(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0149600
(43) 공개일자 2024년10월15일

(51) 국제특허분류(Int. Cl.)
G06N 3/096 (2023.01) *G06N 3/0455* (2023.01)
G06N 3/0475 (2023.01)
(52) CPC특허분류
G06N 3/096 (2023.01)
G06N 3/0455 (2023.01)
(21) 출원번호 10-2023-0045340
(22) 출원일자 2023년04월06일
심사청구일자 2023년04월06일

(71) 출원인
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
(72) 발명자
최동완
인천광역시 미추홀구 용정공원로 33, 108동 1303호(용현동, 인천 SK Sky VIEW)
신현준
서울특별시 성북구 동소문로34길 24, 102동 2203호 (돈암동, 돈암동 삼성아파트)
(74) 대리인
양성보

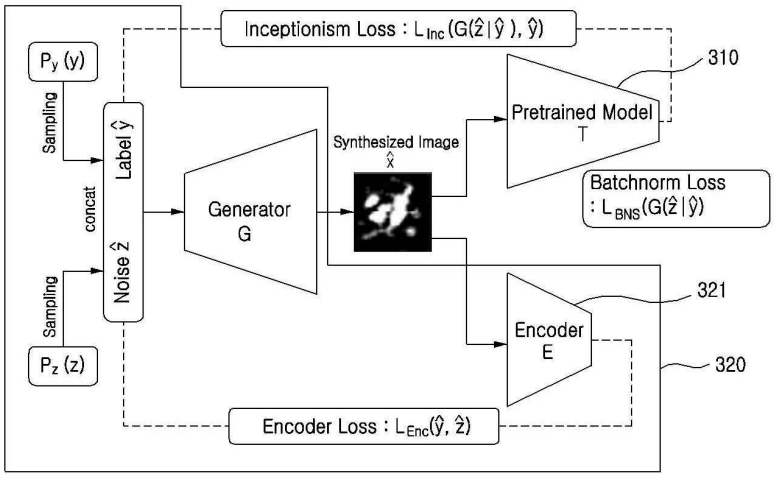
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 **데이터 없는 상황에서 생성형 신경망에 대한 지속학습 방법 및 시스템**

(57) 요 약

데이터 없는 상황에서 생성형 신경망에 대한 지속학습 방법 및 시스템이 제시된다. 본 발명에서 제안하는 데이터 없는 상황에서 생성형 신경망에 대한 지속학습 방법은 사전학습부를 통해 사전 학습된 추론형 신경망에서 데이터에 의존하지 않고 생성형 신경망을 학습하는 단계, 사전학습부를 통해 상기 사전 학습된 추론형 모델에서 데이터를 역으로 추적하고 내부 활성화의 통계값을 제약하는 단계, 상기 사전 학습된 추론형 모델을 이용하여 지속학습부를 통해 복수의 합성 이미지 및 노이즈에 대한 간건성을 위해 인코더 신경망을 추가하여 생성형 신경망을 학습하는 단계 및 지속학습부를 통해 파라미터 정규화 기법 및 재생 메모리 사용 기법을 적용하여 데이터에 의존하지 않고 생성형 신경망을 학습한 결과 및 인코더 신경망을 추가하여 생성형 신경망을 학습한 결과를 비교하는 단계를 포함한다.

대 표 도



(52) CPC특허분류
G06N 3/0475 (2023.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711193695
과제번호 2022-0-00448-002
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 방송통신산업기술개발사업
연구과제명 [Ezbaro] 인간처럼 회상이 가능한 인공 신경망 지속학습 플랫폼 개발(2차년도)
기 여 율 50/100
과제수행기관명 인하대학교 산학협력단
연구기간 2023.01.01 ~ 2023.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1711189477
과제번호 2021R1F1A1060160
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 기본연구
연구과제명 [Ezbaro] 질의 가능한 인공지능: 상호작용형 XAI 서비스를 실현하기 위한 질의 기반 딥러닝 모델 해석 기술
기 여 율 40/100
과제수행기관명 인하대학교 산학협력단
연구기간 2023.03.01 ~ 2024.02.29
이 발명을 지원한 국가연구개발사업
과제고유번호 1711179347
과제번호 00155915
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정보통신인재양성(교육운영수익)
연구과제명 [Ezbaro][정부+민간] 인공지능융합혁신인재양성(2차년도)
기 여 율 10/100
과제수행기관명 인하대학교 산학협력단
연구기간 2023.01.01 ~ 2023.12.31

명 세 서

청구범위

청구항 1

사전학습부를 통해 사전 학습된 추론형 신경망에서 데이터에 의존하지 않고 생성형 신경망을 학습하는 단계;

사전학습부를 통해 상기 사전 학습된 추론형 모델에서 데이터를 역으로 추적하고 내부 활성화의 통계값을 제약하는 단계;

상기 사전 학습된 추론형 모델을 이용하여 지속학습부를 통해 복수의 합성 이미지 및 노이즈에 대한 간건성을 위해 인코더 신경망을 추가하여 생성형 신경망을 학습하는 단계; 및

지속학습부를 통해 파라미터 정규화 기법 및 재생 메모리 사용 기법을 적용하여 데이터에 의존하지 않고 생성형 신경망을 학습한 결과 및 인코더 신경망을 추가하여 생성형 신경망을 학습한 결과를 비교하는 단계

를 포함하는 생성형 신경망에 대한 지속학습 방법.

청구항 2

제1항에 있어서,

상기 사전학습부를 통해 상기 사전 학습된 추론형 모델에서 데이터를 역으로 추적하고 내부 활성화의 통계값을 제약하는 단계는,

데이터 없는 지식 증류(Data-Free Knowledge Distillation; DFKD)를 통해 학습된 생성형 신경망에 대하여 복수의 합성 이미지가 자연 이미지를 모방하는 것에 대한 제약을 위해 입출력을 제약하는 손실 함수와 실제 데이터를 사용했을 때 생성형 신경망의 계층간 통계값과 합성된 데이터를 사용했을 때 생성형 신경망의 계층간 통계값 간의 손실을 정규화 손실 함수를 이용하여 복수의 합성 이미지를 위한 손실함수를 정의하고, 사전 학습된 추론형 모델의 배치 정규화 계층에서 수집된 내부 활성화의 통계값을 제약하는

생성형 신경망에 대한 지속학습 방법.

