



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2025년06월11일  
(11) 등록번호 10-2819296  
(24) 등록일자 2025년06월05일

(51) 국제특허분류(Int. Cl.)  
G06V 10/82 (2022.01) G06N 3/045 (2023.01)  
G06N 3/0464 (2023.01) G06N 3/08 (2023.01)  
G06V 10/44 (2022.01) G06V 20/62 (2022.01)

(52) CPC특허분류  
G06V 10/82 (2022.01)  
G06N 3/045 (2023.01)

(21) 출원번호 10-2024-0133908

(22) 출원일자 2024년10월02일

심사청구일자 2024년10월02일

(56) 선행기술조사문헌

한국 공개특허공보 제10-2022-0073645  
호(2022.06.03.) 1부.\*

한국 공개특허공보 제10-2024-0133384  
호(2024.09.04.) 1부.\*

\*는 심사관에 의하여 인용된 문헌

(73) 특허권자  
수퍼게이트 주식회사

서울특별시 강남구 테헤란로 407, 11층(삼성동,  
이케이타워)

(72) 발명자

문우호

서울특별시 강남구 테헤란로 407, 11층 (삼성동)

김예지

서울특별시 강남구 테헤란로 407, 11층 (삼성동)

김동주

서울특별시 강남구 테헤란로 407, 11층 (삼성동)

(74) 대리인

이장훈, 박정우

전체 청구항 수 : 총 14 항

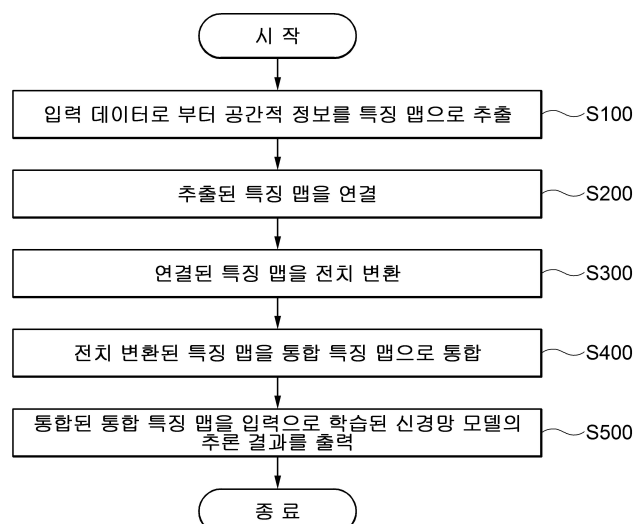
심사관 : 정수진

(54) 발명의 명칭 전치 변환에 기반한 행동 추론 방법 및 이를 수행하는 컴퓨터 장치

(57) 요약

본 발명은 객체의 시공간적 특성을 고려한 행동 추론 방법에 관한 것이다. 본 발명의 일 실시예에 따른 컴퓨터 장치에서 수행되는 행동 추론 방법으로, 입력 데이터로부터 공간적 정보를 특징 맵으로 추출하고, 추출된 특징 맵을 연결하며, 연결된 특징 맵을 전치 변환하고, 전치 변환된 특징 맵을 입력으로 하여 학습된 신경망 모델의 추론 결과를 출력할 수 있다. 본 발명은 추론을 위한 네트워크에서 CNN을 사용하면 공간이 시간별로 가지는 특징점을, 트랜스포머를 사용하면 공간 간 시간에 따른 관계성을 학습할 수 있다. 이러한 변환 과정을 통해 시간적 정보에 초점을 맞추어 공간적 정보를 유지하면서도, 높은 정확도로 행동 인식을 수행할 수 있다.

대표도 - 도2



(52) CPC특허분류

**G06N 3/0464** (2023.01)

**G06N 3/08** (2023.01)

**G06V 10/44** (2023.08)

**G06V 20/62** (2023.08)

이 발명을 지원한 국가연구개발사업

|             |                                              |
|-------------|----------------------------------------------|
| 과제고유번호      | 2410005064                                   |
| 과제번호        | 00443723                                     |
| 부처명         | 산업통상자원부                                      |
| 과제관리(전문)기관명 | 한국산업기술기획평가원                                  |
| 연구사업명       | 소재부품기술개발(R&D)                                |
| 연구과제명       | 비전 트랜스포머 지원 하드웨어 가속기를 탑재한 인공지능 엣지 반도체 및 응용 솔 |
| 루선 개발       |                                              |
| 과제수행기관명     | 수퍼게이트 주식회사                                   |
| 연구기간        | 2024.07.01 ~ 2024.12.31                      |

---

## 명세서

### 청구범위

#### 청구항 1

컴퓨터 장치에서 수행되는 행동 추론 방법에 있어서,  
 시퀀스로 입력되는 입력 데이터로부터 공간적 정보를 특징 맵으로 추출하는 단계;  
 상기 추출된 특징 맵을 상기 시퀀스를 고려하여 연결하는 단계;  
 상기 연결된 특징 맵을 전치 변환하는 단계;  
 상기 전치 변환된 특징 맵을 통합 특징 맵으로 통합하는 단계; 및  
 상기 통합된 통합 특징 맵을 입력으로 학습된 신경망 모델의 추론 결과를 출력하는 단계를 포함하고,  
 상기 전치 변환하는 단계는 상기 연결된 특징 맵을 제1 방향 및 상기 제1 방향과 상이한 제2 방향에 따라 전치 변환하고,  
 상기 통합하는 단계는 상기 제1 방향으로 변환된 제1 변환 특징 맵 및 상기 제2 방향으로 변환된 제2 변환 특징 맵을 통합하는,  
 행동 추론 방법.

#### 청구항 2

제1 항에 있어서,  
 상기 전치 변환하는 단계는,  
 상기 연결된 특징 맵을 상기 특징 맵의 높이 방향을 기준으로 전치 변환하는 행동 추론 방법.

#### 청구항 3

제2 항에 있어서,  
 상기 전치 변환하는 단계는,  
 상기 연결된 특징 맵을 상기 특징 맵의 넓이 방향을 기준으로 전치 변환하는 행동 추론 방법.

#### 청구항 4

제3 항에 있어서,  
 상기 출력하는 단계는,  
 상기 높이 방향을 기준으로 전치 변환된 제1 변환 특징 맵 및 상기 넓이 방향을 기준으로 전치 변환된 제2 변환 특징 맵의 출력을 통합한 통합 출력을 추론 결과를 출력하는 행동 추론 방법.

#### 청구항 5

제3 항에 있어서,  
 상기 전치 변환의 기준은 추론하고자 하는 객체의 행동에 따라 결정되는 행동 추론 방법.

