크기에 대한 상이한 비율에 기초하여 결정될 수 있다. 예를 들어, 정상 증강 데이터는, 증강 데이터의 본래 크기에 대해 0.75 내지 1배의 크기를 가지도록 증강 데이터를 크롭하여 생성될 수 있다. 수도 증강 데이터는 증강 데이터의 본래 크기에 대해 0.25 내지 0.4배의 크기를 가지도록 증강 데이터를 크롭하여 생성될 수 있다. 또한, 제2 데이터 증강 처리에서 색 변환은 랜덤하게 적용될 수 있다.
- [0047] 보다 구체적으로, 전자 장치는 제1 데이터 증강 처리를 통해 생성된 복수의 증강 데이터 각각에 대해, 영상의 크롭 및 색 변환에 기초한 제2 데이터 증강 처리를 수행할 수 있다. 제2 데이터 증강 처리를 통해, 하나의 증강 데이터가 복수의 정상 증강 데이터 및 복수의 수도 증강 데이터로 증강될 수 있다. 이 경우, 하나의 증강 데이터에 대해 동일한 개수의 정상 증강 데이터와 수도 증강 데이터가 생성될 수 있다. 또한, 정상 증강 데이터의 크기와 수도 증강 데이터의 크기는 증강 데이터의 본래 크기에 대해 상이한 비율을 가질 수 있다. 따라서, 하나의 증강 데이터로부터 생성된 복수의 정상 증강 데이터는, 증강 데이터를 서로 동일한 제1 크기로 크롭하면서 크롭의 위치를 달리하여 생성될 수 있다. 마찬가지로, 하나의 증강 데이터로부터 생성된 복수의 수도 증강 데이터는 증강 데이터를 정상 증강 데이터와는 다른 제2 크기로 크롭하면서 크롭의 위치를 달리하여 생성될 수 있다. 예를 들어, 도 4를 참조하면, 도 3에서 생성된 증강 데이터(310 내지 380) 중 증강 데이터(310)에 대해, 4개의 정상 증강 데이터(410 내지 440) 및 4개의 수도 증강 데이터(450 내지 480)가 생성될 수 있다. 이 경우,

증강 데이터(320 내지 380)에 대해서도 증강 데이터(310) 동일한 개수만큼의 정상 증강 데이터 및 수도 증강 데이터가 생성되므로, 제1 입력 데이터(300)는 총 32개의 정상 증강 데이터와 32개의 수도 증강 데이터로 증강될 수 있다.

[0048] 단계 S204에서, 전자 장치는 정상 증강 데이터의 특징 벡터 및 수도 증강 데이터의 특징 벡터를 획득할 수 있다. 전자 장치는 입력 영상, 즉 제1 입력 데이터에 대해 복수의 유형의 2차원 변환에 기초하여 증강된 증강 데이터로부터 생성된 정상 증강 데이터의 특징 벡터 및 수도 증강 데이터의 특징 벡터를 획득할 수 있다. 또한, 전자 장치는 복수의 유형의 2차원 변환 각각의 중심점을 획득할 수 있다. 이때, 복수의 유형의 2차원 변환 각각의 중심점은 2차원 변환의 유형 별로 계산되는 정상 증강 데이터의 특징 벡터의 평균 벡터를 나타낼 수 있다.

[0049] 이와 같이, 증강된 데이터의 특징 벡터와 정상 증강 데이터의 평균 벡터를 획득하는 과정 전반은 특징 추출 과정이라고 지칭될 수 있으며, 전자 장치에서 구현되는 인코더를 포함하는 네트워크를 통해 전자 장치 내부에서 수행될 수 있다. 이때, 네트워크는 뉴럴 네트워크로서, 예를 들어 복수의 층의 CNN(Convolution Neural Network)에 대응할 수 있다. 즉, 네트워크는 제1 입력 데이터로부터 생성된 정상 증강 데이터와 수도 증강 데이터를 입력 받아 이들의 특징 벡터를 획득하고, 정상 증강 데이터의 특징 벡터의 평균 벡터인 2차원 변환의 중심점을 획득할 수 있다.