청구항 3

제1항에 있어서,

상기 사전 학습된 추론형 모델을 이용하여 지속학습부를 통해 복수의 합성 이미지 및 노이즈에 대한 간건성을 위해 인코더 신경망을 추가하여 생성형 신경망을 학습하는 단계는,

복수의 합성 이미지 및 노이즈에 대한 조건부 생성형 신경망을 학습하기 위해 다층 퍼셉트론 기반 인코더 신경망을 추가하여 생성형 신경망을 학습하고,

상기 인코더 신경망은 합성 이미지 복수의 합성 이미지 입력 받아 복수의 합성 이미지의 특징이 압축된 형태인 저차원의 변수를 추출하고, 정규 분포에서 임의 추출한 이미지와 인코더 신경망의 출력을 제약하는 손실 함수를 이용하여 상기 조건부 생성형 신경망을 학습하는

생성형 신경망에 대한 지속학습 방법.

청구항 4

제1항에 있어서,

상기 지속학습부를 통해 파라미터 정규화 기법 및 재생 메모리 사용 기법을 적용하여 데이터에 의존하지 않고 생성형 신경망을 학습한 결과 및 인코더 신경망을 추가하여 생성형 신경망을 학습한 결과를 비교하는 단계는,

상기 파라미터 정규화 기법을 이용하여 이전 태스크에서의 파라미터에 대한 업데이트를 제한하고,

상기 재생 메모리 사용 기법을 이용하여 이전 생성형 신경망에서 합성된 이미지와 현재 생성형 신경망에서 합성된 이미지 픽셀간 손실 함수를 정의하여 사용하는

생성형 신경망에 대한 지속학습 방법.

청구항 5

사전 학습된 추론형 신경망에서 데이터에 의존하지 않고 생성형 신경망을 학습하고, 사전 학습된 추론형 모델에서 데이터를 역으로 추적하고 내부 활성화의 통계값을 제약하는 사전학습부; 및

상기 사전 학습된 추론형 모델을 이용하여 복수의 합성 이미지 및 노이즈에 대한 간건성을 위해 인코더 신경망을 추가하여 생성형 신경망을 학습하고, 파라미터 정규화 기법 및 재생 메모리 사용 기법을 적용하여 데이터에 의존하지 않고 생성형 신경망을 학습한 결과 및 인코더 신경망을 추가하여 생성형 신경망을 학습한 결과를 비교하는 지속학습부

를 포함하는 생성형 신경망에 대한 지속학습 시스템.

청구항 6

제5항에 있어서,

상기 사전학습부는,

데이터 없는 지식 증류(Data-Free Knowledge Distillation; DFKD)를 통해 학습된 생성형 신경망에 대하여 복수의 합성 이미지가 자연 이미지를 모방하는 것에 대한 제약을 위해 입출력을 제약하는 손실 함수와 실제 데이터를 사용했을 때 생성형 신경망의 계층간 통계값과 합성된 데이터를 사용했을 때 생성형 신경망의 계층간 통계값 간의 손실을 정규화 손실 함수를 이용하여 복수의 합성 이미지를 위한 손실함수를 정의하고, 사전 학습된 추론형 모델의 배치 정규화 계층에서 수집된 내부 활성화의 통계값을 제약하는

생성형 신경망에 대한 지속학습 시스템.

청구항 7

제5항에 있어서,

상기 지속학습부는,

복수의 합성 이미지 및 노이즈에 대한 조건부 생성형 신경망을 학습하기 위해 다층 퍼셉트론 기반 인코더 신경망을 추가하여 생성형 신경망을 학습하고,

상기 인코더 신경망은 합성 이미지 복수의 합성 이미지 입력 받아 복수의 합성 이미지의 특징이 압축된 형태인 저차원의 변수를 추출하고, 정규 분포에서 임의 추출한 이미지와 인코더 신경망의 출력을 제약하는 손실 함수를 이용하여 상기 조건부 생성형 신경망을 학습하는

생성형 신경망에 대한 지속학습 시스템.

청구항 8

제5항에 있어서,

상기 지속학습부는,

상기 파라미터 정규화 기법을 이용하여 이전 태스크에서의 파라미터에 대한 업데이트를 제한하고,

상기 재생 메모리 사용 기법을 이용하여 이전 생성형 신경망에서 합성된 이미지와 현재 생성형 신경망에서 합성된 이미지 픽셀간 손실 함수를 정의하여 사용하는

생성형 신경망에 대한 지속학습 시스템.

발명의 설명

기 술 분 야

[0001] 본 발명은 데이터 없는 상황에서 생성형 신경망에 대한 지속학습 방법 및 시스템에 관한 것이다.

배 경 기 술

[0002] 최근 고도화된 딥러닝은 기계 학습의 발전을 가속화하고 있다. 하지만 딥러닝의 신경망 구조는 새로운 지식을 점진적으로 학습하는 경우, 과거 지식에 대한 성능이 급락하는 문제(예를 들어, 치명적 망각, 파괴적 망각(Catastrophic Forgetting))가 발생한다. 해당 문제를 해결 하기 위해 이미지 분류모델을 기반으로 활발한 연구가 진행 중이며, GAN의 등장 이후 과거 데이터를 재구성하는 형태로 확장되고 있다.

[0003] 최근에는 생성형 신경망의 치명적 망각을 방지하기 위한 연구(예를 들어, 생성형 신경망의 지속학습)가 주목 받고 있다. 생성형 신경망의 지속학습 연구는 각 태스크의 입력 데이터에 의존한다.

[0004] 하지만, 현실 세계에서 학습용 데이터는 프라이버시 문제와 한정된 메모리 자원과 같은 제한적인 상황으로 접근이 어려울 수 있다. 대표적인 예로 개인별 의학 데이터를 생각해 볼 수 있다. 이에 지속적으로 학습하는 생성형 모델이 특정 태스크에서 데이터에 접근할 수 없는 상황에 대한 해결 방안을 필요로 한다.