#### 청구항 6

제1 항에 있어서,  
 상기 특징 맵으로 추출하는 단계는,  
 상기 입력 데이터를 제1 모델로 축약한 특징 맵을 추출하는 행동 추론 방법.

## 청구항 7

제1 항에 있어서,

상기 제1 변환 특징 맵 및 상기 제2 변환 특징 맵을 합산하여 중간 맵을 생성하는 단계를 포함하고,

상기 통합하는 단계는 상기 제1 변환 특징 맵, 상기 제2 변환 특징 맵 및 상기 중간 맵을 연결 연산하여 통합 특징 맵으로 통합하는,

행동 추론 방법.

## 청구항 8

프로세서, 및

상기 프로세서와 통신하는 메모리를 포함하고,

상기 메모리는 상기 프로세서로 하여금 동작들을 수행하게 하는 명령들을 저장하고,

상기 동작들은,

시퀀스로 입력되는 입력 데이터로부터 공간적 정보를 특징 맵으로 추출하는 동작,

상기 추출된 특징 맵을 상기 시퀀스를 고려하여 연결하는 동작,

상기 연결된 특징 맵을 전치 변환하는 동작,

상기 전치 변환된 특징 맵을 통합 특징 맵으로 통합하는 동작, 및

상기 통합된 통합 특징 맵을 입력으로 학습된 신경망 모델의 추론 결과를 출력하는 동작을 포함하고,

상기 전치 변환하는 동작은 상기 연결된 특징 맵을 제1 방향 및 상기 제1 방향과 상이한 제2 방향에 따라 전치 변환하고,

상기 통합하는 동작은 상기 제1 방향으로 변환된 제1 변환 특징 맵 및 상기 제2 방향으로 변환된 제2 변환 특징 맵을 통합하는,

컴퓨터 장치.

## 청구항 9

제8 항에 있어서,

상기 전치 변환하는 동작은,

상기 연결된 특징 맵을 상기 특징 맵의 높이 방향을 기준으로 전치 변환하는 컴퓨터 장치.

## 청구항 10

제9 항에 있어서,

상기 전치 변환하는 동작은,

상기 연결된 특징 맵을 상기 특징 맵의 넓이 방향을 기준으로 전치 변환하는 컴퓨터 장치.

## 청구항 11

제10 항에 있어서,

상기 출력하는 동작은,

상기 높이 방향을 기준으로 전치 변환된 제1 변환 특징 맵 및 상기 넓이 방향을 기준으로 전치 변환된 제2 변환 특징 맵의 출력을 통합한 통합 출력을 추론 결과를 출력하는 컴퓨터 장치.

## 청구항 12

제10 항에 있어서,

상기 전치 변환의 기준은 추론하고자 하는 객체의 행동에 따라 결정되는 컴퓨터 장치.

### 청구항 13

제8 항에 있어서,

상기 특징 맵으로 추출하는 동작은,

상기 입력 데이터를 제1 모델로 축약한 특징 맵을 추출하는 컴퓨터 장치.

### 청구항 14

제8 항에 있어서,

상기 제1 변환 특징 맵 및 상기 제2 변환 특징 맵을 합산하여 중간 맵을 생성하는 동작을 포함하고,

상기 통합하는 동작은 상기 제1 변환 특징 맵, 상기 제2 변환 특징 맵 및 상기 중간 맵을 연결 연산하여 통합 특징 맵으로 통합하는 컴퓨터 장치.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 객체의 시공간적 특성을 고려한 행동 추론 방법에 관한 것이다.

### 배경 기술

[0002] 행동 인식은 컴퓨터 비전 분야에서 중요한 연구 주제 중 하나로, 다양한 응용 분야에서 활용될 수 있다. 행동 인식을 위해서는 시간적 정보와 공간적 정보를 모두 포함하는 것이 필수적이다. 시간적 정보는 연속된 동작의 흐름을 이해하는 데 중요하며, 공간적 정보는 특정 순간의 자세나 위치를 파악하는 데 기여한다.

[0003] 기존의 행동 인식 방법들은 주로 두 가지 접근 방식을 취해왔다. 첫 번째 방식은 클립이나 비디오를 활용하여 시공간적 정보를 동시에 학습하려는 접근이다. 이 방식에서는 3D 컨볼루션을 통해 시공간적 정보를 추출하거나, 이미지를 패치로 처리하여 트랜스포머에 넣는 구조로 행동 인식 시스템을 구성하였다. 이러한 방법은 시공간적 정보를 동시에 처리하여 높은 인식률을 달성할 수 있으나, 연산량이 급격히 증가하게 되어 리소스가 제한된 임베디드 시스템에서는 실시간으로 행동을 인식하기 어려운 문제가 발생하였다.

[0004] 공간적 정보를 우선시하는 경우 이미지에서 행동을 인식하고자 하는 객체의 특징점을 추출하고, 이후 시간적 정보를 포함하기 위해 이러한 특징 맵들을 쌓은 뒤 추가적인 처리를 통해 시간적 정보를 더하는 방식을 사용하였다. 예를 들어, GNN(Graph Neural Network)을 통과시키거나 트랜스포머 또는 CNN을 다시 적용하는 방식이 있다.

[0005] 그러나 이러한 방법은 내재적으로 공간적 정보의 우위를 두고 설계되었기 때문에, 시간적 정보의 중요성이 간과되는 경우가 많았다. 예를 들어 앉는 동작과 일어나는 동작처럼 동일한 동작이라도 시간적 정보가 더 중요한 상황에서는 공간적 정보에 치중한 설계가 정확한 행동 인식을 저해하는 문제가 발생할 수 있다.

## 발명의 내용

### 해결하려는 과제

[0006] 본 발명의 목적은, 적은 리소스로도 네트워크의 목적에 부합하는 입력 데이터를 효율적으로 제공함으로써, 연산 부담을 최소화하는 것이다.

[0007] 또한, 본 발명은 시간적 정보를 우선적으로 고려하는 설계를 통해 공간적 정보와의 조화를 이루어, 행동 인식에서 발생할 수 있는 오류를 최소화하고자 한다.

[0008] 이를 통해 본 발명은 실시간 행동 인식 시스템에서 정확도와 효율성을 동시에 향상시키는 것을 목적으로 한다. 나아가, 이 접근법은 임베디드 시스템이나 리소스가 제한된 환경에서도 신속하고 정확한 행동 인식을 가능하게 할 수 있다.

