



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0077845
(43) 공개일자 2023년06월02일

(51) 국제특허분류(Int. Cl.)
G06N 3/04 (2023.01) G06N 3/08 (2023.01)
G06N 5/02 (2023.01)
(52) CPC특허분류
G06N 3/045 (2023.01)
G06N 3/08 (2023.01)
(21) 출원번호 10-2021-0165004
(22) 출원일자 2021년11월26일
심사청구일자 2021년11월26일

(71) 출원인
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
(72) 발명자
송병철
서울특별시 양천구 목동서로 38, 125동 307호(목동, 목동신시가지아파트1단지)
이승현
경기도 부천시 상일로 71, 1803동 906호(상동, 반달마을)
이성욱
경기도 수원시 권선구 권선로682번길 30, B01호(권선동)
(74) 대리인
양성보

전체 청구항 수 : 총 7 항

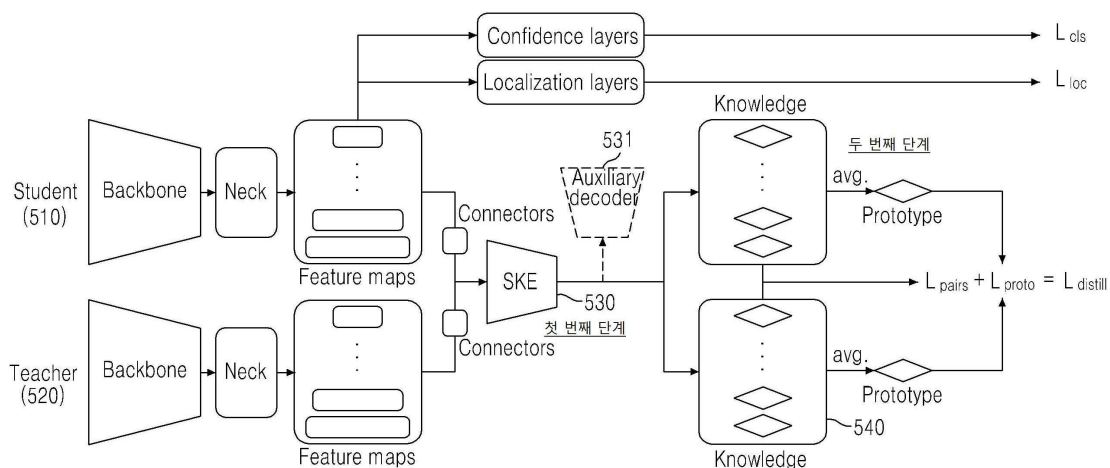
(54) 발명의 명칭 **사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법 및 장치**

(57) 요약

사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법 및 장치가 제시된다. 본 발명의 일 실시예에 따른 컴퓨터 장치에 의해 수행되는 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법은, 공유 지식 증류기를 이용하여 복수개의 특징 지도로부터 지식을 증류하는 단계를 포함하고, 상기 공유 지식 증류기는, 복수개의 상기 특징 지도에 공유될 수 있다.

대표도

500



(52) CPC특허분류
G06N 5/02 (2023.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1415172337
과제번호	20006483
부처명	산업통상자원부
과제관리(전문)기관명	한국산업기술평가관리원
연구사업명	산업용지능융합부품기술개발사업
연구과제명	4K30p급 Deep Learning 기반 Edge Computing IP 카메라용 시스템반도체 개발 (3차년도)
기 여 율	1/1
과제수행기관명	(주)아이닉스
연구기간	2021.01.01 ~ 2021.12.31

명세서

청구범위

청구항 1

컴퓨터 장치에 의해 수행되는 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법에 있어서,
공유 지식 증류기를 이용하여 복수개의 특징 지도로부터 지식을 증류하는 단계
를 포함하고,
상기 공유 지식 증류기는,
복수개의 상기 특징 지도에 공유되는 것
을 특징으로 하는, 균형 잡힌 지식 증류 방법.

청구항 2

제1항에 있어서,
선생 네트워크와 학생 네트워크의 프로토타입을 추가적으로 전달하여, 상기 특징 지도의 쌍에 지식의 전달 정도
마다 서로 다른 패널티를 부여하는 단계
를 더 포함하고,
상기 선생 네트워크와 학생 네트워크의 프로토타입은,
상기 복수개의 특징 지도를 증류하여 얻은 벡터 형태의 지식의 평균 값으로 정의되는 것
을 특징으로 하는, 균형 잡힌 지식 증류 방법.

청구항 3

제1항에 있어서,
상기 복수개의 특징 지도로부터 지식을 증류하는 단계는,
서로 다른 해상도를 갖는 상기 복수개의 특징 지도는 해상도에 따라 정해진 위치에 입력되어 상기 공유 지식 증
류기에 공유되며, 추출되는 지식은 벡터 형태를 갖는 것
을 특징으로 하는, 균형 잡힌 지식 증류 방법.

청구항 4

제1항에 있어서,
상기 공유 지식 증류기는,
대칭 구조를 갖는 임시적인 보조 디코더를 사용하여 오토 인코더(auto-encoder)와 같이 재구성 오류를 최소화하
며 학습되는 것
을 특징으로 하는, 균형 잡힌 지식 증류 방법.

청구항 5

사물 검출기 경량화를 위한 균형 잡힌 지식 증류 장치에 있어서,
복수개의 특징 지도로부터 지식을 증류하는 공유 지식 증류기
를 포함하고,
상기 공유 지식 증류기는,

복수개의 상기 특징 지도에 공유되는 것
을 특징으로 하는, 균형 잡힌 지식 증류 장치.