선행기술문헌

비특허문헌

[0005] (비특허문헌 0001) [1] A. Mordvintsev et al. Inceptionism: Going deeper into neural networks, https://research.googleblog.com/2015/06/ inceptionism-going-deeper-into-neural. html, 2015.

(비특허문헌 0002) [2] Liangchen Luo et al. Large-scale generative data-free distillation.arXiv preprint arXiv:2012.05578, 2020.

(비특허문헌 0003) [3] J. Kirkpatrick, et al., "Overcoming catastrophic forgetting in neural networks," Proceedings of the national academy of sciences 114.13 (2017): 3521-3526.

(비특허문헌 0004) [4] Wu, Chenshen, et al. "Memory replay gans: Learning to generate new categories without forgetting." Advances in Neural Information Processing Systems 31 (2018).

발명의 내용

해결하려는 과제

[0006] 본 발명이 이루고자 하는 기술적 과제는 신경망의 블랙박스 특성과 상대적으로 작은 메모리 크기를 고려하여 접근 불가능한 데이터를 대체하기 위해 각 태스크에서 사전 학습된 추론형 신경망을 입력으로 받아 데이터에 의존하지 않고 생성형 신경망을 지속학습하기 위한 방법 및 시스템(Data-Free Continual Learning of Generative Networks; DF'CLG)을 제안한다.

과제의 해결 수단

[0007] 일 측면에 있어서, 본 발명에서 제안하는 데이터 없는 상황에서 생성형 신경망에 대한 지속학습 방법은 사전학습부를 통해 사전 학습된 추론형 신경망에서 데이터에 의존하지 않고 생성형 신경망을 학습하는 단계, 사전학습부를 통해 상기 사전 학습된 추론형 모델에서 데이터를 역으로 추적하고 내부 활성화의 통계값을 제약하는 단계, 상기 사전 학습된 추론형 모델을 이용하여 지속학습부를 통해 복수의 합성 이미지 및 노이즈에 대한 간건성을 위해 인코더 신경망을 추가하여 생성형 신경망을 학습하는 단계 및 지속학습부를 통해 파라미터 정규화 기법 및 재생 메모리 사용 기법을 적용하여 데이터에 의존하지 않고 생성형 신경망을 학습한 결과 및 인코더 신경망을 추가하여 생성형 신경망을 학습한 결과를 비교하는 단계를 포함한다.

[0008] 상기 사전학습부를 통해 상기 사전 학습된 추론형 모델에서 데이터를 역으로 추적하고 내부 활성화의 통계값을 제약하는 단계는 데이터 없는 지식 증류(Data-Free Knowledge Distillation; DFKD)를 통해 학습된 생성형 신경망에 대하여 복수의 합성 이미지가 자연 이미지를 모방하는 것에 대한 제약을 위해 입출력을 제약하는 손실 함수와 실제 데이터를 사용했을 때 생성형 신경망의 계층간 통계값과 합성된 데이터를 사용했을 때 생성형 신경망의 계층간 통계값 간의 손실을 정규화 손실 함수를 이용하여 복수의 합성 이미지를 위한 손실함수를 정의하고, 사전 학습된 추론형 모델의 배치 정규화 계층에서 수집된 내부 활성화의 통계값을 제약한다.

[0009] 상기 사전 학습된 추론형 모델을 이용하여 지속학습부를 통해 복수의 합성 이미지 및 노이즈에 대한 간건성을 위해 인코더 신경망을 추가하여 생성형 신경망을 학습하는 단계는 복수의 합성 이미지 및 노이즈에 대한 조건부

생성형 신경망을 학습하기 위해 다층 퍼셉트론 기반 인코더 신경망을 추가하여 생성형 신경망을 학습하고, 상기 인코더 신경망은 합성 이미지 복수의 합성 이미지 입력 받아 복수의 합성 이미지의 특징이 압축된 형태인 저차원의 변수를 추출하고, 정규 분포에서 임의 추출한 이미지와 인코더 신경망의 출력을 제약하는 손실 함수를 이용하여 상기 조건부 생성형 신경망을 학습한다.

[0010] 상기 지속학습부를 통해 파라미터 정규화 기법 및 재생 메모리 사용 기법을 적용하여 데이터에 의존하지 않고 생성형 신경망을 학습한 결과 및 인코더 신경망을 추가하여 생성형 신경망을 학습한 결과를 비교하는 단계는 상기 파라미터 정규화 기법을 이용하여 이전 태스크에서의 파라미터에 대한 업데이트를 제한하고, 상기 재생 메모리 사용 기법을 이용하여 이전 생성형 신경망에서 합성된 이미지와 현재 생성형 신경망에서 합성된 이미지 픽셀 간 손실 함수를 정의하여 사용한다.

[0011] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 데이터 없는 상황에서 생성형 신경망에 대한 지속학습 시스템은 사전 학습된 추론형 신경망에서 데이터에 의존하지 않고 생성형 신경망을 학습하고, 사전 학습된 추론형 모델에서 데이터를 역으로 추적하고 내부 활성화의 통계값을 제약하는 사전학습부 및 상기 사전 학습된 추론형 모델을 이용하여 복수의 합성 이미지 및 노이즈에 대한 간건성을 위해 인코더 신경망을 추가하여 생성형 신경망을 학습하고, 파라미터 정규화 기법 및 재생 메모리 사용 기법을 적용하여 데이터에 의존하지 않고 생성형 신경망을 학습한 결과 및 인코더 신경망을 추가하여 생성형 신경망을 학습한 결과를 비교하는 지속학습부를 포함한다.

발명의 효과

[0012] 본 발명의 실시예들에 따르면 데이터 없는 상황에서 생성형 신경망에 대한 지속학습 방법 및 시스템을 통해 신경망의 블랙박스 특성과 상대적으로 작은 메모리 크기를 고려하여 접근 불가능한 데이터를 대체하기 위해 각 태스크에서 사전 학습된 추론형 신경망을 입력으로 받아 데이터에 의존하지 않고 생성형 신경망을 지속학습할 수 있다.