[0050] 보다 구체적으로, 전자 장치는 2차원 변환의 복수의 유형 중 m번째 2차원 변환이 적용된 정상 증강 데이터의 집합 χ_m 에 포함된 각각의 정상 증강 데이터 x_m 의 특징 벡터 $\phi(x_m)$ 를 추출하고, 아래의 수학식 1과 같이 m번째 2차원 변환에 기초한 정상 증강 데이터의 평균 벡터 c_m , 즉 m번째 2차원 변환의 중심점을 계산할 수 있다.

[0051] [수학식 1]

$$c_m = \frac{\left(\sum_{x_m \in \chi_m} \phi(x_m) \right)}{\left\| \sum_{x_m \in \chi_m} \phi(x_m) \right\|}$$

[0052]

[0053] 단계 S205에서, 전자 장치는 정상 증강 데이터의 특징 벡터 및 수도 증강 데이터의 특징 벡터에 대한 특징 분리에 대해 네트워크를 학습시킬 수 있다. 일 실시예에서, 특징 분리는 정상 증강 데이터의 특징 벡터에 대한 제1 특징 분리 및 수도 증강 데이터의 특징 벡터에 대한 제2 특징 분리를 포함할 수 있다. 즉, 정상 증강 데이터와 수도 증강 데이터에 대해 특징 분리가 상이하게 수행될 수 있다.

[0054] 제1 특징 분리는, 복수의 유형 중 일 유형에 따른 2차원 변환에 기초한 정상 증강 데이터의 특징 벡터가 동일한 유형의 2차원 변환의 중심점과 큰 유사도를 가지고, 상이한 유형의 2차원 변환의 중심점과 낮은 유사도를 가지도록 하는 처리를 포함할 수 있다. 이를 통해, 네트워크는 앞서 획득한 정상 증강 데이터의 특징 벡터에 대해 수학식 2에서 정의되는 손실 함수를 계산함으로써 동일한 2차원 변환이 적용된 정상 증강 데이터는 유사도가 높은 특징 벡터를 가지도록 학습될 수 있다. 또한, 네트워크는 앞서 획득한 정상 증강 데이터의 특징 벡터에 대해 수학식 3에서 정의되는 손실 함수를 계산함으로써 상이한 2차원 변환이 적용된 정상 증강 데이터는 서로 유사도가 낮은 특징 벡터를 가지도록 학습될 수 있다.

[0055] M개의 2차원 변환 중 m번째 2차원 변환을 통해 생성된 정상 증강 데이터의 특징 벡터 $\phi(x_m)$ 가 서로 높은 유사도를 가지도록 하기 위한 손실 함수 L_{att} 는 수학식 2에서 정의될 수 있다.

[0056] [수학식 2]

$$L_{att} = - \sum_{m=1}^M \phi(x_m)^T c_m$$

[0057]

[0058] 또한, M개의 2차원 변환 중 m번째 2차원 변환을 통해 생성된 정상 증강 데이터의 특징 벡터 $\phi(x_m)$ 가 m과 상이

한 1번째 2차원 변환을 통해 생성된 정상 증강 데이터와는 서로 낮은 유사도를 가지도록 하기 위한 손실 함수 L_{nor} 는 수학적 식 3에서 정의될 수 있다. 손실 함수 L_{nor} 는 특징 벡터 $\phi(x_m)$, 1번째 2차원 변환의 중심점 c_l , 특징 벡터와 중심점 사이의 유사도의 임계값 θ , 1번째 2차원 변환에 대한 설정 값 w_l 에 기초하여 계산될 수 있다. 아래의 수학적 식 3을 참조하면, 손실 함수 L_{nor} 은 m번째 2차원 변환을 통해 생성된 정상 증강 데이터의 특징 벡터 $\phi(x_m)$ 와 m과 상이한 1번째 2차원 변환의 중심점 c_l 간의 유사도와 임계값 θ 중 더 큰 값에 기초하여 계산될 수 있다.