### 과제의 해결 수단

- [0009]     기술한 목적을 달성하기 위한, 본 발명의 일 실시예에 따른 컴퓨터 장치에서 수행되는 행동 추론 방법으로, 입력 데이터로부터 공간적 정보를 특징 맵으로 추출하고, 추출된 특징 맵을 연결하며, 연결된 특징 맵을 전치 변환하고, 전치 변환된 특징 맵을 입력으로 하여 학습된 신경망 모델의 추론 결과를 출력할 수 있다.
- [0010]     연결된 특징 맵을 높이 방향을 기준으로 전치 변환할 수 있다.
- [0011]     연결된 특징 맵을 넓이 방향을 기준으로 전치 변환할 수 있다.
- [0012]     높이 방향으로 전치 변환된 제1 변환 특징 맵과 넓이 방향으로 전치 변환된 제2 변환 특징 맵의 출력을 통합하여 추론 결과를 출력할 수 있다.
- [0013]     전치 변환의 기준은 추론하고자 하는 객체의 행동에 따라 결정될 수 있다.
- [0014]     입력 데이터를 제1 모델로 축약하여 특징 맵을 추출할 수 있다.
- [0015]     프로세서와 통신하는 메모리를 포함한 컴퓨터 장치로, 입력 데이터로부터 공간적 정보를 특징 맵으로 추출하고, 추출된 특징 맵을 연결하며, 연결된 특징 맵을 전치 변환하고, 전치 변환된 특징 맵을 입력으로 하여 학습된 신경망 모델의 추론 결과를 출력할 수 있다.

### 발명의 효과

- [0016]     본 발명은 공간적 정보를 추출하여 실시간 시스템에서 입력된 프레임의 특징 맵을 생성하고, 이를 기반으로 시간적 정보를 반영하여 보다 고차원적인 행동 인식 결과를 제공할 수 있다.
- [0017]     또한, 본 발명은 시간적 정보에 초점을 맞춘 설계를 통해, 특징 맵을 변형하여 각 공간에 해당하는 시간의 흐름을 고려한 행동의 인식이 가능하다.
- [0018]     본 발명은 추론을 위한 네트워크에서 CNN을 사용하면 공간이 시간별로 가지는 특징점을, 트랜스포머를 사용하면 공간 간 시간에 따른 관계성을 학습할 수 있다. 이러한 변환 과정을 통해 시간적 정보에 초점을 맞추어 공간적 정보를 유지하면서도, 높은 정확도로 행동 인식을 수행할 수 있다.
- [0019]     본 발명은 시간 및 공간적 정보를 균형있게 고려한 구조를 통해 행동 인식 시스템의 정확도를 높일 수 있으며, 실시간 시스템에서의 행동 인식 오차를 줄일 수 있다.

### 도면의 간단한 설명

- [0020]     도 1은 본 발명의 일 실시예에 따른 시스템의 구성을 나타내는 블록도이다.
- 도 2는 본 발명의 일 실시예에 따른 추론 과정을 나타내는 흐름도이다.
- 도 3은 본 발명의 일 실시예에 따른 연결 과정을 예시하는 도이다.
- 도 4는 본 발명의 일 실시예에 따른 전치 변환 과정을 예시하는 도이다.
- 도 5는 본 발명의 일 실시예에 따른 특징 맵의 통합을 나타내는 도이다.
- 도 6은 본 발명의 일 실시예에 따른 출력의 통합을 나타내는 도이다.
- 도 7은 본 발명의 일 실시예에 따른 서비스 서버의 컴퓨팅 장치로의 구현을 나타내는 예시도이다.

### 발명을 실시하기 위한 구체적인 내용

- [0021]     이하의 내용은 단지 발명의 원리를 예시한다. 그러므로 당업자는 비록 본 명세서에 명확히 설명되거나 도시되지 않았지만 발명의 원리를 구현하고 발명의 개념과 범위에 포함된 다양한 장치를 발명할 수 있는 것이다. 또한, 본 명세서에 열거된 모든 조건부 용어 및 실시 예들은 원칙적으로, 발명의 개념이 이해되도록 하기 위한 목적으로만 명백히 의도되고, 이와같이 특별히 열거된 실시 예들 및 상태들에 제한적이지 않는 것으로 이해되어야 한다.
- [0022]     상술한 목적, 특징 및 장점은 첨부된 도면과 관련한 다음의 상세한 설명을 통하여 보다 분명해질 것이며, 그에 따라 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 발명의 기술적 사상을 용이하게 실시할 수 있을 것이다.
- [0023]     또한, 발명을 설명함에 있어서 발명과 관련된 공지 기술에 대한 구체적인 설명이 발명의 요지를 불필요하게 흐

릴 수 있다고 판단되는 경우에 그 상세한 설명을 생략하기로 한다. 이하에는 첨부한 도면을 참조하여 본 발명의 바람직한 실시 예에 대해 상세하게 설명한다.