도면의 간단한 설명

[0013] 도 1은 본 발명의 일 실시예에 따른 DF'CLG 모델 구성도이다.

도 2는 본 발명의 일 실시예에 따른 데이터 없는 상황에서 생성형 신경망에 대한 지속학습 방법을 설명하기 위한 흐름도이다.

도 3은 본 발명의 일 실시예에 따른 데이터 없는 상황에서 생성형 신경망에 대한 지속학습 시스템의 구성을 나타내는 도면이다.

도 4는 본 발명의 일 실시예에 따른 MNIST 5개 태스크 학습 후 합성된 이미지를 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0014] 딥러닝의 치명적 망각문제를 해결하기 위한 지속학습 연구는 최근 생성형 신경망으로 확장되고 있다. 하지만 생성형 신경망의 지속학습은 태스크별 입력 데이터에 의존하며, 이는 현실 세계에서 프라이버시 문제와 한정된 메모리 자원에 따른 제약이 존재한다. 즉, 신경망의 블랙박스 특성과 상대적으로 작은 메모리 크기를 고려하여 접근 불가능한 데이터를 대체할 필요가 있다.

[0015] 따라서, 본 발명에서는 각 태스크에서 사전 학습된 추론형 신경망을 입력으로 받아 데이터에 의존하지 않고 생성형 신경망을 지속학습하는 모델(Data-Free Continual Learning of Generative Networks; DF'CLG)을 제안한다.

[0016] 더욱 상세하게는, 사전 학습된 추론형 모델에서 데이터를 역으로 추적하고 내부 활성화를 제약하는 방법과 합성 이미지의 다양성을 위한 인코더 신경망을 추가하여 생성형 신경망을 학습하는 방법을 제안한다. 이후, 파라미터 정규화 기법과 재생 메모리 방법을 각각 적용하여 제안하는 생성형 모델의 치명적 망각현상이 방지될 수 있음을 보인다. 본 발명의 실시예에 따르면, 상대적으로 저조한 성능으로 알려진 파라미터 정규화 기법이 데이터 없이 학습된 생성형 신경망에서 재생 메모리 방법과 준하는 성능을 보이는 것을 확인할 수 있다. 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0018] 도 1은 본 발명의 일 실시예에 따른 DF'CLG 모델 구성도이다.

[0019] 본 발명의 실시예에 따르면, 모든 태스크 데이터에 접근할 수 없는 상황을 가정하고 생성형 모델의 지속학습에 대한 방법을 제안한다.

[0020] 본 발명은 신경망의 블랙박스 특성과 상대적으로 작은 메모리 크기를 고려하여 접근 불가능한 데이터를 대체하여 사용한다. 즉, 각 태스크에서 사전 학습된 추론형 신경망을 입력 받아 데이터에 의존하지 않고 생성형 신경망을 지속학습하는 모델(Data-Free Continual Learning of Generative Networks; DF'CLG)을 제안한다.

[0021] 먼저, 하나의 사전 학습된 추론형 신경망에서 데이터에 의존하지 않고 생성형 신경망을 학습하는 방법을 고려하고, 제안하는 생성형 신경망의 치명적 망각현상을 완화하는 방안을 제안한다.

[0022] 하나의 사전 학습된 추론형 신경망에서 데이터에 의존하지 않고 생성형 신경망을 학습하기 위해 데이터 없는 지식 증류(Data-Free Knowledge Distillation; DFKD)에서 생성형 신경망을 학습하는 방법을 사용한다.

[0023] 더욱 상세하게는, 사전 학습된 추론형 모델에서 데이터를 역으로 추적하고 내부 활성화의 통계값을 제약하는 방법과 합성 이미지의 다양성과 노이즈에 대한 간건성을 위해 인코더 신경망을 추가하여 생성형 신경망을 학습하는 방법을 제안한다. 이후, 지속학습을 위한 파라미터 정규화 기법(EWC[3]) 및 재생 메모리 사용 기법(MERGAN[4])을 각각 적용하여 그 결과를 비교한다.

[0024] 본 발명의 실시예에 따르면, 상대적으로 저조한 성능으로 알려진 파라미터 정규화 기법이 데이터 없이 학습된 생성형 신경망에서 재생 메모리 방법과 준하는 성능을 보이는 것을 확인할 수 있다.

[0025] 본 발명의 실시예에 따른 데이터 없는 지식 증류 (Data-Free Knowledge Distillation; DFKD)는 실제 데이터에 접근할 수 없을 때 사용하는 지식 증류(Knowledge Distillation; KD) 방법으로, 사전 학습된 추론형 신경망에서 합성 데이터를 생성하여 지식 증류에 활용한다.

[0026] 합성 데이터를 위한 대표적인 방법으로 Inceptionism[1]은 사전 학습된 신경망의 입출력을 제약하여 합성 이미지를 추적한다. 예측 레이블 $\hat{y}$과 사전 학습된 신경망 T이 주어질 때, $\hat{p} = OneHot(\hat{y})$와 $p = T(\hat{x})$사이 크로스-엔트로피 손실 함수를 최소화하는 이미지 $\hat{x}$를 찾는 것이다. 크로스-엔트로피 손실 함수는 식(1)과 같이 나타낸다:

[0027] $$L_{CE}(\hat{x},\hat{y}) = H(\hat{p},p) = -\sum_i \hat{p}_i \log p_i \quad \text{식(1)}$$

[0028] 또한, 합성 이미지가 자연 이미지를 모방하는 제약을 위해, 합성 이미지의 총 변동 손실 (L_{tv})과 L2 손실 (L_{l_2})을 식(2)와 같이 정의하고 이를 정규화 손실 함수로 추가한다. 정규화 손실 함수는 식(3)과 같이 나타낼 수 있다:

[0029] $$L_{Inc}(\hat{x},\hat{y}) = L_{CE}(\hat{x},\hat{y}) + L_{Reg}(\hat{x}) \quad \text{식(2)}$$