[0059] [수학적 식 3]

$$L_{nor} = \sum_{m=1}^M \sum_{l \neq m} w_l \cdot \max(\phi(x_m)^T c_l, \theta)$$

[0061] 손실 함수 L_{att} 는 정상 증강 데이터의 특징 벡터와 대응하는 2차원 변환의 중심점 간 유사도가 클수록 작은 값을 가질 수 있고, 손실 함수 L_{nor} 는 정상 증강 데이터의 특징 벡터와 상이한 2차원 변환의 중심점 간 유사도가 작을수록 작은 값을 가질 수 있다. 따라서, 네트워크는 앞서 획득한 정상 증강 데이터의 특징 벡터 각각에 대해 손실 함수 L_{att} 및 손실 함수 L_{nor} 의 값을 계산하여 목표 값과 비교하고, 정상 증강 데이터에 대해, 동일한 2차원 변환에 대해 손실 함수 L_{att} 이 작은 값을 가지고 상이한 2차원 변환에 대해 손실 함수 L_{nor} 이 작은 값을 가지도록 하는 특징 벡터가 최종적으로 획득되도록 학습될 수 있다. 예를 들어, 손실 함수 L_{att} 및 L_{nor} 은 그 출력 값이 0으로 수렴되도록 학습될 수 있으며, 또는 0이 아닌 특정한 기준을 만족하는 작은 값으로 수렴되도록 학습될 수 있다.

[0062] 또한, 제2 특징 분리는, 수도 증강 데이터의 특징 벡터가 복수의 유형의 2차원 변환 각각의 중심점과 낮은 유사도를 가지도록 하는 처리를 포함할 수 있다. 이를 통해, 네트워크는 앞서 획득한 수도 증강 데이터의 특징 벡터에 대해 수학적 식 4에서 정의되는 손실 함수를 계산함으로써 수도 증강 데이터가 정상 증강 데이터로부터 획득된 모든 2차원 변환의 중심점과 유사도가 낮은 특징 벡터를 가지도록 학습될 수 있다. M개의 2차원 변환 중 m번째 2차원 변환을 통해 생성된 수도 증강 데이터의 특징 벡터 $\overline{x}_m$ 가 2차원 변환의 중심점과 낮은 유사도를 가지도록 하기 위한 손실 함수 L_{abn} 는 아래 수학적 식 4에서 정의될 수 있다.

[0063] [수학적 식 4]

$$L_{abn} = \sum_{m=1}^M \phi(\overline{x}_m)^T c_m$$

[0065] 손실 함수 L_{abn} 는 수도 증강 데이터의 특징 벡터와 2차원 변환의 중심점 간 유사도가 작을수록 작은 값을 가질 수 있다. 따라서, 네트워크는 앞서 획득한 수도 증강 데이터의 특징 벡터 각각에 대해 손실 함수 L_{abn} 의 값을 계산하여 목표 값과 비교하고, 수도 증강 데이터에 대해, 손실 함수 L_{abn} 이 작은 값을 가지도록 하는 특징 벡터가 최종적으로 획득되도록 학습될 수 있다. 예를 들어, 손실 함수 L_{abn} 은 그 출력 값이 0으로 수렴되도록 학습될 수 있으며, 또는 0이 아닌 특정한 기준을 만족하는 작은 값으로 수렴되도록 학습될 수 있다.

[0066] 이와 같이, 본 개시의 실시예는 데이터 증강을 이용하여 정상 데이터인 입력 데이터로부터 정상 증강 데이터와 수도 증강 데이터를 생성하고, 정상 증강 데이터를 입력 데이터에 관한 정상 데이터로, 그리고 수도 증강 데이

터를 입력 데이터에 관한 이상 데이터로 간주하여 네트워크의 학습에 이용할 수 있다. 또한, 네트워크는 손실 함수를 감소시키도록 학습되므로, 정상 증강 데이터는 2차원 변환 유형이 동일할 경우 유사도가 크고, 2차원 변환이 상이할 경우 유사도가 작은 특징 벡터를 가지게 되며, 수도 증강 데이터는 정상 증강 데이터와 유사도가 작은 특징 벡터를 가지게 될 수 있다. 이를 통해, 정상 데이터만으로도 정상 데이터와 이상 데이터의 특징 분리가 효과적으로 학습될 수 있다.