- [0024] 도 1은 본 발명의 일 실시예에 따른 서비스 시스템을 나타내는 예시도이다.
- [0025] 도 1을 참조하면 본 발명의 시스템은 사용자 단말(1000)과 서버(300)로 구성된 시스템으로, 행동 인식을 통해 다양한 응용 분야에서 행동 기반 이벤트 탐지를 가능하게 할 수 있다.
- [0026] 사용자 단말(1000)은 카메라 또는 센서와 같은 장치를 통해 실시간으로 입력 데이터(100)를 수집하는 장치이다.
- [0027] 입력 데이터(100)는 주로 동영상 클립 또는 이미지 시퀀스의 형태로 제공되며, 사용자 단말(1000)은 이를 서버(300)로 전송하기 위해 필요한 전처리 작업을 수행할 수 있다.
- [0028] 사용자 단말(1000)은 프레임 단위로 입력 데이터를 캡처하여 서버(300)로 전송할 수 있다. 또한, 임베디드 시스템과 같은 제한된 리소스 환경에서 작동할 수 있도록 설계되어, 실시간으로 데이터를 수집하고 전송할 수 있다. 이 과정에서 리소스 절약을 위해 프레임의 해상도를 조정하거나, 중요한 정보만 추출하여 전송할 수 있다.
- [0029] 사용자 단말(1000)의 주요 목적은 실시간 행동 인식을 위해 필요한 데이터를 효율적으로 수집하고, 이 데이터를 서버(300)로 전송하는 것이다. 이를 통해 서버(300)에서 정확한 행동 인식이 가능하도록 지원할 수 있다.
- [0030] 서버(300)는 신경망 모델(Neural Network Model)을 실행하는 중앙 처리 장치로, 사용자 단말(1000)에서 전송된 입력 데이터를 기반으로 행동 인식 추론을 수행할 수 있다. 서버(300)는 본 발명에 따른 전치 패치 동작 인식(Transpose Patch Action Recognition)을 위한 다양한 신경망 모델(CNN, Transformer 등)을 포함할 수 있다.
- [0031] 서버(300)는 사용자 단말(1000)에서 전송된 입력 데이터를 수신하고, 이를 기반으로 신경망 모델을 통해 행동 인식을 수행할 수 있다. 이 과정에서, 서버(300)는 입력 데이터의 공간적 정보와 시간적 정보를 동시에 처리하여 최종적으로 행동 인식 결과를 생성할 수 있다. 서버(300)는 이 결과를 사용자 단말(1000)로 반환하여, 사용자에게 실시간 피드백을 제공할 수 있다.
- [0032] 서버(300)의 주요 목적은 정확하고 신속한 행동 인식을 통해 사용자 단말(1000)에서 제공된 입력 데이터에 대해 실시간으로 추론 결과를 생성하는 것이다. 서버(300)는 높은 처리 능력을 바탕으로 복잡한 신경망 모델을 실행하며, 시간적 및 공간적 정보를 모두 고려하여 최적의 인식 결과를 도출할 수 있다.
- [0033] 나아가, 본 실시예에서 서버(300)는 네트워크를 통해 다른 컴퓨터나 장치(클라이언트)관계에서 서비스를 제공하는 컴퓨터 시스템 외에 신경망 모델이 동작 가능한 사용자 단말(1000)이나 엣지 디바이스 등의 장치로 대체될 수 있으며, 따라서 본 실시예에 따른 행동 추론 방법을 장치들 간의 통신을 통해 수행하는 것도 가능하다.
- [0034] 즉, 서버(300) 역할을 수행하는 장치가 꼭 중앙 서버(300)일 필요는 없으며, 사용자 단말(1000)이나 엣지 디바이스 등에서도 신경망 모델을 구동하고, 장치 간의 협력을 통해 행동 패턴을 실시간으로 분석하고 인식할 수 있다.
- [0035] 본 발명에 따른 시스템은 다양한 행동을 실시간으로 인식하여, 특정 행동 패턴에 따라 이벤트를 탐지하거나 경고를 제공할 수 있다. 예를 들어, 공공 안전 모니터링 시스템에서 사람이 넘어지는 등의 행동을 실시간으로 탐지하여, 즉각적인 조치를 취할 수 있다. 환자의 움직임을 모니터링하여, 낙상 예방이나 재활 치료 과정에서 발생할 수 있는 이상 행동을 실시간으로 감지하고, 이를 의료진에게 알릴 수 있다.
- [0036] 또는 스마트 홈 시스템에서는 거주자의 행동을 인식하여, 조명이나 온도 조절 등의 자동화된 환경 제어를 실행할 수 있다. 또한, 비정상적인 행동이 감지되면, 보안 경고를 활성화할 수 있다.
- [0037] 이하, 도 2를 참고하여 구체적인 서비스 과정에 대해 설명한다.
- [0038] 도 2는 본 발명의 일 실시예에 따른 추론 과정을 나타내는 흐름도이다.
- [0039] 도 2를 참고하면, 서버(300)는 먼저 공간적 정보를 특징 맵으로 추출한다.
- [0040] 서버(300)는 입력 데이터로부터 행동 인식에 필요한 공간적 정보를 추출하여 특징 맵(feature map)을 생성한다(S100).
- [0041] 실시간 시스템에서 프레임이 입력으로 들어오면, 제1 네트워크는 해당 프레임의 공간적 특징을 분석하여 중요한 정보를 특징 맵 형태로 추출할 수 있다.
- [0042] 이러한 특징 맵은 행동 인식의 기초 데이터로 사용되며, 특히 임베디드 시스템과 같은 리소스가 제한된 환경에

서 효율적인 처리에 핵심적인 역할을 할 수 있다. 이 과정은 모델이 프레임 내에서 주요 객체나 동작을 식별할 수 있도록 도와줄 수 있다.

- [0043] 이어서 서버(300)는 추출된 특징 맵을 연결한다(S200).
- [0044] 서버(300)는 여러 프레임에서 추출된 특징 맵들을 결합하여, 시간적 정보가 반영된 새로운 특징 맵을 생성할 수 있다.
- [0045] 도 3은 본 발명의 일 실시예에 따른 연결 과정을 예시하는 도이다.
- [0046] 도 3을 참조하면 서버(300)는 공간적 정보를 반영한 각 프레임의 특징 맵을 쌓아 올려, 시간의 흐름에 따른 변화를 포함하는 특성을 부여할 수 있다.
- [0047] 이를 통해, 단순히 공간적 정보를 나열하는 것에서 나아가, 각 프레임 간의 연속적인 변화, 즉 시간적 정보를 포함하게 된다. 이 과정은 행동 인식에서 중요한 요소인 시간적 연관성을 반영하기 위해 필수적이다.
- [0048] 서버(300)는 시퀀스로 입력되는 이미지들을 컨볼루션 레이어(310)를 통해 특징 맵(320)로 축약할 수 있다. 축약된 특징 맵(320)들은 시간 또는 프레임 순으로 적층되면서 연결된 특징 맵(320')으로 생성될 수 있다.
- [0049] 축약된 특징 맵(320)들은 각 프레임에서 추출된 정보로, 여러 프레임에서 발생한 특징 맵들이 시간 순으로 배열된다.
- [0050] 축약된 특징 맵(320)들은 데이터의 시간적 흐름을 반영하여 각 프레임의 특징을 시간 축을 따라 적층(stackin g)함으로써, 시간적 변화나 동작 패턴을 포착할 수 있게 만듭니다
- [0051] 적층된 특징 맵들은 시간적 또는 프레임 순으로 연결되어 하나의 통합된 연결된 특징 맵(320') 형태로 생성될 수 있다.
- [0052] 서버(300)는 연결된 특징 맵을 전치(transpose) 변환한다(S300).
- [0053] 서버(300)는 연결된 특징 맵을 전치 변환하여, 각 공간적 위치에서의 시간적 변화를 보다 명확하게 반영할 수 있도록 입력 데이터의 형태를 변경할 수 있다.
- [0054] 전치 변환(transpose)은 특징 맵의 차원을 재배치하여, 신경망 모델이 시간적 정보를 더 효과적으로 학습할 수 있도록 할 수 있다. 예를 들어, CNN 기반의 네트워크에서는 각 공간 위치에서 시간에 따른 특징을 포착하고, 트 랜스포머 기반 네트워크에서는 각 공간 위치 간의 시간적 관계성을 학습할 수 있다.
- [0055] 변환은 높이 방향 또는 너비 방향으로 진행되어 공간적 정보의 보존과 시간적 정보의 강조를 동시에 이룰 수 있 다.
- [0056] 도 4는 본 발명의 일 실시예에 따른 전치 변환 과정을 예시하는 도이다.
- [0057] 도 4를 참조하면 서버(300)는 결정된 전치 변환의 방향(높이 방향 또는 넓이 방향)에 따라 연결된 특징 맵 (320')을 전치 변환하여 전치 변환된 특징 맵(320 T)를 생성한다.
- [0058] 서버(300)는 전치 변환의 방향을 추론하고자 하는 특정 객체의 행동 유형에 따라 동적으로 결정할 수 있다.
- [0059] 서버(300)는 전치 변환의 방향을 고정된 방식으로 수행하는 것이 아니라, 객체의 행동 유형에 따라 동적으로 결 정할 수 있다.
- [0060] 예를 들어, 객체의 행동이 주로 수직적인 움직임을 포함하는 경우(예: 점프나 앉기/일어서기), 높이 방향을 기 준으로 전치 변환을 수행하는 것이 적절할 수 있다. 반대로, 객체의 행동이 수평적인 움직임을 포함하는 경우 (예: 달리기나 좌우로 걷기), 넓이 방향을 기준으로 전치 변환을 수행하는 것이 더 효과적일 수 있다.
- [0061] 즉, 객체의 행동에 따른 전치 변환 예시로 수직적 움직임이 강조되는 행동의 경우, 높이 방향을 기준으로 전치 변환을 수행하는 것이 적합할 수 있다. 이는 시간에 따라 몸의 위치가 수직으로 변하는 특징을 더 잘 포착하게 할 수 있다. 신경망 모델은 이 전치 변환된 특징 맵을 사용하여 수직적 변화를 기반으로 한 행동을 더 정확하게 인식할 수 있다.
- [0062] 사람이 좌우로 걸거나 달리는 행동은 수평적인 변화가 두드러지는 행동의 경우, 넓이 방향을 기준으로 전치 변 환을 수행하는 것이 적합할 수 있다. 이는 신경망이 시간에 따라 좌우로 움직이는 동작을 효과적으로 학습할 수 있도록 한다. 예를 들어, 좌우로 이동하는 동안 신체의 위치와 자세 변화 등을 잘 반영할 수 있다.