[0030] $$L_{Reg}(\hat{x}) = \lambda_{tv} L_{tv}(\hat{x}) + \lambda_{l_2} L_{l_2}(\hat{x}) \quad \text{식(3)}$$

[0031] 하지만 Inceptionism[1] 손실 함수 (L_{Inc})는 오직 입출력을 제약하는 식으로, 실제 학습한 데이터와 다른 비정상적인 내부 활성화가 생길 수 있다. 이를 해결하기 위해 사전 학습된 신경망의 배치 정규화 계층에서 수집된 통계값(다시 말해, 평균과 표준 편차)을 활용한다. 즉, 실제 데이터를 사용했을 때 신경망의 계층간 통계값 $(\hat{\mu}_l \ , \ \hat{\sigma}_l^2)$과 합성된 데이터를 사용했을 때 신경망의 계층간 통계값$(\mu(\hat{x}), \sigma_l^2(\hat{x}))$간 L2 손실을 정규화 손실 함수로 추가하고 식(4)와 같이 나타낼 수 있다. 모든 λ는 손실 함수의 크기를 조절하는 값이다.

[0032] $$L_{BNS}(\hat{x}) = \lambda_S \sum_l [\, \| \mu_l(\hat{x}) - \hat{\mu}_l \|_2 + \| \sigma_l^2(\hat{x}) - \hat{\sigma}_l^2 \|_2 \,] \quad \text{식(4)}$$

[0033] 결과적으로 입출력을 제약하는 $L_{Inc}(\hat{x},\hat{y})$와 배치 정규화 계층의 통계값을 제약한 $L_{BNS}(\hat{x})$를 합쳐 이미지

$\hat{x}$를 위한 손실 함수를 식(5)와 같이 정의할 수 있다:

[0034] $$L_{Image}(\hat{x},\hat{y}) = L_{Inc}(\hat{x},\hat{y}) + L_{BNS}(\hat{x})$$ 식(5)

[0035] 이때 이미지 $\hat{x}$를 조건부 생성형 신경망을 통해 합성하기 위해 $\hat{x} = G(z|\hat{y})$로 변환한다(LS-GDFD[2]). 따라서, 조건부 생성형 신경망 학습에 사용된 최종 손실 함수는 식(6)과 같이 나타낼 수 있다:

[0036] $$L_G = E_{z\sim p_z(z),\hat{y}\sim p_y(y)}[L_{Inc}(G(z|\hat{y}),\hat{y}) + L_{BNS}(G(z|\hat{y}))]$$ 식(6)

[0037] 본 발명의 실시예에 따른 생성 모델의 지속 학습(Continual Learning of Generative Networks)이 생성형 신경망의 망각을 완화하기 위해 사용된다.

[0038] 첫째는 파라미터 정규화 방법(EWC[3])으로 이전 태스크에서 중요했던 파라미터의 업데이트를 제한한다. 이전 태스크가 학습이 완료된 시점에서 데이터의 사후 분포를 활용하여, 중요 파라미터를 피셔 정보 행렬의 대각 성분 F_i으로 정의하고 파라미터의 평균 $\theta^*_{A,i}$과 멀어지지 않게 하는 손실 함수를 정의하여 현재 태스크를 위한 손실 함수 $L_B(\theta)$에 추가하여 사용하고, 이를 식(7)과 같이 나타낼 수 있다:

[0039] $$L_{EWC}(\theta) = L_B(\theta) + \sum_i \frac{\lambda}{2} F_i(\theta_i - \theta^*_{A,i})^2$$ 식(7)

[0040] 둘째는 생성형 신경망의 지속학습 분야에서 우수한 성능으로 알려진 재생 메모리를 활용하는 방법이다. 그 중, Replay Alignment는 MERGAN[4]에서 소개된 방법으로, 이전 생성형 신경망에서 합성된 이미지와 현재 생성형 신경망에서 합성된 이미지 픽셀간 L2 거리 손실을 정의하여 사용하고, 이를 식(8)과 같이 나타낼 수 있다:

[0041] $$L_{RA}(\theta_t) = E_{z\sim P_Z, c\sim U\{1,t-1\}}[\| G_{\theta_t^G}(z,c) - G_{\theta_{t-1}^G}(z,c) \|_2]$$ 식(8)

[0043] 도 2는 본 발명의 일 실시예에 따른 데이터 없는 상황에서 생성형 신경망에 대한 지속학습 방법을 설명하기 위한 흐름도이다.

[0044] 본 발명의 실시예에 따른 데이터 없는 상황에서 생성형 신경망에 대한 지속학습 방법은 사전학습부를 통해 사전 학습된 추론형 신경망에서 데이터에 의존하지 않고 생성형 신경망을 학습하는 단계(210), 사전학습부를 통해 상기 사전 학습된 추론형 모델에서 데이터를 역으로 추적하고 내부 활성화의 통계값을 제약하는 단계(220), 상기 사전 학습된 추론형 모델을 이용하여 지속학습부를 통해 복수의 합성 이미지 및 노이즈에 대한 간건성을 위해 인코더 신경망을 추가하여 생성형 신경망을 학습하는 단계(230) 및 지속학습부를 통해 파라미터 정규화 기법 및 재생 메모리 사용 기법을 적용하여 데이터에 의존하지 않고 생성형 신경망을 학습한 결과 및 인코더 신경망을 추가하여 생성형 신경망을 학습한 결과를 비교하는 단계(240)를 포함한다.

[0045] 본 발명의 실시예에 따르면, 각 태스크에서 사전 학습된 추론형 신경망이 주어질 때, 데이터에 의존하지 않고 생성형 신경망을 지속학습하여, 추론형 신경망 학습에 사용된 모든 데이터를 재합성한다.