[0067] 도 3은 일 실시예에 따른 제1 데이터 증강에 관한 예시도를 나타낸다. 도 3은 제1 입력 데이터(300)가 영상인 경우를 예시한다. 도 3을 참조하면, 제1 입력 데이터(300)에 대해 8개 유형의 2차원 변환이 적용됨으로써 총 8개의 증강 데이터(310 내지 380)이 생성될 수 있다. 이때, 2차원 변환의 유형은 영상의 좌우 반전 여부 및 회전 정도(시계 방향으로 0도, 90도, 180도, 270도)의 조합에 기초하여 결정될 수 있다. 예를 들어, 2차원 변환의 제1 유형 내지 제4 유형은 좌우 반전 없이 각각 0도, 90도, 180도, 및 270도 회전이 조합된 변환으로서, 증강 데이터(310 내지 340)를 참고하여 확인될 수 있다. 또한, 2차원 변환의 제5 유형 내지 제8 유형은 좌우 반전을 적용하고 각각 0도, 90도, 180도, 및 270도 회전이 조합된 변환으로서, 증강 데이터(350 내지 380)를 참고하여 확인될 수 있다. 도 3에서 설명된 2차원 변환 유형의 수와 조합의 내용은 단지 설명의 편의를 위해 예시되었으며, 본 개시의 실시예는 이에 한정되지 않을 것이다.

[0068] 도 4는 일 실시예에 따른 제2 데이터 증강에 관한 예시도를 나타낸다. 도 4는 도 3에서 설명된 영상인 제1 입력 데이터로부터 생성된 증강 데이터(310 내지 380) 중 증강 데이터(310)에 대한 경우를 예시한다. 도 4를 참조하면, 증강 데이터(310)으로부터 4개의 정상 증강 데이터(410 내지 440) 및 4개의 수도 증강 데이터(450 내지 480)가 생성될 수 있다. 이때, 정상 증강 데이터(410 내지 440)는 증강 데이터(310)를 제1 크기로 크롭하고 랜덤한 색 변환을 적용하여 생성될 수 있고, 수도 증강 데이터(450 내지 480)는 증강 데이터(310)를 제2 크기로 크롭하고 랜덤한 색 변환을 적용하여 생성될 수 있다. 예를 들어, 정상 증강 데이터(410 내지 440)는 각각 증강 데이터(310)의 좌측 상단, 우측 하단, 좌측 하단, 우측 상단을 기준으로 증강 데이터(310) 크기의 0.75 내지 1배의 크기를 가지도록 크롭될 수 있다. 또한, 수도 증강 데이터(450 내지 480)는 각각 증강 데이터(310)의 좌측 상단, 우측 하단, 좌측 하단, 우측 상단을 기준으로 증강 데이터(310) 크기의 0.25 내지 0.4배의 크기를 가지도록 크롭될 수 있다. 동일한 제2 데이터 증강 처리가 증강 데이터(320 내지 380)에 대해 수행될 수 있다. 도 4에서 설명된 데이터 크롭의 수치와 크롭되는 특정 위치에 관한 내용은 단지 설명의 편의를 위해 예시되었으며, 본 개시의 실시예는 이에 한정되지 않을 것이다.

[0069] 전술한 도 2 내지 4에서 설명된 데이터 증강 처리, 특징 추출, 특징 분리를 통해 네트워크의 학습이 완료되면, 학습된 네트워크를 이용하여 새로운 데이터에 대한 특징 추출과 이상 데이터 검출이 수행될 수 있다. 이하 도 5는 일 실시예에 따른 전자 장치에 의해 수행되는 이상 데이터 검출의 흐름도를 나타낸다. 도 5는 도 1에 도시된 전자 장치(100)의 동작 방법을 예시할 수 있다.

[0070] 단계 S501에서, 전자 장치는 제2 입력 데이터를 수신할 수 있다. 제2 입력 데이터는 도 2에서 설명된 네트워크의 학습을 위해 이용되는 입력 데이터와의 구별을 위한 명칭으로서, 학습 완료 후 이상 데이터 여부를 판별하기 위해 입력되는 새로운 입력 데이터를 의미할 수 있다.