- [0063] 또는 복합적 움직임으로 수직적 및 수평적 변화를 모두 포함하는 복합적인 행동이 발생할 수 있다.
- [0064] 이 경우, 높이 방향과 넓이 방향을 모두 고려한 전치 변환이 필요할 수 있다. 먼저 높이 방향을 기준으로 전치 변환을 수행하고, 이어서 넓이 방향을 기준으로 전치 변환을 수행함으로써, 복합적인 행동을 더욱 정교하게 분석할 수 있다.
- [0065] 본 실시예에서 서버(300)는 객체의 특정 행동 유형에 따라 전치 변환의 기준을 동적으로 결정할 수 있는 유연성을 제공할 수 있다. 이를 통해, 객체의 행동 특성에 따라 맞춤형 데이터 변환이 가능해지며, 이는 행동 인식의 정확도를 크게 향상시킬 수 있다.
- [0066] 다음 서버(300)는 전치 변환된 특징 맵을 입력으로 학습된 신경망 모델의 추론 결과를 출력한다.
- [0067] 서버(300)는 전치 변환된 특징 맵을 학습된 신경망 모델에 입력하여, 최종적으로 행동 인식 결과를 도출할 수 있다.
- [0068] 전치 변환된 특징 맵은 학습된 신경망 모델(CNN, 트랜스포머 등)의 입력으로 사용된다. 신경망 모델은 이 입력을 처리하여, 행동 인식의 최종 결과를 생성할 수 있다. 이 단계에서는 모델이 시간적 정보와 공간적 정보를 모두 고려한 상태에서 행동을 정확히 인식할 수 있게 되며, 출력된 결과는 행동 인식 시스템의 최종 판단으로 활용된다.
- [0069] 구체적으로 본 실시예에 따른 서버(300)는 전치 변환된 특징 맵을 통합 특징 맵으로 통합한다(S400).
- [0070] 도 5를 참고하면 서버(300)는 2개의 특징 맵을 통합할 수 있다.
- [0071] 제1 변환 특징 맵(320T')은 예를 들어 높이 방향을 기준으로 전치 변환된 특징 맵일 수 있다. 입력 데이터의 높이 방향에서 중요한 패턴이나 특징을 추출하고 이를 변환하여 생성된 맵일 수 있다.
- [0072] 또한, 제2 변환 특징 맵(320T)은 넓이 방향을 기준으로 전치 변환된 특징 맵이다. 입력 데이터의 넓이 방향에서 중요한 패턴이나 특징을 추출하고 이를 변환하여 생성된 맵이다.
- [0073] 이 두 변환은 입력 데이터의 공간적 구조를 다르게 해석하여, 서로 다른 방향에서의 특징을 학습하게 할 수 있다. 따라서 각각의 특징 맵의 통합을 통해 신경망이 데이터의 높이와 넓이 방향 모두에서 중요한 정보를 놓치지 않고 추론할 수 있게 한다.
- [0074] 서버(300)는 제1 변환 특징 맵(높이 방향)과 제2 변환 특징 맵(넓이 방향)의 결합 방식을 결정할 수 있다.
- [0075] 제1 통합 방식으로 연결(Concatenate)은 두 특징 맵을 단순히 연결하는 방식일 수 있다. 예를 들어, 높이 방향의 특징과 넓이 방향의 특징을 새로운 차원으로 이어 붙일 수 있다. 이를 통해 두 특징 맵의 정보를 모두 유지하면서, 신경망이 두 방향에서 얻은 패턴을 학습할 수 있게 할 수 있다. 이 방식은 정보 손실 없이 두 변환 특징 맵의 모든 정보를 그대로 보존하게 된다.
- [0076] 또는 제2 통합 방식으로 합산(Add)은 두 특징 맵을 더하는 것으로 수행될 수 있다. 이 경우, 동일한 차원의 두 특징 맵을 요소별로 더해, 하나의 통합된 맵을 생성할 수 있다.
- [0077] 합산 연산을 사용하면 두 특징 맵이 합쳐지면서 각 특징의 중요도를 더 직관적으로 반영할 수 있다. 이 방식은 정보가 압축되면서도 두 방향의 패턴을 함께 고려할 수 있는 장점을 제공할 수 있다.
- [0078] 제3 통합 방식으로 제1 변환 특징 맵(320T')과 제2 변환 특징 맵(320T)과 함께 전치 맵에 대한 통합 특징 맵을 연결하는 하이브리드 방식을 적용하는 것도 가능하다.
- [0079] 예를 들어 제1 변환 특징 맵(높이 방향)과 제2 변환 특징 맵(넓이 방향)을 합산 연산을 통해 통합 맵을 생성할 수 있다.
- [0080] 즉, 두 맵의 요소별 값을 더하여 하나의 통합된 중간 맵을 생성할 수 있으며 이를 통해 두 방향의 패턴 정보를 통합하는 동시에 각 특징 맵의 중요도를 고려해 정보를 결합한다.
- [0081] 이후 서버(300)는 중간 맵과 각 전치 변환된 특징맵을 연결 연산을 통한 최종 통합을 수행한다.
- [0082] 서버(300)는 합산 연산으로 결합된 중간맵을 제1 변환 특징 맵(320T')과 제2 변환 특징 맵(320T)과 새로운 차원에서 연결함으로써 최종 통합한다. 이 과정에서는 두 맵의 정보를 결합하여 하나의 통합된 맵을 만들며, 두 방향에서 얻어진 특징 정보를 모두 보존될 수 있다.