[0046] 먼저, 태스크 t에서 주어진 데이터 $D_t = (X_t, Y_t) = (x_t^i, y_t^i)_{i=1}^{n_t}$($x_t^i \in X_t$와 $y_t^i \in Y_t$는 각각 i번째 이미지와 해당 클래스, n_t는 태스크 t의 데이터 수)로 사전 학습된 추론형 신경망 M_t이 입력으로 주어진다. 이때 조건부 생성형 신경망 G_t은 오직 사 전 학습된 신경망 M_t만을 사용하여, 학습에 사용된 데이터 D_t를 합성해낼 수 있도록 학습된다. 모든 태스크를 학습한 조건부 생성형 신경망 $G_{T(0:t)}$은 이전까지 학습에 사용된 모든 태 스크의 데이터 $(X')_{j=0}^t$를 재생성한다.

[0047] 단계(210)에서, 사전학습부를 통해 사전 학습된 추론형 신경망에서 데이터에 의존하지 않고 생성형 신경망을 학습한다.

[0048] 단계(220)에서, 사전학습부를 통해 상기 사전 학습된 추론형 모델에서 데이터를 역으로 추적하고 내부 활성화의 통계값을 제약한다.

[0049] 본 발명의 실시예에 따르면, 데이터 없는 지식 증류(Data-Free Knowledge Distillation; DFKD)를 통해 학습된 생성형 신경망에 대하여 복수의 합성 이미지가 자연 이미지를 모방하는 것에 대한 제약을 위해 입출력을 제약하는 손실 함수와 실제 데이터를 사용했을 때 생성형 신경망의 계층간 통계값과 합성된 데이터를 사용했을 때 생성형 신경망의 계층간 통계값 간의 손실을 정규화 손실 함수를 이용하여 복수의 합성 이미지를 위한 손실함수를 정의하고, 사전 학습된 추론형 모델의 배치 정규화 계층에서 수집된 내부 활성화의 통계값을 제약한다.

[0050] 단계(230)에서, 상기 사전 학습된 추론형 모델을 이용하여 지속학습부를 통해 복수의 합성 이미지 및 노이즈에 대한 간건성을 위해 인코더 신경망을 추가하여 생성형 신경망을 학습한다.

[0051] 본 발명의 실시예에 따르면, 복수의 합성 이미지 및 노이즈에 대한 조건부 생성형 신경망을 학습하기 위해 다층 퍼셉트론 기반 인코더 신경망을 추가하여 생성형 신경망을 학습한다.

[0052] 본 발명의 실시예에 따르면, 상기 인코더 신경망은 합성 이미지 복수의 합성 이미지 입력 받아 복수의 합성 이미지의 특징이 압축된 형태인 저차원의 변수를 추출하고, 정규 분포에서 임의 추출한 이미지와 인코더 신경망의 출력을 제약하는 손실 함수를 이용하여 상기 조건부 생성형 신경망을 학습한다.

[0053] 본 발명의 실시예에 따르면, 사전 학습된 추론형 신경망에서 데이터에 의존하지 않고 조건부 생성형 신경망을 학습하기 위해 LS-GDFD[2]를 활용한다. 하지만 기존 LS-GDFD[2]는 합성된 이미지가 모두 비슷해지는 모드 파괴(Mode Collapse) 문제가 존재한다. 따라서, 본 발명에서는 합성된 이미지의 다양성을 확장하고 노이즈에 강건한 조건부 생성형 신경망을 학습하기 위해 다층 퍼셉트론 기반 인코더 신경망 E를 추가한다. 인코더 신경망은 합성 이미지 $\hat{x} = G(\hat{z}|\hat{y})$를 입력 받아 저차원의 변수 $\bar{z}$를 추출하고, 이는 합성 이미지 $\hat{x}$의 특징이 압축된 형태로 볼 수 있다. 따라서 정규 분포에서 임의 추출한 $\hat{z}$와 인코더의 출력 $\bar{z} = E(G(\hat{z}|\hat{y}))$을 L2 거리 손실로 제약하는 손실 함수 L_{Enc}를 추가하여 조건부 생성형 신경망과 인코더를 학습한다. 손실 함수 L_{Enc}는 식(9)와 같이 나타낼 수 있다:

[0054] $$L_{Enc}(\hat{y},\hat{z}) = \| \hat{z} - E(G(\hat{y},\hat{z})) \|_2 \quad \text{식(9)}$$

[0055] 최종적으로, 사전 학습된 추론형 신경망에서 데이터 없이 조건부 생성형 신경망을 학습하기 위한 손실 함수는 식(10)과 같이 나타낼 수 있다:

[0056] $$L_G = E_{\hat{z} \sim p_z(z),\, \hat{y} \sim p_y(y)} [L_{Inc}(G(\hat{z}|\hat{y}),\hat{y}) + L_{BNS}(G(\hat{z}|\hat{y})) + L_{Enc}(\hat{y},\hat{z})] \quad \text{식(10)}$$

[0057] 추가적으로 생성형 신경망의 치명적 망각을 완화하기 위해 지속학습의 대표 방법인 파라미터 정규화 기법(EWC[3])과 재생 메모리 사용 방법(Replay Alignment)을 각각 적용한다.

[0058] 단계(240)에서, 지속학습부를 통해 파라미터 정규화 기법 및 재생 메모리 사용 기법을 적용하여 데이터에 의존하지 않고 생성형 신경망을 학습한 결과 및 인코더 신경망을 추가하여 생성형 신경망을 학습한 결과를 비교한다.

[0059] 본 발명의 실시예에 따르면, 상기 파라미터 정규화 기법을 이용하여 이전 태스크에서의 파라미터에 대한 업데이트를 제한한다.

[0060] 본 발명의 실시예에 따르면, 상기 재생 메모리 사용 기법을 이용하여 이전 생성형 신경망에서 합성된 이미지와 현재 생성형 신경망에서 합성된 이미지 픽셀간 손실 함수를 정의하여 사용한다.

[0062] 도 3은 본 발명의 일 실시예에 따른 데이터 없는 상황에서 생성형 신경망에 대한 지속학습 시스템의 구성을 나타내는 도면이다.

[0063] 제안하는 데이터 없는 상황에서 생성형 신경망에 대한 지속학습 시스템은 사전학습부(310) 및 지속학습부(320)

를 포함한다.