청구항 6

제5항에 있어서,

선생 네트워크와 학생 네트워크의 프로토타입을 추가적으로 전달하여, 상기 특징 지도의 쌍에 지식의 전달 정도마다 서로 다른 패널티를 부여하는 프로토타입 전달부

를 더 포함하고,

상기 선생 네트워크와 학생 네트워크의 프로토타입은,

상기 복수개의 특징 지도를 증류하여 얻은 벡터 형태의 지식의 평균 값으로 정의되는 것

을 특징으로 하는, 균형 잡힌 지식 증류 장치.

청구항 7

제5항에 있어서,

상기 공유 지식 증류기는,

서로 다른 해상도를 갖는 상기 복수개의 특징 지도는 해상도에 따라 정해진 위치에 입력되어 상기 공유 지식 증류기에 공유되며, 추출되는 지식은 벡터 형태를 갖는 것

을 특징으로 하는, 균형 잡힌 지식 증류 장치.

발명의 설명

기술 분야

[0001] 아래의 본 발명의 실시예들은 사물 검출기를 위한 지식 증류 알고리즘에 관한 것으로, 더욱 상세하게는 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법 및 장치에 관한 것이다.

배경 기술

[0002] 최근 딥러닝 기반의 사물 검출기는 다양한 산업 분야에서 전면모를 발휘하며, 영상에서 필요한 정보들을 인식하는데 유용하게 사용되고 있다. 더 작고 빠른 사물 검출기는 낮은 비용의 디바이스에서도 작동할 수 있게 하는 경량화 기술도 대거 등장했다. 그 중에서도 지식 증류에 관한 연구는 선생 네트워크에서 학생 네트워크로 양질의 지식을 전수하며 학생 네트워크가 더 좋은 성능을 가질 수 있도록 공헌했다.

[0003] 하지만 거의 모든 지식 증류 알고리즘은 분류기로의 적용을 목적으로 제안되어 왔고, 사물 검출기에 적용이 되더라도 네트워크의 구조와 학습 방법 등 다양한 제약 조건이 뒤따라야 했다. 결과적으로 지식 증류 알고리즘이 적용된 분류기는 큰 폭의 성능 향상을 이루었지만, 사물 검출기의 경우 소폭 성능이 향상되거나 조건이 충족되지 않은 경우에는 성능이 되려 감소하는 부작용을 감수해야 했다.

[0004] 기존의 지식 증류 알고리즘은 대부분 분류기에 적용되어 분류 성능을 향상하는데 중점을 두었다. 딥러닝 기반의 분류기의 경우 최종 출력을 바탕으로 입력 영상의 분류를 수행하는 특성이 있다. 이러한 이유로, 학생 네트워크는 선생 네트워크의 출력을 잘 모사하는 것만으로도 향상된 분류 성능을 가질 수 있었다. 여러 개의 특징 지도를 사용하는 알고리즘의 경우에도 서로 짝을 이루는 특징 지도들 사이에서 지식 증류 및 전달되는 과정을 고려할 필요가 없이, 학생 네트워크가 선생 네트워크의 최종 출력을 얼마나 잘 모사하는지에 주력했다.

[0005] 하지만, 실시간성을 목적으로 하는 단층 사물 검출기의 경우 앞서 언급한 분류기와 달리, 딥러닝 네트워크에서 얻어지는 여러 개의 특징 지도를 사용하여 물체를 검출한다. 따라서, 선생 네트워크의 최종 출력을 잘 모사하기에만 특화된 기존 알고리즘은 단층 사물 검출기에서 효과적으로 작동하지 않는다. 기존 알고리즘이 사용되더라도 매우 국한된 상황에서만 적용이 가능했다. 이는 기존 알고리즘이 여러 개의 특징 지도에 지식을 고르게 증류하지 못하고, 전달하는 능력이 부재하기 때문이다.

[0006] 지식 증류 및 전달 과정에서 발생하는 불균형을 효과적으로 해소하지 못하면, 분류기의 경우와 같이 학생 네트

워크가 선생 네트워크의 특정 특징 지도만 집중적으로 학습하게 되는 문제가 발생한다. 이러한 현상은 지식 증류 알고리즘의 적용 이후에도 학생 네트워크를 갖는 검출기의 성능이 미비하게 향상되거나 부정적인 영향으로 저하되는 결과를 초래한다.

선행기술문헌

비특허문헌

- [0007] (비특허문헌 0001) Zagoruyko, Sergey, and Nikos Komodakis. "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer." arXiv preprint arXiv:1612.03928 (2016).
- (비특허문헌 0002) Tung, Frederick, and Greg Mori. "Similarity-preserving knowledge distillation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.
- (비특허문헌 0003) Chen, Defang, et al. "Cross-layer distillation with semantic calibration." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 8. 2021.

발명의 내용

해결하려는 과제

- [0008] 본 발명의 실시예들은 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법 및 장치에 관하여 기술하며, 보다 구체적으로 사물 검출기를 위한 지식 증류 알고리즘 기술을 제공한다.
- [0009] 본 발명의 실시예들은 사물이 검출되는 네트워크에 여러 특징 지도를 대상으로 지식을 고르게 증류할 뿐만 아니라 고르게 전달함으로써, 선생 네트워크로부터 학생 네트워크가 양질의 지식을 전달받을 수 있는 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법 및 장치를 제공하는데 있다.
- [0010] 또한, 본 발명의 실시예들은 추가적인 제약 조건이 없이 다양한 구조를 갖는 사물 검출기 네트워크에 손쉽게 적용이 가능한 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법 및 장치를 제공하는데 있다.