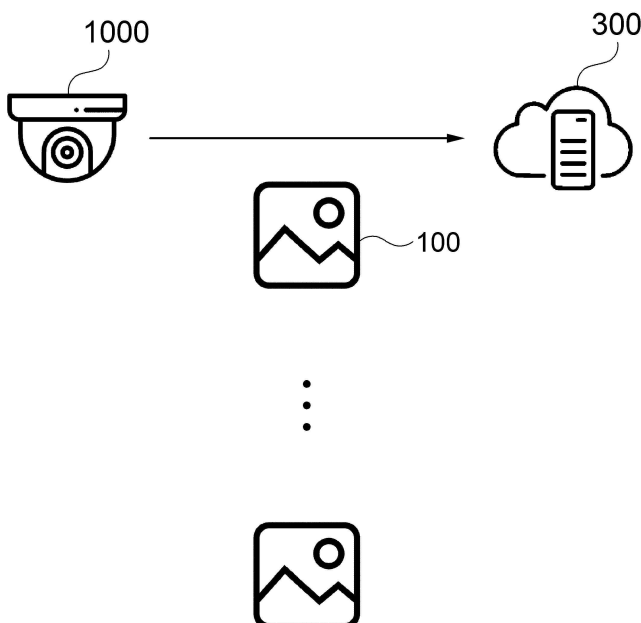
- [0083] 단순 합산 연산만 수행할 경우, 각 방향의 특징 정보가 압축되면서 일부 정보가 손실될 수 있어 합산 연산 후 연결함으로써 새로운 차원에서 두 방향의 맵을 통합함으로써, 정보 손실을 방지하고 각 특징 맵의 원래 정보를 보존할 수 있다.
- [0084] 또한 추가적인 실시예에서 서버(300)는 특징 맵으로부터 생성된 각각의 출력을 통합하여 최종 출력을 제공할 수 있다(S500).
- [0085] 도 6을 참고하면 본 실시예에서 서버(300)는 제1 변환 특징 맵(320T')과 제2 변환 특징 맵(320T)을 각각 신경망(330)에 입력하고 신경망의 추론 결과를 이용하여 최종적으로 하나의 통합된 출력을 생성할 수 있다.
- [0086] 나아가, 본 실시예에 따라 추론을 위한 신경망 모델은 통합된 특징 맵 내 특징값들의 시간적인 관계성을 추출하는 트랜스포머 기반의 모델일 수 있다.
- [0087] 즉, 트랜스포머 모델이 전치 변환된 특징 맵에서 시간적 관계를 효과적으로 학습하여 행동을 추론하는 것도 가능하다.
- [0088] 트랜스포머는 입력된 데이터(여기서는 전치 변환된 특징 맵)에서 시간적 관계성을 추출할 수 있다.
- [0089] 전치 변환된 특징 맵은 원래의 데이터가 특정 방향(높이 또는 넓이 방향)으로 전치 변환되어 시간적 순서에 따른 공간적 변화가 명확히 드러나도록 처리된 것이다. 트랜스포머 모델은 이 전치 변환된 입력을 받아, 시간적 순서에 따른 각 특징값들 간의 상관관계를 학습할 수 있다.
- [0090] 어텐션 메커니즘은 입력된 데이터의 각 요소가 다른 요소와 얼마나 중요한 관계를 가지는지를 동적으로 계산할 수 있다. 입력된 각 특징값(feature) 간의 상호작용을 평가하여, 그 관계성을 반영한 가중치(weight)를 할당할 수 있다.
- [0091] 어텐션은 전치 변환된 특징 맵 내의 각 시간적 특징값이 다른 시간적 특징값과 어떤 관계를 가지는지 계산할 수 있다. 예를 들어, 과거의 동작이 현재 동작에 어떤 영향을 미치는지, 또는 현재 동작이 미래의 행동을 어떻게 예측할 수 있는지를 학습하게 된다.
- [0092] 트랜스포머 모델의 어텐션 메커니즘은 모든 시간적 특징값 간의 관계를 동시에 고려할 수 있으므로, 복잡한 시간적 패턴을 효과적으로 인식하고, 행동 인식의 정확도를 높인다.
- [0093] 트랜스포머 모델은 짧은 시간 간격 내에서 일어나는 빠른 움직임뿐만 아니라, 긴 시간 간격에 걸쳐 발생하는 행동의 변화를 모두 학습할 수 있다.
- [0094] 어텐션 메커니즘은 특정 행동이 발생하는 시간적 맥락을 인식할 수 있다. 예를 들어, 달리기 중 팔의 흔들림이 발생하는 시점과 속도에 따른 다리의 움직임 변화를 동시에 학습하여, 더 정교한 행동 인식을 가능하게 할 수 있다.
- [0095] 이상 본 발명은 시간적 정보를 우선적으로 고려한 구조를 통해 행동 인식 시스템의 정확도를 높일 수 있으며, 실시간 시스템에서의 행동 인식 오차를 줄일 수 있다.
- [0096] 이하 도 7을 참고하여 본 발명의 일 실시예에 따른 서비스 서버(300)의 하드웨어 구성에 대하여 설명한다.
- [0097] 도 7을 참조하면, 본 발명의 몇몇 실시예들에서 서비스 서버(300)는 컴퓨팅 장치의 형태로 구현될 수 있다. 서비스 서버(300)를 구성하는 각각의 모듈 중 하나 이상은 범용 컴퓨팅 프로세서 상에서 구현되며 따라서 프로세서(processor)(308), 입출력 I/O(302), 메모리 (memory)(340), 인터페이스(interface)(306) 및 버스(314, bus)를 포함할 수 있다. 프로세서(308), 입출력 장치(302), 메모리 (340) 및/또는 인터페이스(306)는 버스(314)를 통하여 서로 결합될 수 있다. 버스(314)는 데이터들이 이동되는 통로(path)에 해당한다.
- [0098] 구체적으로, 프로세서(308)는 CPU(Central Processing Unit), MPU(Micro Processor Unit), MCU(Micro Controller Unit), GPU(Graphic Processing Unit), 마이크로프로세서, 디지털 신호 프로세서, 마이크로컨트롤러, 어플리케이션 프로세서(AP, application processor) 및 이들과 유사한 기능을 수행할 수 있는 논리 소자들 중에서 적어도 하나를 포함할 수 있다.
- [0099] 입출력 장치(302)는 키패드(keypad), 키보드, 터치스크린 및 디스플레이 장치 중 적어도 하나를 포함할 수 있다. 메모리 장치(340)는 데이터 및/또는 프로그램 등을 저장할 수 있다.
- [0100] 인터페이스(306)는 통신 네트워크로 데이터를 전송하거나 통신 네트워크로부터 데이터를 수신하는 기능을 수행