[0064] 본 발명의 실시예에 따른 사전학습부(310)는 사전 학습된 추론형 신경망에서 데이터에 의존하지 않고 생성형 신경망을 학습하고, 사전 학습된 추론형 모델에서 데이터를 역으로 추적하고 내부 활성화의 통계값을 제약한다.

[0065] 본 발명의 실시예에 따른 사전학습부(310)는 데이터 없는 지식 증류(Data-Free Knowledge Distillation; DFKD)를 통해 학습된 생성형 신경망에 대하여 복수의 합성 이미지가 자연 이미지를 모방하는 것에 대한 제약을 위해 입출력을 제약하는 손실 함수와 실제 데이터를 사용했을 때 생성형 신경망의 계층간 통계값과 합성된 데이터를 사용했을 때 생성형 신경망의 계층간 통계값 간의 손실을 정규화 손실 함수를 이용하여 복수의 합성 이미지를 위한 손실함수를 정의하고, 사전 학습된 추론형 모델의 배치 정규화 계층에서 수집된 내부 활성화의 통계값을 제약한다.

[0066] 본 발명의 실시예에 따른 지속학습부(320)는 상기 사전 학습된 추론형 모델을 이용하여 복수의 합성 이미지 및 노이즈에 대한 간건성을 위해 인코더 신경망(321)을 추가하여 생성형 신경망을 학습하고, 파라미터 정규화 기법 및 재생 메모리 사용 기법을 적용하여 데이터에 의존하지 않고 생성형 신경망을 학습한 결과 및 인코더 신경망을 추가하여 생성형 신경망을 학습한 결과를 비교한다.

[0067] 본 발명의 실시예에 따른 지속학습부(320)는 복수의 합성 이미지 및 노이즈에 대한 조건부 생성형 신경망을 학습하기 위해 다층 퍼셉트론 기반 인코더 신경망을 추가하여 생성형 신경망을 학습한다.

[0068] 상기 인코더 신경망(321)은 합성 이미지 복수의 합성 이미지 입력 받아 복수의 합성 이미지의 특징이 압축된 형태인 저차원의 변수를 추출하고, 정규 분포에서 임의 추출한 이미지와 인코더 신경망의 출력을 제약하는 손실 함수를 이용하여 상기 조건부 생성형 신경망을 학습한다.

[0069] 본 발명의 실시예에 따른 지속학습부(320)는 상기 파라미터 정규화 기법을 이용하여 이전 태스크에서의 파라미터에 대한 업데이트를 제한한다.

[0070] 본 발명의 실시예에 따른 지속학습부(320)는 상기 재생 메모리 사용 기법을 이용하여 이전 생성형 신경망에서 합성된 이미지와 현재 생성형 신경망에서 합성된 이미지 픽셀간 손실 함수를 정의하여 사용한다.

[0072] 도 4는 본 발명의 일 실시예에 따른 MNIST 5개 태스크 학습 후 합성된 이미지를 나타내는 도면이다.

[0073] 본 발명의 일 실시예에서, MNIST 데이터셋을 5개의 태스크(0,1/2,3/4,5/6,7/8,9)로 나누어 학습한 이미지 분류 신경망이 입력으로 주어질 때 조건부 생성형 신경망을 학습하였다. 추론형 신경망으로 Lenet5에 배치 정규화 계층을 추가하였으며, 생성형 신경망으로 DCGAN 구조를 사용하였다. 지속학습 방법의 효과를 비교하기 위해, 단순히 연속적으로 학습한 경우(Fine-tuning)와 지속학습 방법(EWC, Replay Alignment)을 적용한 경우를 각각 시각화하였다. 도 4(a)는 Fine-tuning을 적용하여 학습한 결과를 나타내는 도면이고, 도 4(b)는 EWC를 적용하여 학습한 결과를 나타내는 도면이고, 도 4(c)는 Replay Alignment를 적용하여 학습한 결과를 나타내는 도면이고, 도 4(b)는 EWC + Replay Alignment를 적용하여 학습한 결과를 나타내는 도면이다.

[0074] 도 4(a)의 경우 마지막 태스크인 8,9만 생성하여 치명적 망각현상이 존재하였고, 도 4(b) 및 도 4(c)의 경우는 비교적 이전의 성능을 유지하였다.

[0075] 학습된 생성형 신경망의 정량적 비교를 위해 Fitting Capacity(합성된 데이터로 분류기를 학습했을 때, 분류기의 실 제 데이터 정확도)를 10회 반복 측정하였고, 결과는 표1과 같다.

[0076] <표 1>

[0077]

	MNIST (5 Tasks)			
	Fine-tuning	EWC	RA	EWC+RA
Acc (%)	43.18 (±2.78)	63.23 (±1.09)	67.45 (±0.78)	70.43 (±0.86)

[0078] 재생 메모리를 사용하였을 때 성능이 67.45%로 파라미터 정규화 기법을 적용한 성능보다 우수했고, 두 방법을 동시에 사용했을 때 성능이 70.43%로 가장 우수했다. 주목할 점은, 상대적으로 저조 한 성능으로 알려진 파라미터 정규화 기법(EWC[3])이 제안한 생성형 신경망에서 재생 메모리 방법(Replay Alignment)과 준하는 성능

(63.2%)을 보였다는 것이다.

[0079] 본 발명의 실시예에 따르면, 생성형 신경망의 지속학습이 입력 데이터에 의존하는 한계를 해결하기 위해, 신경망의 블랙박스 특성과 상대적으로 작은 메모리 크기를 고려하여 접근 불가능한 데이터를 대체하였다. 즉, 사전 학습된 추론형 신경망 모델에서 생성형 신경망을 연속적으로 학습하는 모델(DF'CLG)을 제안하였다. 이를 위해 인코더 신경망을 추가하여 다양성과 노이즈에 강건 하도록 생성형 신경망을 학습하였고, 파라미터 정규화 기법과 재생 메모리 기법을 각각 적용하여, 제안한 생성형 신경망의 치명적 망각현상이 완화될 수 있음을 보였다. 또한, 저조한 성능으로 알려진 파라미터 정규화 기법이 데이터 없이 학습된 생 성형 신경망에서 재생 메모리 방법과 준하는 성능을 보임을 확인할 수 있다.