과제의 해결 수단

- [0011] 본 발명의 일 실시예에 따른 컴퓨터 장치에 의해 수행되는 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법은, 공유 지식 증류기를 이용하여 복수개의 특징 지도로부터 지식을 증류하는 단계를 포함하고, 상기 공유 지식 증류기는, 복수개의 상기 특징 지도에 공유될 수 있다.
- [0012] 선생 네트워크와 학생 네트워크의 프로토타입을 추가적으로 전달하여, 상기 특징 지도의 쌍에 지식의 전달 정도마다 서로 다른 패널티를 부여하는 단계를 더 포함하고, 상기 선생 네트워크와 학생 네트워크의 프로토타입은, 상기 복수개의 특징 지도를 증류하여 얻은 벡터 형태의 지식의 평균 값으로 정의될 수 있다.
- [0013] 상기 복수개의 특징 지도로부터 지식을 증류하는 단계는, 서로 다른 해상도를 갖는 상기 복수개의 특징 지도는 해상도에 따라 정해진 위치에 입력되어 상기 공유 지식 증류기에 공유되며, 추출되는 지식은 벡터 형태를 가질 수 있다.
- [0014] 상기 공유 지식 증류기는, 대칭 구조를 갖는 임시적인 보조 디코더를 사용하여 오토 인코더(auto-encoder)와 같이 재구성 오류를 최소화하며 학습될 수 있다.
- [0015] 본 발명의 다른 실시예에 따른 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 장치는, 복수개의 특징 지도로부터 지식을 증류하는 공유 지식 증류기를 포함하고, 상기 공유 지식 증류기는, 복수개의 상기 특징 지도에 공유될 수 있다.
- [0016] 선생 네트워크와 학생 네트워크의 프로토타입을 추가적으로 전달하여, 상기 특징 지도의 쌍에 지식의 전달 정도마다 서로 다른 패널티를 부여하는 프로토타입 전달부를 더 포함하고, 상기 선생 네트워크와 학생 네트워크의 프로토타입은, 상기 복수개의 특징 지도를 증류하여 얻은 벡터 형태의 지식의 평균 값으로 정의될 수 있다.
- [0017] 상기 공유 지식 증류기는, 서로 다른 해상도를 갖는 상기 복수개의 특징 지도는 해상도에 따라 정해진 위치에

입력되어 상기 공유 지식 증류기에 공유되며, 추출되는 지식은 벡터 형태를 가질 수 있다.

발명의 효과

[0018] 본 발명의 실시예들에 따르면 사물이 검출되는 네트워크에 여러 특징 지도를 대상으로 지식을 고르게 증류할 뿐만 아니라 고르게 전달함으로써, 선생 네트워크로부터 학생 네트워크가 양질의 지식을 전달받을 수 있는 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법 및 장치를 제공할 수 있다.

[0019] 또한, 본 발명의 실시예들에 따르면 추가적인 제약 조건이 없이 다양한 구조를 갖는 사물 검출기 네트워크에 손쉽게 적용이 가능한 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법 및 장치를 제공할 수 있다.

도면의 간단한 설명

[0020] 도 1은 기존의 지식 증류 알고리즘을 설명하기 위한 도면이다.

도 2는 본 발명의 일 실시예에 따른 지식 증류 알고리즘을 설명하기 위한 도면이다.

도 3은 본 발명의 일 실시예에 따른 특징 지도들을 위한 공유 지식 증류기를 나타내는 도면이다.

도 4는 본 발명의 일 실시예에 따른 균형 잡힌 지식 전달을 표현한 예시를 나타내는 도면이다.

도 5는 본 발명의 일 실시예에 따른 사물 검출기의 학습 과정을 설명하기 위한 도면이다.

도 6은 본 발명의 일 실시예에 따른 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법을 나타내는 흐름도이다.

도 7은 본 발명의 일 실시예에 따른 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 장치를 나타내는 블록도이다.

도 8은 본 발명의 일 실시예에 따른 경계 박스 시각화 결과를 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0021] 이하, 첨부된 도면을 참조하여 본 발명의 실시예들을 설명한다. 그러나, 기술되는 실시예들은 여러 가지 다른 형태로 변형될 수 있으며, 본 발명의 범위가 이하 설명되는 실시예들에 의하여 한정되는 것은 아니다. 또한, 여러 실시예들은 당해 기술분야에서 평균적인 지식을 가진 자에게 본 발명을 더욱 완전하게 설명하기 위해서 제공되는 것이다. 도면에서 요소들의 형상 및 크기 등은 보다 명확한 설명을 위해 과장될 수 있다.

[0023] 기존의 DNN(Deep Neural Network) 기반 지식 증류 알고리즘은 분류기만을 위해 단일 특징 지도를 사용하거나, 혹은 여러 개의 특징 지도를 사용하여 선생 네트워크로부터 학생 네트워크에 지식을 전달하는 연구가 주를 이루었다. 하지만, 분류기만을 위한 지식 증류 알고리즘을 사물 검출기에 적용할 경우, 분류기와 사물 검출기의 구조적 차이에 의해 여러 가지 문제점이 나타난다.

[0024] 도 1은 기존의 지식 증류 알고리즘을 설명하기 위한 도면이다.

[0025] 도 1을 참조하면, 기존의 지식 증류 알고리즘은 여러 개의 특징 지도(11)를 사용하여 각각 증류 네트워크(12)를 거친 후 선생 네트워크로부터 학생 네트워크에 지식(13)을 전달할 수 있다. 이 경우, 지식(13)이 여러 개의 특징 지도(11)로부터 고르게 추출되지 않거나, 여러 개의 특징 지도들(11) 사이의 지식 전달 과정에서 불균형이 쉽게 발생할 수 있다. 이러한 문제점은 지식(13)을 전달 받을 학생 네트워크가 정량적, 정성적으로 우수한 성능을 발휘하는 데 방해 요소가 된다.

[0026] 따라서 본 발명의 실시예들은 사물 검출기에서 지식 증류 알고리즘을 사용할 때 발생하는 불균형을 효과적으로 해소하기 위한 두 가지 기법을 제안한다. 첫 번째는 증류 불균형을 해결하기 위해 여러 개의 특징 지도에서 지식을 고르게 추출하는 기법이고, 두 번째는 전달 불균형을 해결하기 위해 이 추출된 지식을 균형 있게 전달하는 기법이다. 결과적으로, 본 발명의 실시예들은 균형 있는 지식 전달을 기반으로 타 알고리즘과 대비하여 우수한 정성적, 정량적 성능을 보인다.