[0071] 단계 S502에서, 전자 장치는 제2 입력 데이터에 대한 2차원 변환에 기초하여, 제2 입력 데이터에 대응하는 증강 데이터를 생성할 수 있다. 보다 구체적으로, 전자 장치는 전술한 네트워크 학습 과정에서 수행된 제1 데이터 증강 처리와 동일하게, 복수의 유형 각각에 따른 2차원 변환에 기초하여 제2 입력 데이터를 증강시킬 수 있다. 따라서, 전자 장치는 제2 입력 데이터에 대응하는 복수의 증강 데이터를 생성할 수 있고, 예를 들어, 도 3과 마찬가지로 제2 입력 데이터에 대응하는 8개의 증강 데이터가 생성될 수 있다.

[0072] 단계 S503에서, 전자 장치는 제2 입력 데이터에 대응하는 증강 데이터의 특징 벡터를 획득할 수 있다. 보다 구체적으로, 제2 입력 데이터에 대응하는 복수의 증강 데이터가 생성되면, 수학식 2 내지 4의 손실 함수를 최소화시키는 특징 벡터를 획득하도록 학습된 네트워크를 통해 이들 복수의 증강 데이터 각각의 특징 벡터가 획득될 수 있다.

[0073] 단계 S504에서, 전자 장치는 제2 입력 데이터에 대응하는 증강 데이터의 특징 벡터 및 2차원 변환의 중심점 간의 유사도로부터 정상치 점수를 획득할 수 있다. 보다 구체적으로, 전자 장치는 네트워크의 학습 과정에서 수학식 1을 통해 획득된 각 유형의 2차원 변환의 중심점과 단계 S503에서 획득된 복수의 증강 데이터의 특징 벡터 간의 유사도를 계산할 수 있다. 이와 같이 계산된 유사도에 대해 수학식 5에 의해 정의되는 소프트맥스(Softmax) 함수를 적용함으로써, m 번째 2차원 변환이 적용된 증강 데이터 x_m 가 2차원 변환의 각각의 유형 k 에

대응할 확률 P 가 계산될 수 있다. 수학식 5의 소프트맥스 함수에서, 분모는 증강 데이터 x_m 과 2차원 변환의 모든 유형의 유사도에 관한 함수이며, 분자는 증강 데이터 x_m 과 k 번째 2차원 변환의 유사도에 관한 함수를 의미할 수 있다.

[수학식 5]

$$P(k | x_m) = \frac{e^{f(x_m)^T c_k}}{\sum_l e^{f(x_m)^T c_l}}$$

또한, 전자 장치는 수학식 5에서 계산된 결과를 수학식 6에 적용하여 정상치 점수 S 를 획득할 수 있다. 즉, 전자 장치는 증강 데이터 x_m 가 m 번째 유형의 2차원 변환에 대응할 확률만을 log sum 함으로써 최종 정상치 점수를 계산할 수 있다.

[수학식 6]

$$S(x) = \sum_m \log P(m | x_m)$$

단계 S505에서, 전자 장치는 정상치 점수가 임계값보다 큰지 여부를 판단할 수 있다. 정상치 점수는 값이 클수록 해당 데이터가 정상 데이터일 확률이 높음을 의미할 수 있다. 따라서 정상 데이터는 사용된 2차원 변환과의 중심점과 큰 유사도를 가지므로, 확률 P 또한 큰 값을 가지며, 이에 따라 정상치 점수 S 도 큰 값을 가질 수 있다. 반면 이상 데이터는 사용된 2차원 변환과의 중심점과 낮은 유사도를 가지므로, 확률 P 또한 작은 값을 가지며, 이에 따라 정상치 점수 S 도 작은 값을 가질 수 있다.

따라서, 정상치 점수가 임계값보다 큰 경우, 단계 S506에서 전자 장치는 제2 입력 데이터를 정상 데이터로 판단할 수 있고, 정상치 점수가 임계값보다 크지 않은 경우, 단계 S507에서 전자 장치는 제2 데이터를 이상 데이터로 판단할 수 있다.