[0081] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0082] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

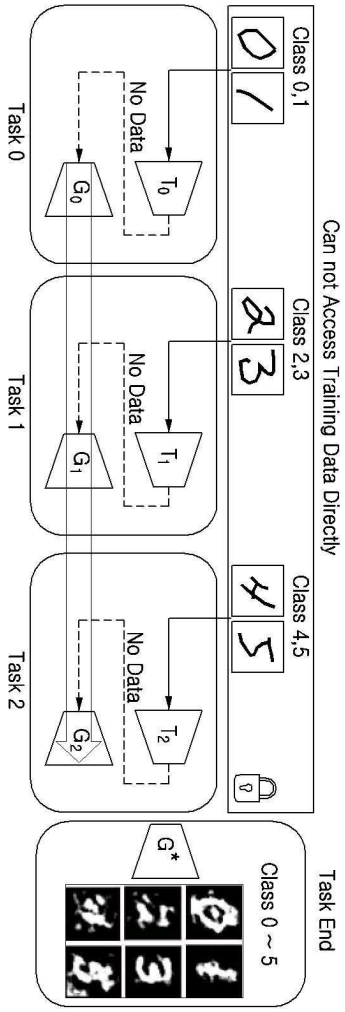
[0083] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0084] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

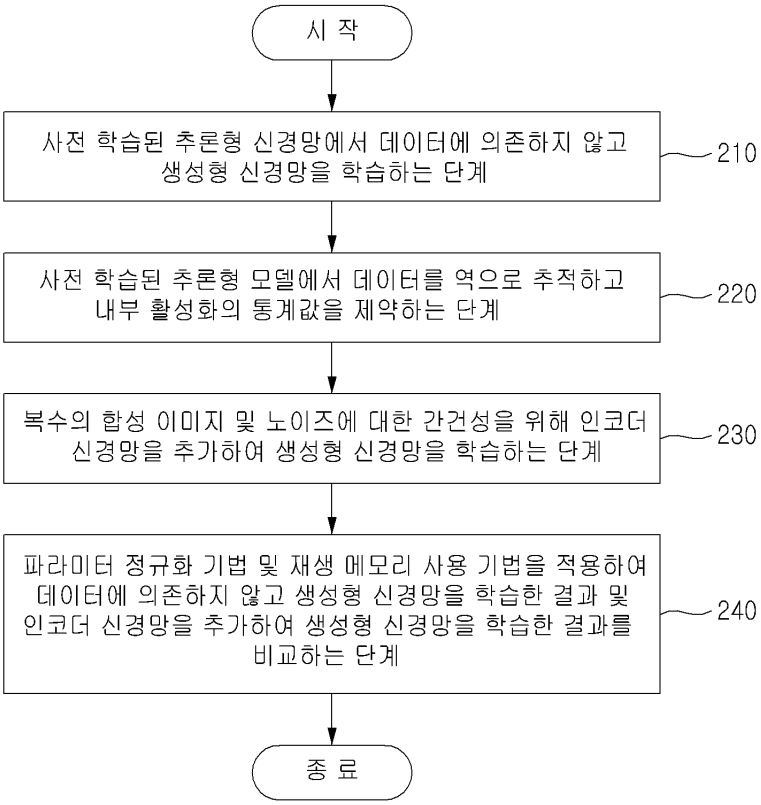
[0085] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

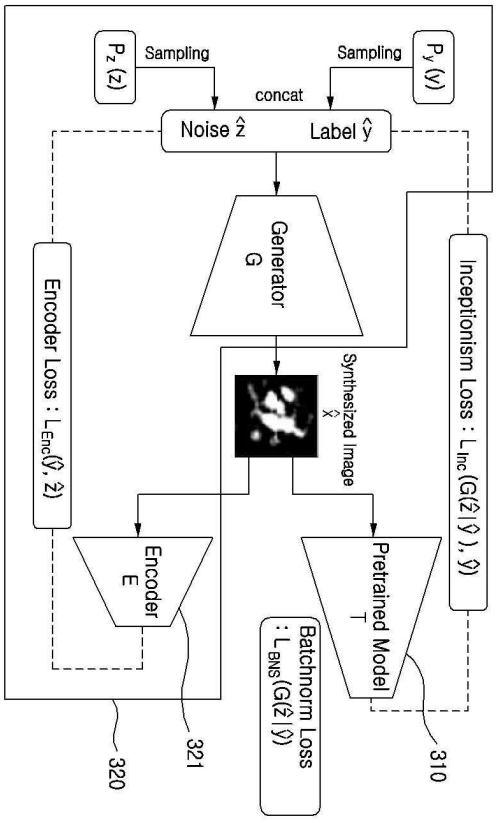
도면1


Task 0
Task 1
Task 2
Task End
No Data
Class 0,1
Class 2,3
Class 4,5
Class 0 ~ 5
Can not Access Training Data Directly

도면2


시 작
사전 학습된 추론형 신경망에서 데이터에 의존하지 않고 생성형 신경망을 학습하는 단계
210
사전 학습된 추론형 모델에서 데이터를 역으로 추적하고 내부 활성화의 통계값을 제약하는 단계
220
복수의 합성 이미지 및 노이즈에 대한 간건성을 위해 인코더 신경망을 추가하여 생성형 신경망을 학습하는 단계
230
파라미터 정규화 기법 및 재생 메모리 사용 기법을 적용하여 데이터에 의존하지 않고 생성형 신경망을 학습한 결과 및 인코더 신경망을 추가하여 생성형 신경망을 학습한 결과를 비교하는 단계
240
종 료

도면3


Sampling
Sampling
concat
Noise ẑ
Label ŷ
Generator G
Synthesized Image
Encoder Loss : L_Enc(ŷ, ẑ)
Inceptionism Loss : L_Inc(G(ẑ|ŷ), ŷ)
Encoder E
Pretrained Model T
Batchnorm Loss : L_BNS(G(ẑ|ŷ))
320
321
310

도면4