[0028] 도 2는 본 발명의 일 실시예에 따른 지식 증류 알고리즘을 설명하기 위한 도면이다.

- [0029] 도 2를 참조하면, 본 발명의 일 실시예에 따른 지식 증류 알고리즘은 두 가지 부분으로 구성될 수 있다.
- [0030] 첫 번째는, 여러 개의 특징 지도(210)로부터 지식(230)을 고르게 증류하기 위한 증류 네트워크(220)이다. 이 증류 네트워크(220)는 타 알고리즘과 달리, 모든 특징 지도에 공유된다는 특징을 가진다. 증류 불균형을 유발하는 독립적인 증류기와 달리, 일 실시예에 따른 증류기는 모든 특징 지도의 지식이 같은 특징 공간에 표현되는 특징이 있다.
- [0031] 두 번째는, 지식을 고르게 전달하기 위한 프로토타입 전달 함수이다. 일 실시예에 따른 지식 증류 알고리즘에서는 학생 네트워크의 프로토타입이 선생 네트워크의 프로토타입과 같아지도록 유도하여, 전달 불균형을 유발하는 특징 지도(210) 쌍에 추가적인 패널티를 부여한다.
- [0032] 이 두 가지 구성을 갖는 일 실시예에 따른 지식 증류 알고리즘은 지식 증류 과정에서 발생하는 불균형을 효과적으로 억제하여 사물 검출기에서도 효과적으로 선생 네트워크의 지식(230)을 학생 네트워크로 전달할 수 있다.
- [0033] 한편, 아래에서 설명되는 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법은 지식 증류 알고리즘에 포함될 수 있다.
- [0035] 도 3은 본 발명의 일 실시예에 따른 특징 지도들을 위한 공유 지식 증류기를 나타내는 도면이다.
- [0036] 도 3을 참조하면, 본 발명의 일 실시예에 따른 지식 추출 과정을 설명하기 위한 것으로, 특징 지도들을 위한 공유 지식 증류기(Shared Knowledge Encoder)(310)를 나타낸다.
- [0037] 일 실시예에 따른 공유 지식 증류기(310)는 모든 특징 지도에 공유되는 특징을 갖는다. 이 때, 서로 다른 해상도를 갖는 특징 지도들에 공유되기 위해서 각 특징 지도는 해상도에 따라 정해진 위치에 입력된다. 예를 들어, 가장 큰 해상도를 갖는 특징 지도는 가장 많은 레이어를 거치게 되고, 가장 작은 해상도를 갖는 특징 지도는 가장 적은 레이어를 거치게 된다. 이로써, 추출되는 지식(330)은 모두 같은 벡터 형태를 가지며 같은 특징 공간에 표현될 수 있게 된다.
- [0038] 이 공유 지식 증류기(310)는 대칭 구조를 갖는 임의 네트워크를 사용해 오토 인코더(auto-encoder)와 같이 재구성 오류(reconstruction error) L_{recon} 를 최적화하며 학습되고, 다음 식과 같이 표현될 수 있다. 즉, 공유 지식 증류기(Shared Knowledge Encoder, SKE)(310)는 대칭 구조를 갖는 보조 디코더(auxiliary decoder)(320)를 사용하여 재구성 오류를 최적화하며 학습된다.
- [0039] [수학식 1]
- [0040]
$$z_i = SKE_{i \rightarrow N}(F_i)$$
- [0041] [수학식 2]
- [0042]
$$\mathcal{L}_{recon} = \sum_{i=1}^N \|F_i - Dec_{N \rightarrow i}(z_i)\|_2^2$$
- [0043] 이로써, 벡터 형태로 표현되는 각 지식은 입력 특징 지도의 암시적 정보(implicit information)를 갖는다.
- [0045] 도 4는 본 발명의 일 실시예에 따른 균형 잡힌 지식 전달을 표현한 예시를 나타내는 도면이다.
- [0046] 도 4를 참조하면, 지식 전달의 균형 조절(Averaged Prototype Transfer)을 나타낸다. 본 발명의 일 실시예에서는 여러 개의 특징 지도를 사용해 공유 지식 증류기(410)를 통해 지식을 증류하고 이를 전달할 때 발생하는 불균형을 억제하기 위한 프로토타입 전달 함수 L_{proto} 를 사용할 수 있으며, 다음 식과 같이 표현할 수 있다.

[0047] [수학식 3]

$$p = \frac{1}{N} \sum_{i=1}^N z_i$$

[0048]

[0049] [수학식 4]

$$\mathcal{L}_{proto} = \|p^t - p^s\|_2^2$$

[0050]

[0051] 선생 네트워크(420)와 학생 네트워크(430)의 프로토타입은 모든 특징 지도를 증류하여 얻은 벡터 형태의 지식의 평균 값으로 정의된다. 미리 정의된 특징 지도 쌍 간의 지식 전달 과정에서 불균형이 발생할 경우 특정 쌍에서만 지식이 충분히 전달되고 나머지 쌍은 그러지 못하게 된다.

[0052] 이 때, 지식 벡터의 평균으로 정의되는 각 네트워크(420, 430)의 프로토타입을 추가적으로 전달하면, 지식의 전달 정도마다 서로 다른 패널티가 각 쌍들에 부여될 수 있다. 이로써 지식 전달 불균형으로 인해 소외된 특징 지도 쌍의 학습에 집중하도록 유도할 수 있다.

[0054] 도 5는 본 발명의 일 실시예에 따른 사물 검출기의 학습 과정을 설명하기 위한 도면이다.

[0055] 도 5를 참조하면, 본 발명의 적용을 위한 사물 검출기 학습 과정을 나타낸다. 본 발명의 일 실시예에서 제안한 지식 증류 알고리즘을 적용하기 위해서 학생 네트워크가 사용되는 사물 검출기(500)를 두 단계로 나누어 학습할 수 있다.