할 수 있다. 인터페이스(306)는 유선 또는 무선 형태일 수 있다. 예컨대, 인터페이스(306)는 안테나 또는 유무선 트랜시버 등을 포함할 수 있다. 메모리 (340)는 프로세서(308)의 동작을 향상시키되, 개인정보의 보호를 위한 휘발성의 동작 메모리로서, 고속의 디램 및/또는 에스램 등을 더 포함할 수도 있다.

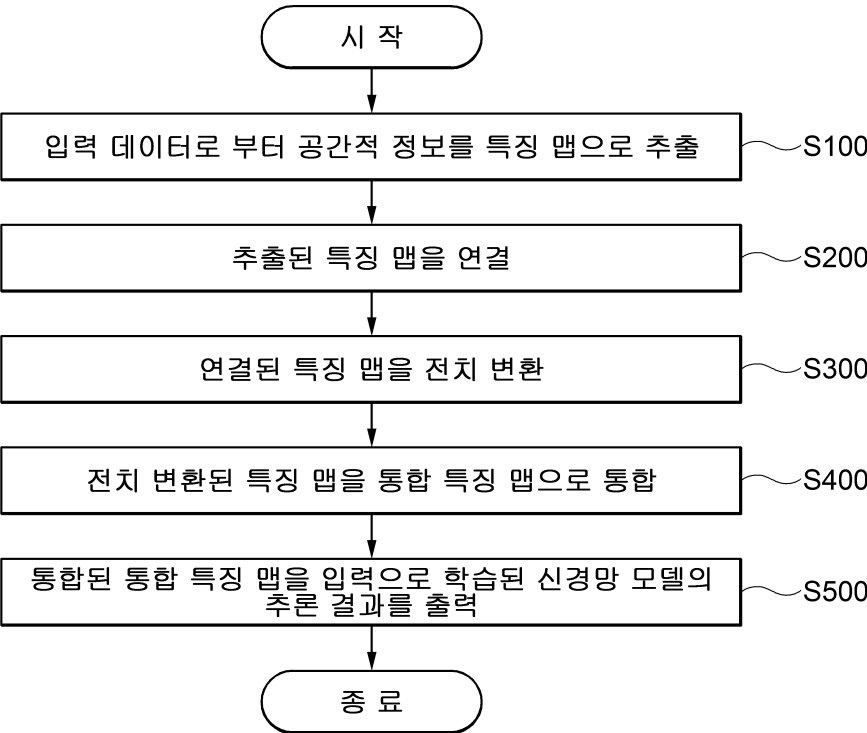
- [0101] 또한, 메모리(340) 내에는 여기에 설명된 일부 또는 모든 모듈의 기능을 제공하는 프로그래밍 및 데이터 구성을 저장한다. 예를 들어, 상술한 학습 방법의 선택된 양태들을 수행하도록 하는 로직을 포함할 수 있다.
- [0102] 메모리 (340)에 저장된 상술한 방법을 수행하는 각 단계를 포함하는 명령어들의 집합으로 프로그램 또는 어플리케이션을 로드하고 프로세서가 각 단계를 수행할 수 있도록 한다.
- [0103] 이상 여기에 설명되는 다양한 실시예는 예를 들어, 소프트웨어, 하드웨어 또는 이들의 조합된 것을 이용하여 컴퓨터 또는 이와 유사한 장치로 읽을 수 있는 기록매체 내에서 구현될 수 있다.
- [0104] 하드웨어적인 구현에 의하면, 여기에 설명되는 실시예는 ASICs (application specific integrated circuits), DSPs (digital signal processors), DSPDs (digital signal processing devices), PLDs (programmable logic devices), FPGAs (field programmable gate arrays, 프로세서(processors), 제어기(controllers), 마이크로 컨트롤러(micro-controllers), 마이크로 프로세서(microprocessors), 기타 기능 수행을 위한 전기적인 유닛 중 적어도 하나를 이용하여 구현될 수 있다. 일부의 경우에 본 명세서에서 설명되는 실시예들이 제어 모듈 자체로 구현될 수 있다.
- [0105] 소프트웨어적인 구현에 의하면, 본 명세서에서 설명되는 절차 및 기능과 같은 실시예들은 별도의 소프트웨어 모듈들로 구현될 수 있다. 상기 소프트웨어 모듈들 각각은 본 명세서에서 설명되는 하나 이상의 기능 및 작동을 수행할 수 있다. 적절한 프로그램 언어로 쓰여진 소프트웨어 어플리케이션으로 소프트웨어 코드가 구현될 수 있다. 상기 소프트웨어 코드는 메모리 모듈에 저장되고, 제어모듈에 의해 실행될 수 있다.
- [0106] 이상의 설명은 본 발명의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명이 속하는 기술 분야에 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위 내에서 다양한 수정, 변경 및 치환이 가능할 것이다.
- [0107] 따라서, 본 발명에 개시된 실시 예 및 첨부된 도면들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시 예 및 첨부된 도면에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 보호 범위는 아래의 청구 범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리 범위에 포함되는 것으로 해석되어야 할 것이다.