이와 같이 전자 장치는 제2 입력 데이터에 대응하는 증강 데이터의 특징 벡터 및 복수의 유형의 2차원 변환 각각의 중심점에 기초하여, 제2 입력 데이터가 정상 데이터인지 또는 이상 데이터인지 여부를 판단할 수 있으며, 제2 입력 데이터가 정상 데이터인지 또는 이상 데이터인지 여부는 제2 입력 데이터에 대응하는 증강 데이터의 특징 벡터 및 복수의 유형의 2차원 변환 각각의 중심점 간의 유사도로부터 획득되는 정상치 점수에 기초하여 판단될 수 있다.

전술한 본 개시의 실시예에 따르면, 하나의 정상 데이터에 대해 더 많은 특징 정보를 획득하기 위하여 데이터 증강이 수행될 수 있다. 따라서 하나의 정상 데이터로부터 증강된 개수만큼의 특징 벡터를 획득할 수 있으며, 이들의 특징 분리를 효과적으로 수행하기 위해 새로운 손실 함수가 정의될 수 있다. 이를 통해, 본 개시의 실시예는 부족한 양의 데이터를 갖는 상황에서도 이상 데이터의 검출을 위한 네트워크의 학습을 효과적으로 수행할 수 있다.

전술한 실시예들에 따른 전자 장치 또는 단말은, 프로세서, 프로그램 데이터를 저장하고 실행하는 메모리, 디스크 드라이브와 같은 영구 저장부(permanent storage), 외부 장치와 통신하는 통신 포트, 터치 패널, 키(key), 버튼 등과 같은 사용자 인터페이스 장치 등을 포함할 수 있다. 소프트웨어 모듈 또는 알고리즘으로 구현되는 방법들은 상기 프로세서상에서 실행 가능한 컴퓨터가 읽을 수 있는 코드들 또는 프로그램 명령들로서 컴퓨터가 읽을 수 있는 기록 매체 상에 저장될 수 있다. 여기서 컴퓨터가 읽을 수 있는 기록 매체로 마그네틱 저장 매체(예컨대, ROM(read-only memory), RAM(random-access memory), 플로피 디스크, 하드 디스크 등) 및 광학적 판독 매체(예컨대, 시디롬(CD-ROM), 디브이디(DVD: Digital Versatile Disc)) 등이 있다. 컴퓨터가 읽을 수 있는 기록 매체는 네트워크로 연결된 컴퓨터 시스템들에 분산되어, 분산 방식으로 컴퓨터가 판독 가능한 코드가 저장되고 실행될 수 있다. 매체는 컴퓨터에 의해 판독가능하며, 메모리에 저장되고, 프로세서에서 실행될 수 있다.

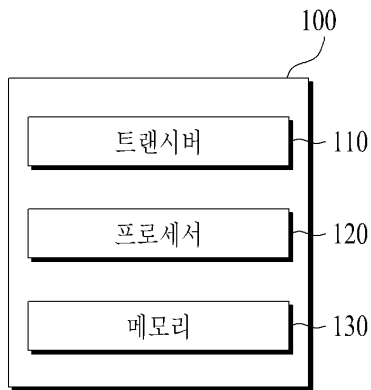
본 실시예는 기능적인 블록 구성들 및 다양한 처리 단계들로 나타내어질 수 있다. 이러한 기능 블록들은 특정