[0056] 먼저, 첫 번째 단계에서 공유 지식 증류기(530)로부터 양질의 지식(540)을 얻어내기 위해 좋은 성능을 갖는 선생 네트워크(520)를 사용하여 앞서 설명한 학습 방법을 통해 공유 지식 증류기(530)의 파라미터를 최적화할 수 있다. 이 때, 보조 디코더(531)가 사용될 수 있다.

[0057] 다음으로, 두 번째 단계에서 공유 지식 증류기(530)의 학습이 완료되면, 선생 네트워크(520)로부터 얻어낸 양질의 지식(540)을 추가로 사용하여 학생 네트워크(510)를 사용하는 사물 검출기(500)의 파라미터를 최적화할 수 있다. 이 때, 사물 검출기(500)의 학습을 위해 사용하는 손실 함수는 일반적인 사물 검출기를 따르며, 추가적인 하이퍼파라미터 α 와 β 를 사용하여 본 발명에서 제안된 두 가지 손실 함수를 더할 수 있으며, 다음 식과 같이 나타낼 수 있다.

[0058] [수학식 5]

$$\mathcal{L}_{pairs} = \sum_{i=1}^N \|z_i^t - z_i^s\|_2^2$$

[0059]

[0060] [수학식 6]

$$\mathcal{L}_{total} = \mathcal{L}_{det} + \alpha \mathcal{L}_{pairs} + \beta \mathcal{L}_{proto}$$

[0061]

[0063] 도 6은 본 발명의 일 실시예에 따른 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법을 나타내는 흐름도이다.

[0064] 도 6을 참조하면, 본 발명의 일 실시예에 따른 컴퓨터 장치에 의해 수행되는 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법은, 공유 지식 증류기를 이용하여 복수개의 특징 지도로부터 지식을 증류하는 단계(S110)를 포함할 수 있다.

[0065] 또한, 선생 네트워크와 학생 네트워크의 프로토타입을 추가적으로 전달하여, 특징 지도의 쌍에 지식의 전달 정

도마다 서로 다른 패널티를 부여하는 단계(S120)를 더 포함할 수 있다.

- [0066] 아래에서 본 발명의 일 실시예에 따른 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법을 보다 구체적으로 설명한다.
- [0068] 본 발명의 일 실시예에 따른 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 방법은 본 발명의 일 실시예에 따른 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 장치를 예를 들어 설명할 수 있다.
- [0069] 도 7은 본 발명의 일 실시예에 따른 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 장치를 나타내는 블록도이다.
- [0070] 도 7을 참조하면, 본 발명의 일 실시예에 따른 사물 검출기 경량화를 위한 균형 잡힌 지식 증류 장치(700)는 공유 지식 증류기(710)를 포함하여 이루어질 수 있으며, 실시예에 따라 프로토타입 전달부(720)를 더 포함할 수 있다.
- [0071] 단계(S110)에서, 공유 지식 증류기(710)는 공유 지식 증류기를 이용하여 복수개의 특징 지도로부터 지식을 증류할 수 있다. 여기서, 공유 지식 증류기는 복수개의 특징 지도에 공유될 수 있다.
- [0072] 공유 지식 증류기(710)는 서로 다른 해상도를 갖는 복수개의 특징 지도는 해상도에 따라 정해진 위치에 입력되어 공유 지식 증류기에 공유되며, 추출되는 지식은 벡터 형태를 가질 수 있다. 예를 들어, 가장 큰 해상도를 갖는 특징 지도는 가장 많은 레이어를 거치게 되고, 가장 작은 해상도를 갖는 특징 지도는 가장 적은 레이어를 거치게 된다. 이에 따라 추출되는 지식은 모두 같은 벡터 형태를 가지며 같은 특징 공간에 표현될 수 있다.
- [0073] 이러한 공유 지식 증류기(710)는 대칭 구조를 갖는 임시적인 보조 디코더를 사용하여 오토 인코더(auto-encoder)와 같이 재구성 오류를 최소화하며 학습될 수 있다.
- [0074] 단계(S120)에서, 프로토타입 전달부(720)는 선생 네트워크와 학생 네트워크의 프로토타입을 추가적으로 전달하여, 특징 지도의 쌍에 지식의 전달 정도마다 서로 다른 패널티를 부여함으로써, 공유 지식 증류기(710)를 통해 지식을 증류하고 전달할 때 발생하는 불균형을 억제할 수 있다. 여기서, 선생 네트워크와 학생 네트워크의 프로토타입은 복수개의 특징 지도를 증류하여 얻은 벡터 형태의 지식의 평균 값으로 정의될 수 있다.
- [0075] 프로토타입 전달부(720)는 지식 벡터의 평균으로 정의되는 선생 네트워크와 학생 네트워크의 프로토타입의 프로토타입을 추가적으로 전달함으로써, 지식의 전달 정도마다 서로 다른 패널티가 각 쌍들에 부여하여 지식 전달 불균형을 억제할 수 있다.
- [0076] 이와 같이 본 발명의 실시예에서는 사물 검출기를 위한 지식 증류 알고리즘을 제안한다. 본 발명의 실시예들에 따르면 사물이 검출되는 네트워크에 여러 특징 지도를 대상으로 지식을 고르게 증류할 뿐만 아니라, 고르게 전달함으로써 기존 알고리즘들과 차별점을 갖는다. 또한, 실시예들에 따르면 추가적인 제약 조건이 없이도 다양한 구조를 갖는 사물 검출기 네트워크에 적용이 손쉽게 가능하다. 이에 따라 선생 네트워크로부터 양질의 지식을 전달받은 학생 네트워크는 정성적, 정량적으로 향상된 결과를 갖는다.
- [0078] 본 발명의 실시예들에 따른 정량적 결과를 비교하기 위해 대표적인 사물 검출 데이터셋인 MS COCO에 대해 성능을 평가하였다.
- [0079] 표 1은 MS COCO 2017 test-dev에서의 성능을 나타낸다.