## 도면

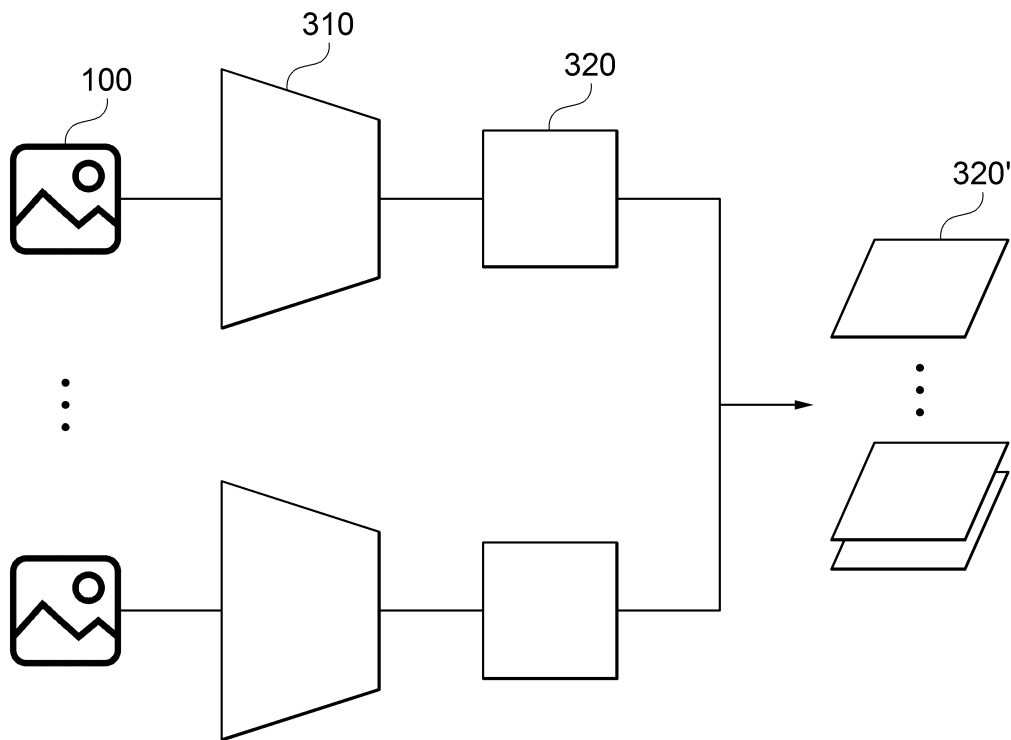
### 도면1



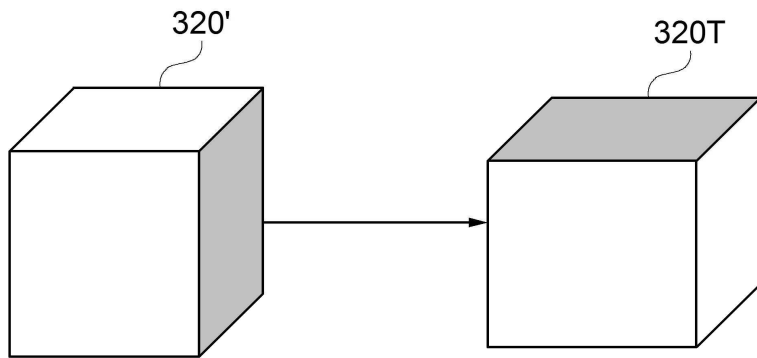
도면2



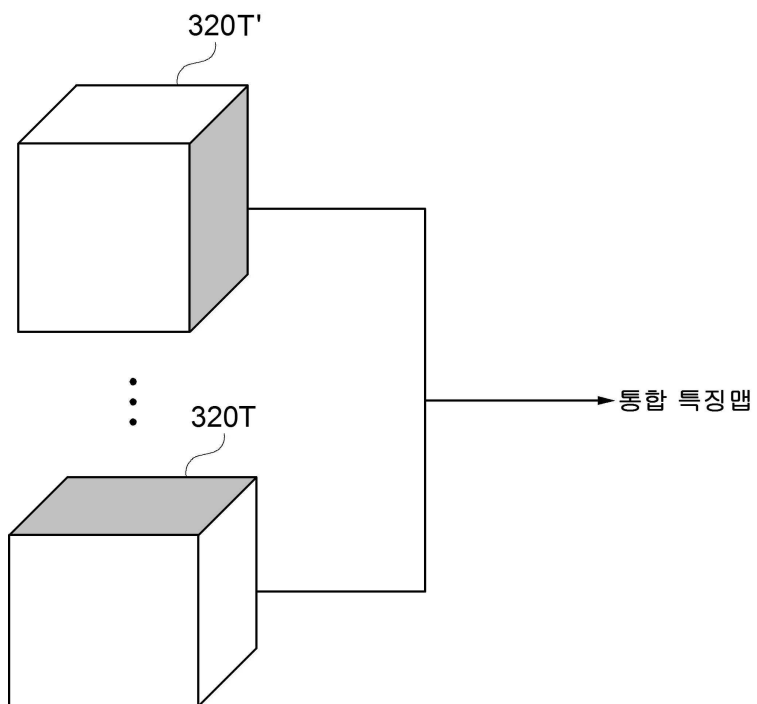
도면3



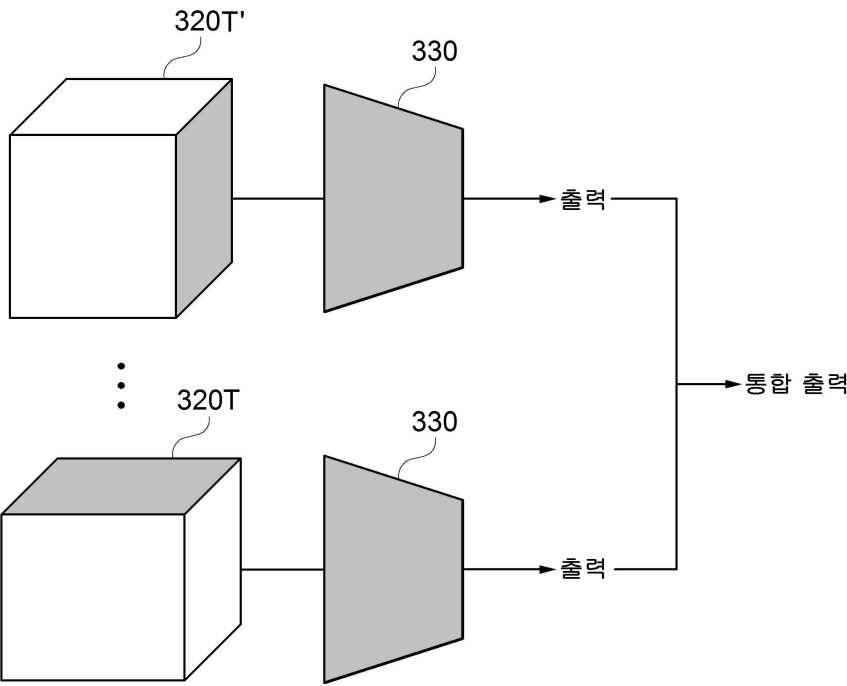
도면4



도면5



도면6



도면7

300