기능들을 실행하는 다양한 개수의 하드웨어 또는/및 소프트웨어 구성들로 구현될 수 있다. 예를 들어, 실시 예는 하나 이상의 마이크로프로세서들의 제어 또는 다른 제어 장치들에 의해서 다양한 기능들을 실행할 수 있는, 메모리, 프로세싱, 로직(logic), 룩 업 테이블(look-up table) 등과 같은 직접 회로 구성들을 채용할 수 있다. 구성 요소들이 소프트웨어 프로그래밍 또는 소프트웨어 요소들로 실행될 수 있는 것과 유사하게, 본 실시 예는 데이터 구조, 프로세스들, 루틴들 또는 다른 프로그래밍 구성들의 조합으로 구현되는 다양한 알고리즘을 포함하여, C, C++, 자바(Java), 어셈블러(assembler) 등과 같은 프로그래밍 또는 스크립팅 언어로 구현될 수 있다. 기능적인 측면들은 하나 이상의 프로세서들에서 실행되는 알고리즘으로 구현될 수 있다. 또한, 본 실시 예는 전자적인 환경 설정, 신호 처리, 및/또는 데이터 처리 등을 위하여 종래 기술을 채용할 수 있다. “매커니즘”, “요소”, “수단”, “구성”과 같은 용어는 넓게 사용될 수 있으며, 기계적이고 물리적인 구성들로서 한정되는 것은 아니다. 상기 용어는 프로세서 등과 연계하여 소프트웨어의 일련의 처리들(routines)의 의미를 포함할 수 있다.

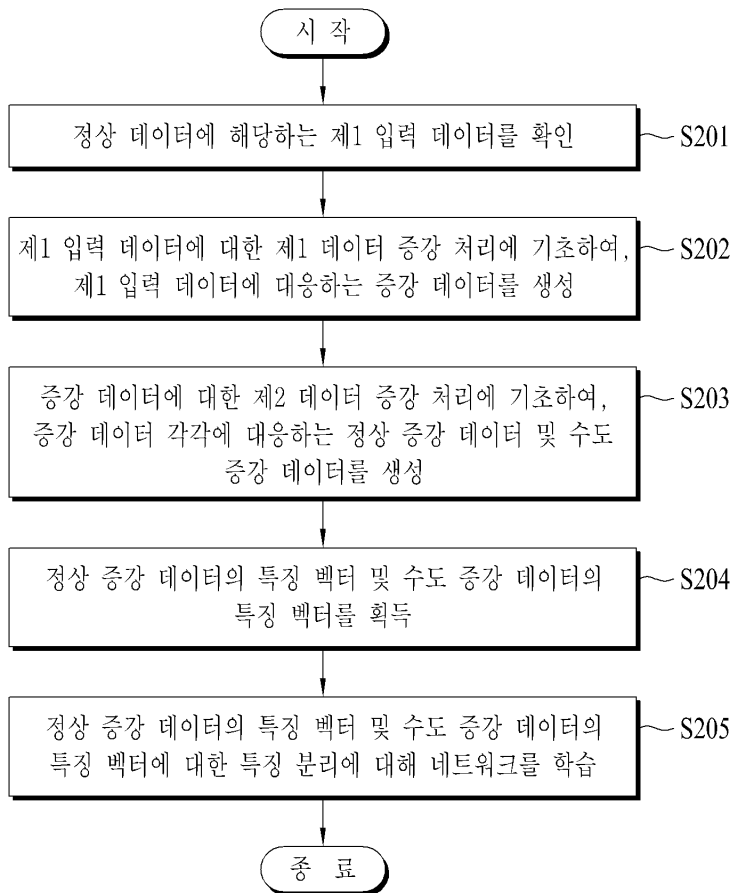
[0085] 전술한 실시예들은 일 예시일 뿐 후술하는 청구항들의 범위 내에서 다른 실시예들이 구현될 수 있다.