[0080] [표 1]

T	S	Neck	Distill	AP@0.5	AP	APs	APm	API	AR	ARs	ARm	ARI
R-101	R-18	F → F	-	39.5	21.9	5.9	23.5	35.7	34.3	10.9	36.5	53.7
			AT	40.9	23.0	6.1	24.7	37.8	35.2	11.1	37.5	55.4
			SP	40.5	22.7	6.1	24.5	37.1	35.1	11.4	37.5	55.2
			SKD	40.6	22.8	6.2	24.6	37.6	35.0	11.5	37.6	55.3
			본 발명	41.7	23.3	6.5	24.8	38.1	35.5	11.6	37.9	55.5
R-50	M-v2	F → F	-	37.6	20.1	4.6	22.0	34.0	33.4	10.3	36.6	51.7
			AT	38.1	20.4	5.3	22.1	34.3	33.2	10.0	35.2	52.9
			SP	38.0	20.5	5.4	22.1	33.9	33.3	10.1	35.3	52.2
			SKD	38.2	20.5	5.4	22.2	34.0	33.3	10.1	35.5	53.0
			본 발명	38.8	20.6	5.1	22.6	35.1	34.0	11.2	37.1	53.0

[0081]

[0082] 본 발명이 적용된 사물 검출기(ResNet18-FPN, MobileNetv2-FPN)는 대표적인 지식 증류 알고리즘 AT(비특허문헌 1), SP(비특허문헌 2) 및 SKD(비특허문헌 3)에 비해 높은 성능을 달성하였으며, 지식이 증류되기 이전보다 각각 2.2 point, 1.2 point 향상된 성능을 가지는 것을 확인할 수 있다.

[0083] 도 8은 본 발명의 일 실시예에 따른 경계 박스 시각화 결과를 나타내는 도면이다.

[0084] 도 8을 참조하면, 본 발명의 실시예들에 따른 정성적인 결과 비교를 위해 작은 물체, 겹쳐 있는 물체, 얇은 물체 등 사물 검출이 어려운 대표적인 경우를 예로 검출 시각화를 수행했다.

[0085] 그 결과, 짝수 행에 표시된 본 발명(Ours)이 적용된 사물 검출기는 홀수 행의 기본 사물 검출기보다 더 많은 양의 사물을 안정적으로 검출해내는 것을 확인할 수 있다.

[0087] 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, 컨트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPA(field programmable 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 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0088] 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.

[0089] 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),

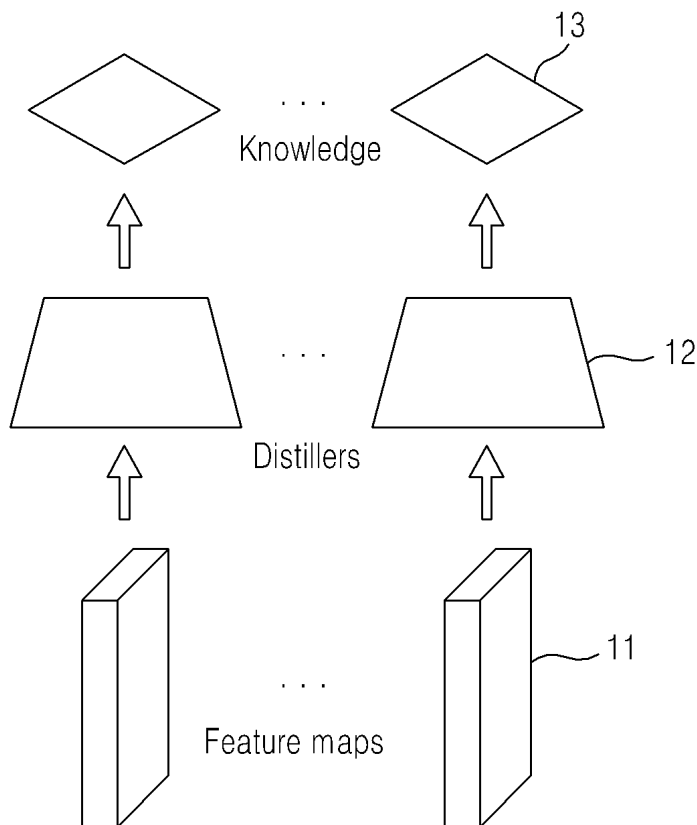
CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0090] 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.

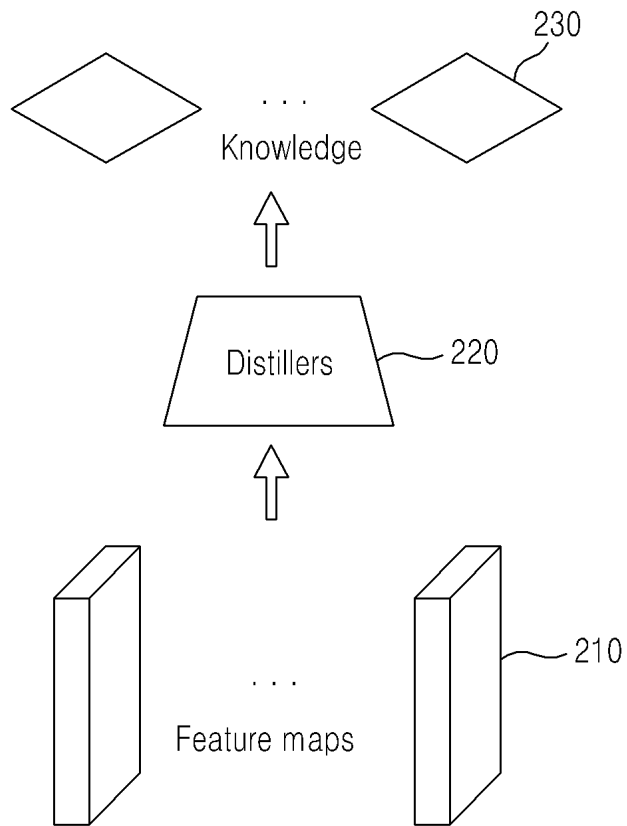
[0091] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

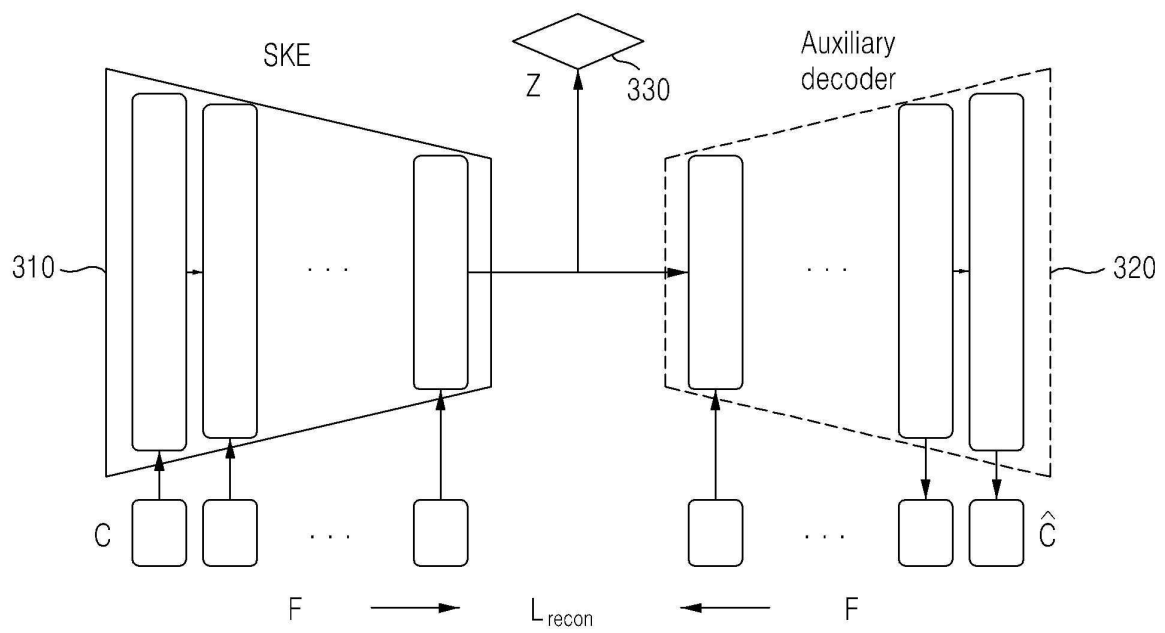
도면1



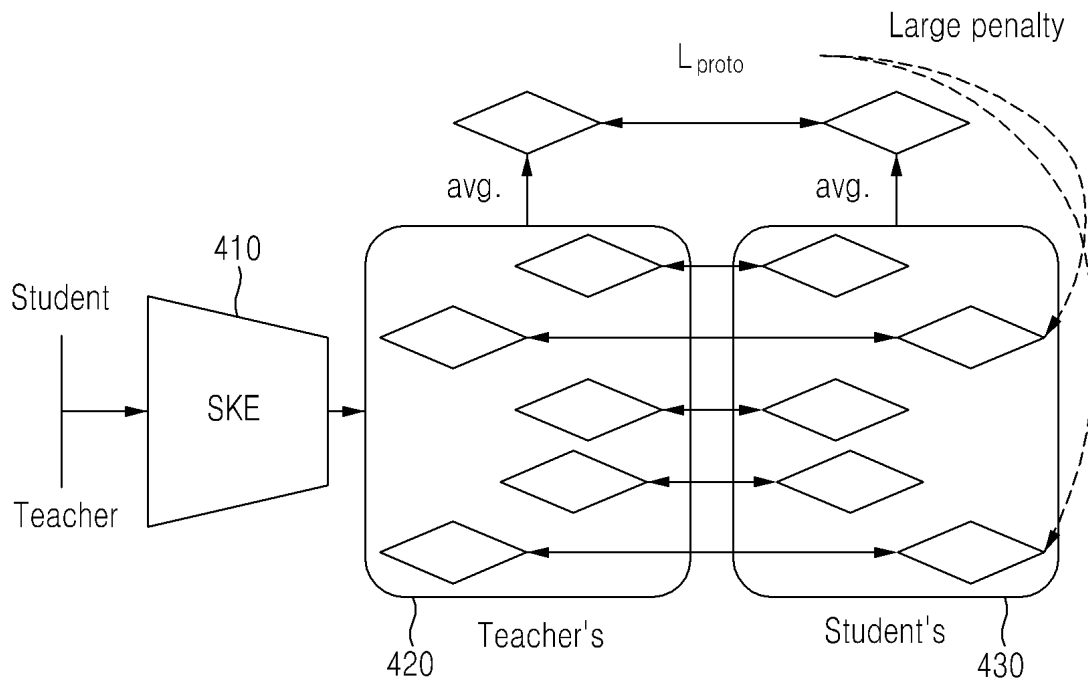
도면2



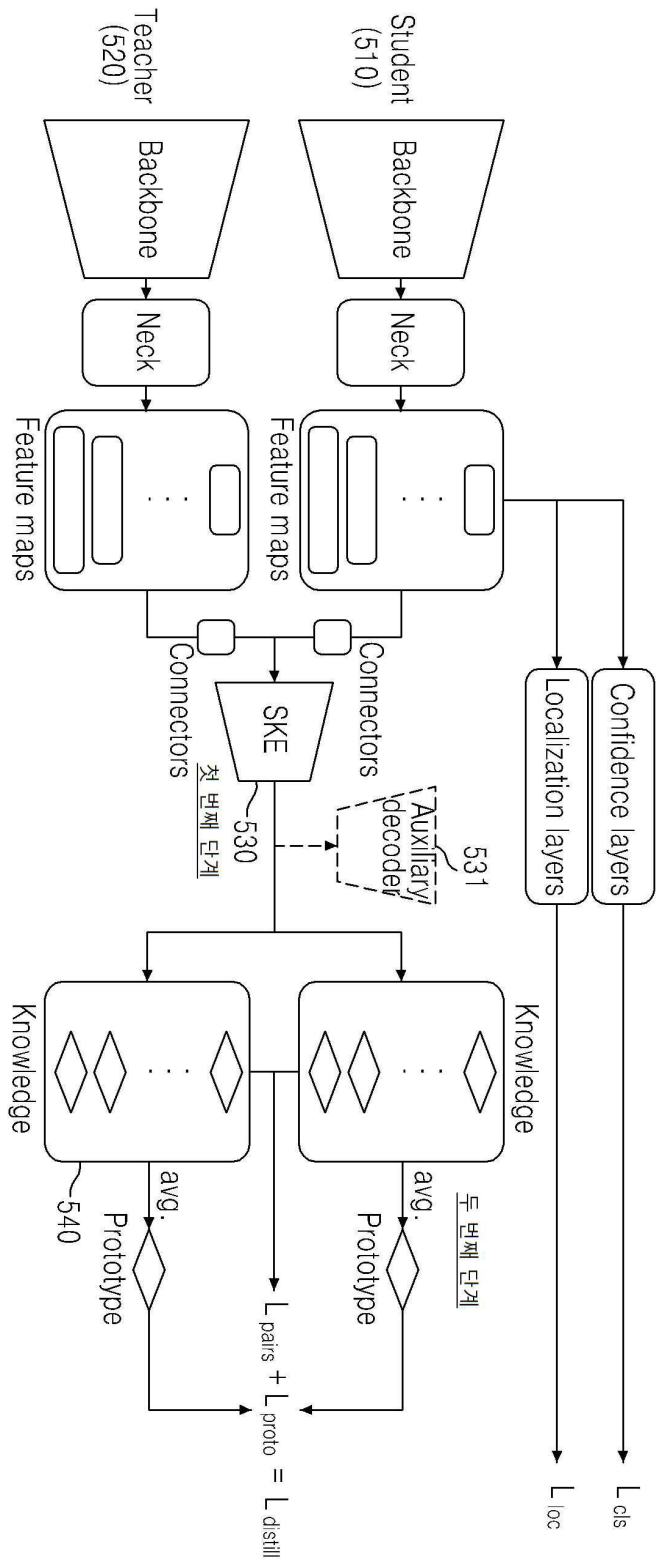