도면

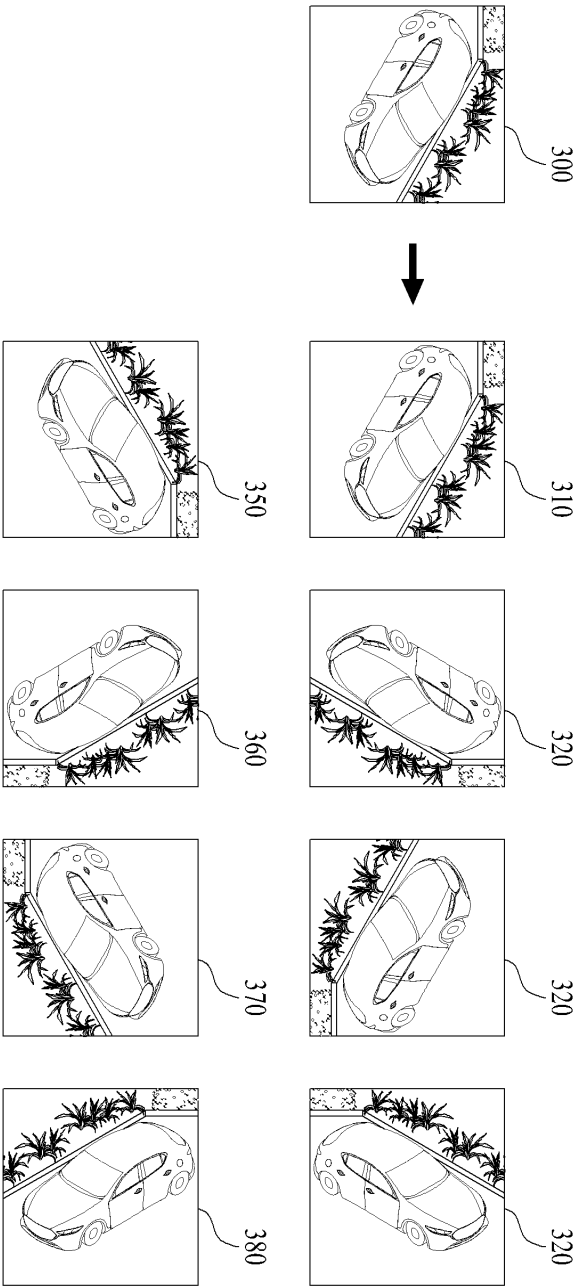
도면1



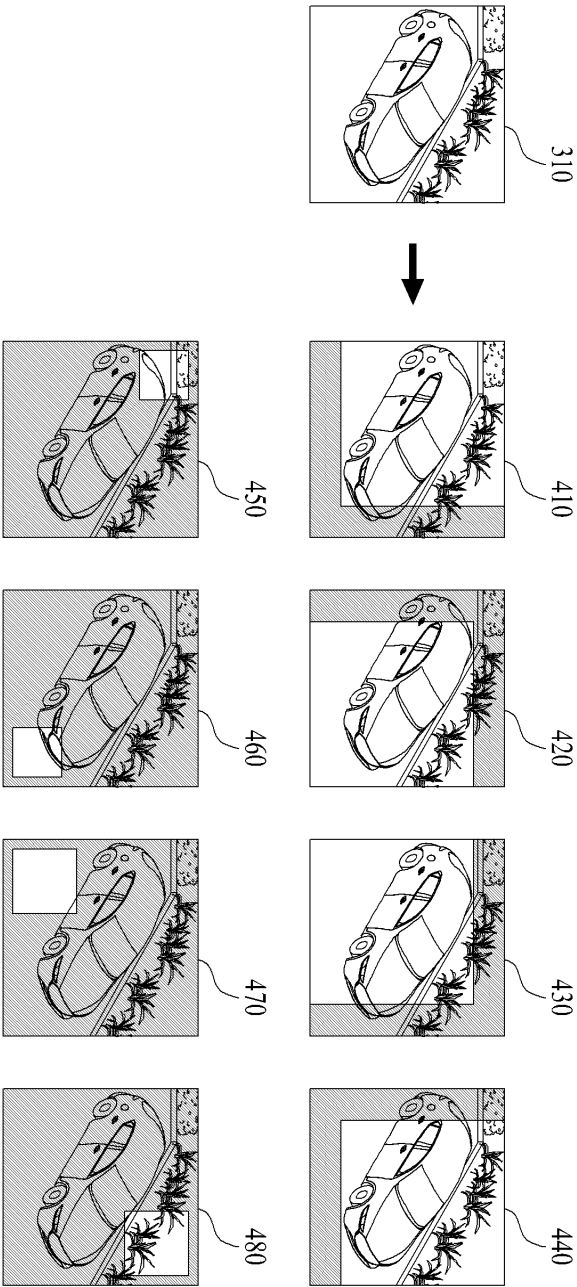
도면2



도면3



도면4



도면5