도면3



도면4

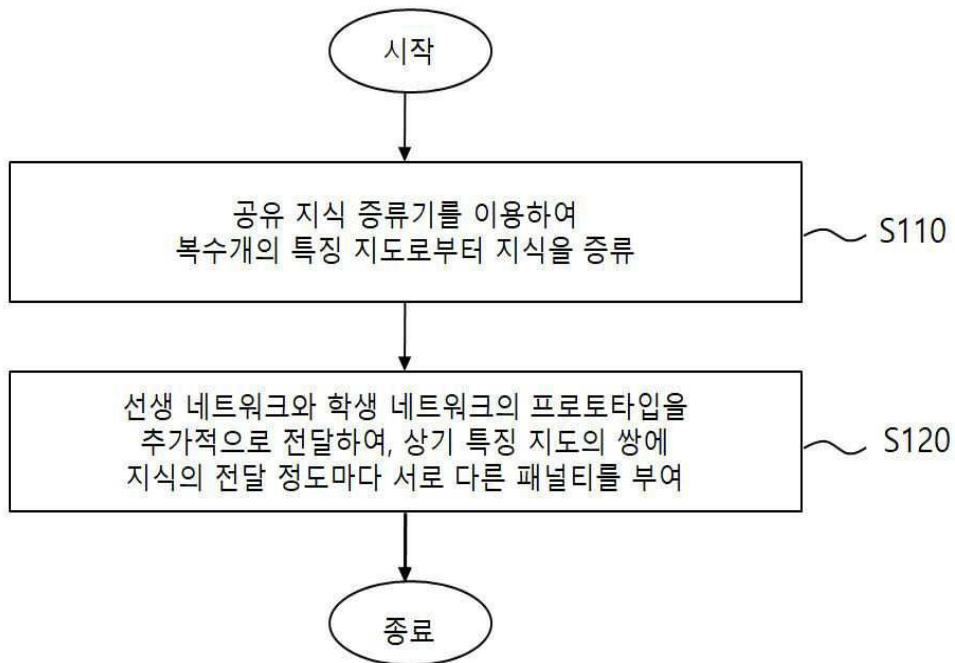


도면5



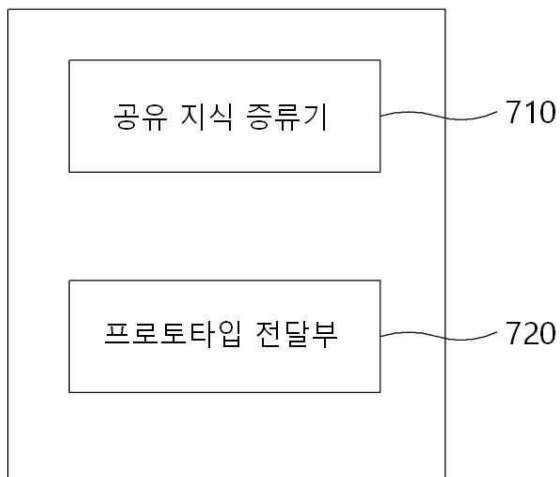
500

도면6



도면7

700



도면8

